A novel, biologically-informed polygenic score reveals role of mesocorticolimbic insulin receptor gene network on impulsivity and addiction

Kathryn McCracken[1,2] #, Shantala A. Hari Dass, PhD[4] #, Irina Pokhvisneva, MSc[4]; Lawrence M. Chen, MSc[4]; Elika Garg[4]; Thao T. T. Nguyen, MSc[4]; Moein Yaqubi, MSc[4]; Lisa M. McEwen, MSc[8]; Julie L. MacIsaac, PhD[8]; Josie Diorio, BSc[4]; Michael S. Kobor, PhD[8]; Kieran J. O'Donnell, PhD[3,4,5,6]; Michael J. Meaney, PhD, CQ, FRSC[3,4,5,6,7]; Patricia P. Silveira, MD, PhD[3,4,5*].

1. John Abbott College, Sainte-Anne-de-Bellevue, QC, Canada.
2. McGill University, Montreal, QC, Canada.
3. Department of Psychiatry, Faculty of Medicine, McGill University, Montreal, QC, Canada.
4. Ludmer Centre for Neuroinformatics and Mental Health, Douglas Mental Health University Institute, McGill University, Montreal, QC, Canada
5. Sackler Program for Epigenetics & Psychobiology at McGill University
6. Child and Brain Development Program, Canadian Institute for Advanced Research (CIFAR), Toronto, ON, Canada
7. Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research (A*STAR)
8. Centre for Molecular Medicine and Therapeutics, BC Children's Hospital Research Institute, Department of Medical Genetics, The University of British Columbia, Vancouver, BC, Canada

#These authors contributed equally to the work.

*Corresponding author:

Patricia Pelufo Silveira, MD, PhD

Department of Psychiatry, Faculty of Medicine, McGill University

Douglas Hospital Research Centre, 6875 Boulevard LaSalle, Montreal, QC, H4H 1R3,

Canada.

Phone: 514-761-6131 (ext.2776)

Fax: 514-761-6131

patricia.silveira@mcgill.ca

Word count:3000

Key points

Question: Considering the modulation of mesocorticolimbic dopaminergic pathways by insulin through the action on its receptors (IR), we investigated if a novel, region specific polygenic score on the IR-related gene network (ePRS-IR) is associated with dopamine-related behaviors (impulsivity and addiction).

Findings: The ePRS-IR showed improved prediction of childhood impulsivity and risk for early addiction onset in comparison to conventional polygenic risk scores for ADHD or addiction.

Meaning: This novel genomic approach reveals insulin action as a biological process involved in the risk for dopamine-related psychopathology.

Abstract

Importance: Activation of brain insulin receptors occurs on mesocorticolimbic regions, modulating reward sensitivity and inhibitory control. Variations in the functioning of this mechanism likely associate with individual differences in the risk for related psychopathologies (attention-deficit hyperactivity disorder, addiction), an idea that agrees with the high co-morbidity between insulin resistant states and psychiatric conditions. While genetic studies comprise an interesting tool to explore neurobiological mechanisms in community samples, the conventional genome-wide association studies and polygenic risk score methodologies completely ignore the fact that genes operate in networks, and code for precise biological functions in specific tissues.

Objective: We propose a novel, biologically informed genetic score reflecting the mesocorticolimbic insulin receptor-related gene network, and investigate if it predicts dopamine-related psychopathology (impulsivity and addiction) in community samples.

Design: Birth cohort (Maternal Adversity, Vulnerability and Neurodevelopment, MAVAN) and adult cohort (Study of Addiction, Genes and Environment, SAGE).

Setting: General community.

Participants: 212 4-year-old children (MAVAN), and 1626 adults (SAGE).

Exposure: The biologically informed, mesocorticolimbic specific, insulin receptor polygenic score was created based on levels of co-expression with the insulin receptor in striatum and prefrontal cortex, and calculated in the two samples using the genotype data (Psychip/Psycharray).

Main outcome: childhood impulsivity in the Information Sampling task, and risk for early addiction onset.

Results: The insulin receptor polygenic score showed improved prediction of childhood impulsivity in boys and risk for early addiction onset in males in comparison to conventional polygenic risk scores for attention-deficit hyperactivity disorder or addiction.

Conclusions and relevance: This novel genomic approach reveals insulin action as a relevant biological process involved in the risk for dopamine-related psychopathology.

Introduction

The co-morbidity between metabolic and psychiatric disorders is well-established, but the mechanisms are poorly understood. It is particularly interesting the high co-occurrence of several psychiatric conditions with insulin resistance[1-3]. Insulin receptors are expressed throughout the brain[4], in areas such as the ventral tegmental area (VTA), prefrontal cortex (PFC)[5,6] and hippocampus[7,8]. Insulin is actively transported across the blood-brain barrier[9]. Insulin action on its mesocorticolimbic receptors modulates synaptic plasticity in dopaminergic neurons, affecting DA-related behaviors such as response to reward[10], impulsivity[11], mood[12,13], cognition[14] and decision-making[15].

Genetic studies can be an interesting tool to investigate the neurobiological mechanisms that explain the co-morbidity between metabolic and psychiatric conditions. Genome-wide association studies (GWAS) provide the basis for novel, cumulative scores that reflect genetic predispositions[16]. The effect sizes from these studies can be used to estimate the genetic risk of the individual through polygenic risk scores (PRS), by multiplying the measured number of risk alleles at a locus by the effect size of the association between a particular genotype and the outcome, and summing over all single nucleotide polymorphisms (SNPs) at a certain significance threshold [17]. One problem of the GWAS and PRS methodology is that it identifies statistically significant associations between scattered SNPs and a certain condition or trait, completely ignoring the fact that genes operate in networks, and code for precise biological functions in specific tissues.

Our hypothesis was that variations in the functioning of the mesocorticolimbic insulin receptor gene network would be associated with differences in dopamine-related behavioral outcomes, namely impulsivity and risk for addiction. Although it is described an association

between conventional polygenic risk scores (PRS) for ADHD and impulsivity problems in population-based samples of children[18], theses scores do not provide information about the underlying network in which these genes operate, nor the neurobiology of the disease. We propose a novel genomics approach that provides a biologically-informed genetic score based on genes co-expressed with the insulin receptor in specific mesocorticolimbic regions to foster neurobiological analysis of behavioral phenotypes linked to psychopathology.

Methods

Samples

Main Cohort: We used data from the prospective Maternal Adversity, Vulnerability and Neurodevelopment (MAVAN) birth cohort[19] that followed the children at different time points in the first years of life in Montreal (Quebec) and Hamilton (Ontario), Canada. Approval for the MAVAN project was obtained from the study hospitals and by the ethics committees and university affiliates (McGill University and Université de Montréal, the Royal Victoria Hospital, Jewish General Hospital, Centre hospitalier de l'Université de Montréal, Hôpital Maisonneuve-Rosemount, St Joseph's Hospital and McMaster University). Informed consent was obtained from all subjects. A total of 212 children had complete data (genotype and impulsivity task at 4 years of age, see below) and were included in this analysis.

Replication Cohort: We used the Study of Addiction: Genetics and Environment (SAGE) repository[20-26], acquired from dbGaP (https://www.ncbi.nlm.nih.gov/gap, Accession number: phs000092.v1.p). The SAGE dataset was compiled from three studies: Collaborative Study on the Genetics of Alcoholism (COGA), Family Study of Cocaine Dependence (FSCD), and Collaborative Genetic Study of Nicotine Dependence (COGEND). The SAGE dataset contains

genotyping and clinical phenotypes related to substance dependence for adult subjects. To maximize consistency with the MAVAN cohort we only used a subset of subjects of caucasian ethnicity (n = 1627). We received access to the SAGE dataset based on the approval of our Data Access Request (DAR) by the NIH Data Access Committee. We agree with the stipulations of the Data Use Certification.

Genotyping: In MAVAN, we genotyped 242,211 autosomal SNPs using genome-wide platforms (PsychArray/PsychChip, Illumina) according to manufacturer's guidelines with 200ng of genomic DNA derived from buccal epithelial cells and our quality control procedures. Specifically, we removed SNPs with a low call rate (<95%), low p-values on Hardy-Weinberg Equilibrium exact test (p<1e-40), and minor allele frequency (<5%). Afterward, we performed imputation using the Sanger Imputation Service[27] resulting in 20,790,893SNPs with an info score >0.80 and posterior genotype probabilities >0.90.

ePRS-IR calculation (Figure 1): The polygenic risk score based on genes co-expressed with the insulin receptor (ePRS-IR) was created using gene co-expression databases including 1) GeneNetwork (http://genenetwork.org), 2) BrainSpan (http://brainspan.org), 3) NCBI Variation Viewer (https://www.ncbi.nlm.nih.gov/variation/view/). These resources allowed us to identify genes co-expressed with the IR in the striatum and prefrontal cortex (PFC) regions in mice (GeneNetwork) and humans (BrainSpan), and to identify SNPs for these genes in humans (NCBI Variation Viewer). The PRS was constructed as follows: 1) we used GeneNetwork to generate co-expression matrix with IR in the i) ventral striatum, ii) PFC in mice (absolute value of the co-expression correlation $r \geq 0.5$)[28], 2) we then used BrainSpan to identify consensus transcripts from this list with a child and fetal enrichment within the human brain. We selected autosomal transcripts differentially expressed in these brain regions at ≥1.5 fold during child and fetal

development as compared to adult samples[29]. The final list included 184 genes. Based on their functional annotation in the National Center for Biotechnology Information, U.S. National Library of Medicine (https://www.ncbi.nlm.nih.gov/variation/view/) using GRCh37.p13 we gathered all the existing SNPs from these genes (total = 17,258) and subjected this list of SNPs to linkage disequilibrium clumping, which uses the lowest association p-values in the ADHD GWAS to inform removal of highly correlated SNPs (r^2>0.2) across 500kb regions[30], resulting in 371 independent functional SNPs based on the children's genotype data from MAVAN. Out of the 371 SNPs, 318 (86%) were genotyped in the SAGE dataset and used to compute the ePRS-IR in the replication cohort. The median of the ePRS-IR computed in MAVAN and SAGE datasets were comparable (ePRS-IR; MAVAN median = 0.065, SAGE median = 0.061). We used a count function of the number of alleles at a given SNP weighted by the effect size of the association between the individual SNP and ADHD[31]. All SNPs were subjected to linkage disequilibrium clumping (r^2>0.2 across 500kb) so only independent SNPs that are most associated to ADHD, based on the association p-values in the ADHD GWAS, comprised the PRS (Figure 1).

Other genetic scores: We generated other PRSs using our accelerated pipeline (https://github.com/MeaneyLab/PRSoS)[32], for each subject: (a) A random list of 184 genes that were differentially expressed in the PFC/Striatum in the prenatal and childhood developmental stages, selected using BrainSpan[29]. Adopting the same methodology that was implemented to generate the ePRS-IR, we computed a "random genes" ePRS based on the ADHD GWAS[31]; (b) conventional PRS for ADHD[31] (termed $PRS_{ADHD2010}$); (c) another PRS for ADHD based in a more recent GWAS[33] (termed $PRS_{ADHD2017}$) and (d) PRS for onset of tobacco smoking[34] (termed $PRS_{Tobacco}$) of MAVAN children. All PRSs were created based on their SNPs and meta-analysis

available at the Psychiatric Genomics Consortium[17,31]. The PRS are cumulative summary scores computed as the sum of the allele count weighted by the effect size across SNPs at selected p-value thresholds (P_T) based on the relevant GWAS[30].

Reflexive impulsivity measure in children: The Information Sampling Task (IST) from the CANTAB was designed to measure reflection impulsivity and decision making[35]. Rather than relying on speed-accuracy indices, IST measures reflection impulsivity by calculating the probability of the subject selecting the correct answer after making a decision based on the information sampled prior to making that decision. On each trial, children are presented with a 5×5 matrix of grey boxes on the computer screen, and two larger colored panels at the foot of the screen. They are told that it is a game for points, won by correctly choosing the color under the majority of the grey boxes. Touching a grey box immediately opens that box to reveal one of the two colors displayed at the bottom of the screen. Subjects could open boxes at their own will with no time limit before deciding between one of the two colors, indicating their decision by touching one of the two panels at the bottom of the screen. When they do, the remaining boxes are uncovered and a message is displayed to inform them whether or not they were correct. We particularly focused on the "fixed win" condition during IST, in which subjects are awarded 100 points for a correct decision regardless of the number of boxes opened (no penalties according to the number of boxes). The primary performance outcome measure was the mean probability of being correct at the point of decision (P Correct). P Correct is the probability that the color chosen by the subject at the point of decision would be correct, based only on the evidence available to the subject at the time (i.e., dependent on the amount of information they had sampled). There was a recent update in the mean P Correct formula, which was endorsed by the original authors of the measure[36,37], therefore in this study we calculated and used the new mean

P Correct. We also excluded the trials where discrimination errors occurred since the presence of this type of error runs contrary to the task instructions and influences the mean P Correct values, being common in children[38].

Addiction onset: The age of onset of substance dependence was computed as the earliest age of onset of alcohol, nicotine, marijuana, cocaine, opiate or other substance use. The clinical assessment of substance dependency was based on a Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA II)[39] and adapted versions of the SSAGAII, which assesses the physical, psychological and social manifestations of substance dependence.

Gene Network Analysis: We extracted a list of the 184 genes from the SNPs with the lowest p-values based on the post clumped results of the ADHD GWAS[31]. RNA-sequencing data was downloaded from BrainSpan, including samples from 8 postconceptional weeks to 11 years old within prefrontal cortex (dorsolateral, ventrolateral, anterior cingulate cortex and orbitofrontal cortex) and striatum[29], for three gene lists: (a) ePRS-IR (see supplementary Table 1); (b) Random gene list (as detailed above) and (c) The top 184 genes associated with the most significant SNPs in the ADHD GWAS[31]. A median expression value was computed across the mentioned brain regions. The protein-protein interaction data were retrieved from STRING[40] (https://string-db.org/) and GeneMANIA[41] (https://genemania.org) databases and the protein-protein interaction networks were constructed and visualized in the Cytoscape software[42]. One-way ANOVA was used to compare the values of the number of connections across the three gene lists.

Statistical Analysis:

Main Cohort: Statistical analysis of the baseline characteristics was performed using Spearman's correlation. All PRSs were coded into quartile variables. The population structure of

the MAVAN cohort was evaluated using principal component analysis of all autosomal SNPs that passed the quality control, without low allele frequency (MAF>5%) and are not in high linkage disequilibrium (r^2>0.2) within a window of 50 SNPs at each step size of 5[43]. Based on the inspection of the screeplot, the first three principal components were the most informative of population structure in this cohort and were included in all analysis. Linear regression analysis was performed to explore if the gender by ePRS-IR interaction was associated with the IST outcome, adjusting for population structure. In order to assess if there was a linear trend between ePRS-IR and the outcome, i.e mean P Correct derived from the IST, the ePRS-IR was categorized into quartiles. The main outcome measure was the mean P Correct derived from the IST during the fixed win condition (see above).

Replication Cohort: Statistical analysis of the potential confounders was performed using Spearman correlation. Similar to the analysis on MAVAN cohort, the ePRS-IR was categorized into quartile to assess if there was a linear trend between ePRS-IR and the outcome, i.e age of onset of substance dependence. The age of onset of dependency was categorized into “early”, “mid” and “late” based on the distribution tertiles. The population structure of the SAGE cohort was evaluated using principal component analysis of all autosomal SNPs that passed the quality control, and these scores were provided in the SAGE data repository. Based on the inspection of the screeplot, the first three principal components were the most informative of population structure in this cohort and included in all subsequent analysis. Ethnic outliers (>6 standard deviations) were excluded from the analysis. Linear regression analysis was performed to explore if the gender by ePRS-IR interaction was associated with the age of onset of dependence, adjusting for the population structure and study source. Simple slopes were analyzed to test the significance of the association in males and females. The ratio of the probability of early onset to

late onset dependence was computed. A value >1 describes a population with increased prevalence of early onset dependence compared to late onset dependence. Data were analyzed using the Statistical Package for the Social Sciences (SPSS) version 20.0 software (SPSS Inc., Chicago, IL, USA) and R. Significance levels for all measures were set at $\alpha <= 0.05$.

Results

Baseline comparisons

In both the MAVAN and SAGE datasets, major potential confounders did not vary according to ePRS-IR, or were used as co-variates in the main analysis (Table 1). The number of boxes opened on the IST was significantly correlated with incorrect judgments in the fixed win condition ($r(198) = -0.839$, $p<0.0001$), demonstrating that response accuracy is a function of the extent of information analysis, a feature of reflection impulsivity.

Relationship between ePRS-IR and childhood impulsivity – MAVAN cohort

There was a significant interaction effect between ePRS-IR and gender on IST ($\hat{\beta} = 0.036$, $p = 0.02$); while a simple slopes analysis showed no relationship between the ePRS-IR score and mean P Correct values in girls ($\hat{\beta} = 0.01$, $p = 0.31$), a high ePRS-IR was significantly related to lower mean P Correct (less certainty when coming up to a decision or higher reflexive impulsivity) in boys ($\hat{\beta} = -0.02$, $p = 0.02$) (Figure 2A). We next examined whether a conventional PRSs for ADHD of comparable size in terms of SNP number would also predict impulsivity in these children. As shown in Table 2, using two different GWAS's for ADHD[31,33], we found no association between the conventional PRS and reflection impulsivity on the CANTAB in this sample. Considering the clinical overlap between impulsive phenotypes and

risk for addiction[44], we also created a conventional PRS for tobacco smoking, but this was also not associated with reflection impulsivity on the IST task. Finally, to demonstrate that the results of the ePRS-IR analysis are not merely due to the presence of random genes that are differentially expressed in the PFC/striatum of prenatal and childhood developmental stages, but are instead dependent on selecting a biologically relevant network, we computed a PRS based on a random selection of genes differentially expressed in the PFC/striatum of neonates and children. This random PRS did not interact with gender to predict mean P Correct values in the MAVAN cohort (Table 2).

Relationship between ePRS-IR and age of addiction onset – SAGE cohort

We then hypothesized that a high ePRS-IR, associated with childhood impulsivity in boys as shown above, would predict early age of addiction onset. We used data from the SAGE repository to validate the above interaction between sex and ePRS-IR in an independent cohort. The analysis revealed a significant interaction between the ePRS-IR score and gender (interaction effect; $\hat{\beta} = -0.08$, p = 0.01). Further simple slopes analysis showed that a higher ePRS-IR was significantly associated with earlier onset of dependence only in males (males, simple slope $\hat{\beta} = -0.05$, p = 0.03; females, simple slope $\hat{\beta} = 0.03$, p = 0.16). In males, a higher ePRS-IR was significantly associated with earlier onset of dependence (Figure 2B and C).

ePRS-IR gene network analysis

We compared the list of 184 genes included in the ePRS-IR with (1) a random selection of 184 genes differently expressed in the PFC in fetal/childhood phases and (2) the top 184 genes associated with the most significant SNPs from the ADHD GWAS[31] in a gene network analysis

(Figure 3A). The number of interactions between the genes that comprise the ePRS-IR is significantly higher than the control random list or the genes form the ADHD GWAS ($p<0.05$, Figure 3B). Enrichment analysis of the SNPs comprising ePRS-IR using Metacore® (Thomson Reuters) showed statistically significant enrichment for pathways involved in cell cycle regulation (FDR q = 3.91e-3). Two process networks were highly significant: translation initiation (FDR q = 2.37e-23) and elongation termination (FDR q = 1.81e-17).

Discussion

We showed that a biologically-informed polygenic risk score based on genes co-expressed with the IR in mesocorticolimbic regions is more strongly associated with impulsivity in boys and the risk for early onset of substance dependence in men from an independent cohort than is the conventional PRS for either ADHD or addiction. The sex-specificity of our findings was expected, given the increased prevalence of ADHD and behavioral alterations associated with this condition (such as impulsivity) in boys compared to girls[45,46].

Our network analysis shows that the ePRS-IR represents a cohesive gene network with significantly more connections than the list of genes extracted from the GWAS for ADHD, or a random list of developmentally relevant genes. This robust approach therefore goes beyond describing associations between single gene variants and the outcomes, but captures information about the whole gene network, and its function, in specific brain regions.

Conduct disorder and impulsivity are the foremost risk traits for alcohol use disorder among the 80 personality disorder criteria of DSM-IV [47]. There is a relationship between childhood ADHD and the risk for developing drug addiction later in life [48], especially considering the impulsivity component, rather than inattention [49], in agreement to the findings

described here. Insulin function is associated with the risk for drug addiction [50]. Diminished insulin sensitivity is related to less endogenous dopamine at D2/3 receptors in the ventral striatum[51], reinforcing the idea that metabolic processes are involved in dysfunctions of the mesocorticolimbic system, such as drug dependence. The cell cycle regulation process enriched in our ePRS-IR is consistent with findings showing that insulin is involved in proliferation of many cell types, and activates several cell-cycle regulators.

Our genomic approach integrates information from molecular neurobiology with GWAS technology to develop a biologically-informed polygenic score based on gene co-expression data from specific brain regions. This approach creates a novel genomic measure to identify genetic vulnerability for childhood behavioral phenotypes that predict later psychiatric conditions in community-based samples, highlighting possible targets for drug development.

Figure legends

Figure 1. Flowchart depicting the steps involved in creating the expression-based and mesocorticolimbic-specific polygenic risk score based on genes co-expressed with the insulin receptor (ePRS-IR) using gene co-expression databases. 1) GeneNetwork was used to generate a co-expression matrix with insulin receptor (IR) in the ventral striatum and in the prefrontal cortex in mice (absolute value of the co-expression correlation $r \geq 0.5$). 2) BrainSpan was then used to identify consensus autosomal transcripts from this list with developmental enrichment within the same brain areas, selecting transcripts differentially expressed at ≥1.5 fold during child and fetal development as compared to adult samples. The final list included 184 genes. 3) Based on their functional annotation in the National Center for Biotechnology Information, U.S. National Library of Medicine using GRCh37.p13, we gathered all the existing SNPs from these genes (total = 17,258) and subjected this list of SNPs to linkage disequilibrium clumping, to inform removal of highly correlated SNPs ($r^2>0.2$) across 500kb regions, resulting in 371 independent functional SNPs based on the children's genotype data from MAVAN (Study Sample ids). 4) We used a count function of the number of alleles at a given SNP (rs1, rs2…) weighted by the effect size (ES) of the association between the individual SNP and ADHD. The sum of these values from the total number of SNPs provides the ePRS-IR score. Numbers in superscript correspond to citations in the References list.

Figure 2. A) Performance in the Information Sampling Task (IST, CANTAB) according to gender and ePRS-IR categories. There is a significant interaction between the genetic score and gender ($p = 0.02$, $\hat{\beta} = 0.03$) in which boys with high ePRS-IR show decreased mean P Correct (or sample less information before taking the decision, being more impulsive) (Males: $p = 0.02$ and simple slope = −0.02, Females: $p = 0.31$ and simple slope = 0.01). In males, but not females,

there is a linear dose dependent association between the ePRS-IR and P Correct. (Linear trend analysis; Males: F = 4.4, p = 0.04, Females: F = 0.76, p = 0.38). Males are depicted in blue and females in red; B**)** Age onset of substance dependence according to gender and ePRS-IR categories. There is a significant interaction between the genetic score and gender (p = 0.01, $\hat{\beta}$= −0.13) in which males with high ePRS-IR show greater probability of early onset and lower probability of late onset dependencies; males, but not females, show linear dose dependent association between the ePRS-IR and age of onset of substance dependence; linear trend analysis; Males: F = 6.7, p = 0.01, Females: F = 1.1, p = 0.29). The dotted line at y = 1 represents a population with equal number of subjects with early and late onset dependency; C) Males (blue): p = 0.01 and simple slope = −0.08, Females (red): p = 0.33 and simple slope = 0.03].

Figure 3. Gene Network analysis in the ePRS-IR, random and ADHD (top 184 genes associated with SNPs from the GWAS) gene lists. A) Schematic representation of the gene networks and B) comparison of the number of connections between the genes in each network. The ePRS-IR represents a network with significantly higher connectivity than a random list of genes differently expressed in the same brain areas and developmental period. The ePRS-IR gene network has also higher connectivity than the top genes from the ADHD GWAS.

Tables

Study participants' characteristics correlation with ePRS-IR

Sample descriptives	Spearman correlation coefficient	P-value
MAVAN cohort		
Gender	-0.12	0.04
Birth weight (grams)	0.03	0.53
Gestational age (weeks)	-0.08	0.18
Family income below Low Income Cut Off[52]	-0.20	0.25
SAGE cohort		
Gender	10.01	0.36
Family income below $20K	-0.06	0.07
Age at interview	-0.006	0.69
Study source	0.052	0.0007

Table 2 - Description of the polygenic risk scores used and analysis of the interactions.

PRS	**P value threshold**	**Number of SNPs**	**PRS vs. sex interaction p-value**	**PRS vs. sex Beta estimate**
ePRS-IR	N/A	371	0.022	3.6e-2
Random genes ePRS	N/A	794	0.219	2.0e-2
$PRS_{ADHD2010}$	0.001	461	0.470	-1.1e-2
$PRS_{ADHD2017}$	0.0001	409	0.326	-1.6e-2
$PRS_{Tobacco}$	0.001	787	0.551	9.5e-3

Acknowledgements:
This work was funded by the Toxic Stress Research network of the JPB Foundation. The MAVAN Cohort was funded by the Canadian Institutes for Health Research, the Ludmer Family Foundation, the Norlien Foundation (Calgary, Canada), the WOCO Foundation (London, Canada), the Blema & Arnold Steinberg Family Foundation, and the Faculty of Medicine of McGill University. Additional funding was provided by the Jacobs Foundation (Switzerland). LMC was funded by the Brain Canada and Kids Brain Health Network Developmental Neurosciences Research Training Award.

Disclosure statement:
The authors have no conflicts of interest to declare.

References:

1. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. *Diabetes care.* 2001;24(6):1069-1078.
2. Silarova B, Giltay EJ, Van Reedt Dortland A, et al. Metabolic syndrome in patients with bipolar disorder: comparison with major depressive disorder and non-psychiatric controls. *J Psychosom Res.* 2015;78(4):391-398.
3. Accardi G, Caruso C, Colonna-Romano G, Camarda C, Monastero R, Candore G. Can Alzheimer disease be a form of type 3 diabetes? *Rejuvenation research.* 2012;15(2):217-221.
4. Havrankova J, Schmechel D, Roth J, Brownstein M. Identification of insulin in rat brain. *Proc Natl Acad Sci U S A.* 1978;75(11):5737-5741.
5. Figlewicz DP. Insulin, food intake, and reward. *Semin Clin Neuropsychiatry.* 2003;8(2):82-93.
6. Sipols AJ, Bayer J, Bennett R, Figlewicz DP. Intraventricular insulin decreases kappa opioid-mediated sucrose intake in rats. *Peptides.* 2002;23(12):2181-2187.
7. Figlewicz DP. Expression of receptors for insulin and leptin in the ventral tegmental area/substantia nigra (VTA/SN) of the rat: Historical perspective. *Brain research.* 2016;1645:68-70.
8. Baskin DG, Schwartz MW, Sipols AJ, D'Alessio DA, Goldstein BJ, White MF. Insulin receptor substrate-1 (IRS-1) expression in rat brain. *Endocrinology.* 1994;134(4):1952-1955.
9. Bingham EM, Hopkins D, Smith D, et al. The role of insulin in human brain glucose metabolism: an 18fluoro-deoxyglucose positron emission tomography study. *Diabetes.* 2002;51(12):3384-3390.
10. Figlewicz DP, Bennett JL, Aliakbari S, Zavosh A, Sipols AJ. Insulin acts at different CNS sites to decrease acute sucrose intake and sucrose self-administration in rats. *Am J Physiol-Reg I.* 2008;295(2):R388-R394.
11. Schoffelmeer AN, Drukarch B, De Vries TJ, Hogenboom F, Schetters D, Pattij T. Insulin modulates cocaine-sensitive monoamine transporter function and impulsive behavior. *J Neurosci.* 2011;31(4):1284-1291.
12. Kleinridders A, Cai W, Cappellucci L, et al. Insulin resistance in brain alters dopamine turnover and causes behavioral disorders. *Proc Natl Acad Sci U S A.* 2015;112(11):3463-3468.
13. Kemp DE, Ismail-Beigi F, Ganocy SJ, et al. Use of insulin sensitizers for the treatment of major depressive disorder: A pilot study of pioglitazone for major depression accompanied by abdominal obesity. *J Affect Disorders.* 2012;136(3):1164-1173.
14. Stoeckel LE, Arvanitakis Z, Gandy S, et al. Complex mechanisms linking neurocognitive dysfunction to insulin resistance and other metabolic dysfunction. *F1000Res.* 2016;5:353.
15. Mucellini AB, Fonseca N, Manfro GG, Silveira PP. Hippocampal insulin resistance and altered food decision-making as players on obesity risk. *Neurosci Biobehav Rev.* 2017;77:165-176.

16. Consortium CAD, Deloukas P, Kanoni S, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature genetics.* 2013;45(1):25-33.
17. Wray NR, Goddard ME. Multi-locus models of genetic risk of disease. *Genome Med.* 2010;2(2):10.
18. Groen-Blokhuis MM, Middeldorp CM, Kan K-J, et al. Attention-Deficit/Hyperactivity Disorder Polygenic Risk Scores Predict Attention Problems in a Population-Based Sample of Children. *Journal of the American Academy of Child & Adolescent Psychiatry.* 2014;53(10):1123-1129.e1126.
19. O'Donnell KG, H.; Colalillo, S.; Steiner, M.; Atkinson, L.; Moss, E.; Karama, S.; Matthews, S.; Lydon, J.; Silveira, P.P. Wazana, A.; Levitan, R. Sokolowski, M.; Kennedy J.; Fleming, A. Meaney, M. . The Maternal Adversity Vulnerability and Neurodevelopment (MAVAN) Project: Theory and methodology. *Can J Psychiatry.* 2014;59(9):497-508.
20. Bierut LJ, Saccone NL, Rice JP, et al. Defining alcohol-related phenotypes in humans. The Collaborative Study on the Genetics of Alcoholism. *Alcohol Res Health.* 2002;26(3):208-213.
21. Bierut LJ, Madden PA, Breslau N, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet.* 2007;16(1):24-35.
22. Bierut LJ. Genetic variation that contributes to nicotine dependence. *Pharmacogenomics.* 2007;8(8):881-883.
23. Saccone SF, Hinrichs AL, Saccone NL, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet.* 2007;16(1):36-49.
24. Bierut LJ, Strickland JR, Thompson JR, Afful SE, Cottler LB. Drug use and dependence in cocaine dependent subjects, community-based individuals, and their siblings. *Drug Alcohol Depend.* 2008;95(1-2):14-22.
25. Edenberg HJ, Xuei X, Chen HJ, et al. Association of alcohol dehydrogenase genes with alcohol dependence: a comprehensive analysis. *Hum Mol Genet.* 2006;15(9):1539-1549.
26. Edenberg HJ. The collaborative study on the genetics of alcoholism: an update. *Alcohol Res Health.* 2002;26(3):214-218.
27. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics.* 2016;48(10):1279-1283.
28. Mulligan MK, Mozhui K, Prins P, Williams RW. GeneNetwork: A Toolbox for Systems Genetics. *Methods in molecular biology.* 2017;1488:75-120.
29. Miller JA, Ding SL, Sunkin SM, et al. Transcriptional landscape of the prenatal human brain. *Nature.* 2014;508(7495):199-206.
30. Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: Polygenic methods and their application to psychiatric traits. *Journal of child psychology and psychiatry, and allied disciplines.* 2014;55(10):1068-1087.
31. Neale BM, Medland SE, Ripke S, et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry.* 2010;49(9):884-897.
32. Chen LM, Yao N, Garg E, et al. 2017.
33. Demontis D, Walters RK, Martin J, et al. 2017.

34. Tobacco, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature genetics.* 2010;42(5):441-447.
35. Luciana M, Nelson CA. Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery: performance in 4- to 12-year-old children. *Developmental neuropsychology.* 2002;22(3):595-624.
36. Bennett D, Oldham S, Dawson A, Parkes L, Murawski C, Yu Cel M. Systematic Overestimation of Reflection Impulsivity in the Information Sampling Task. *Biol Psychiatry.* 2016.
37. Clark L, Robbins TW. Reply to: Systematic Overestimation of Reflection Impulsivity in the Information Sampling Task. *Biol Psychiatry.* 2016.
38. Pokhvisneva I, Léger É, Meaney MJ, Silveira PP. Systematic Overestimation of Reflection Impulsivity in the Information Sampling Task: Age Dependency in Children. *Biol Psychiat.* 2018;83(2):e33-e34.
39. Bucholz KK, Cadoret R, Cloninger CR, et al. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol.* 1994;55(2):149-158.
40. Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research.* 2013;41(Database issue):D808-815.
41. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research.* 2010;38(Web Server issue):W214-220.
42. Saito R, Smoot ME, Ono K, et al. A travel guide to Cytoscape plugins. *Nature methods.* 2012;9(11):1069-1076.
43. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics.* 2006;38(8):904-909.
44. Fatseas M, Hurmic H, Serre F, et al. Addiction severity pattern associated with adult and childhood Attention Deficit Hyperactivity Disorder (ADHD) in patients with addictions. *Psychiatry Res.* 2016;246:656-662.
45. Vasiliadis HM, Diallo FB, Rochette L, et al. Temporal Trends in the Prevalence and Incidence of Diagnosed ADHD in Children and Young Adults between 1999 and 2012 in Canada: A Data Linkage Study. *Can J Psychiatry.* 2017:706743717714468.
46. Silveira PP, Agranonik M, Faras H, et al. Preliminary evidence for an impulsivity-based thrifty eating phenotype. *Pediatric research.* 2012;71(3):293-298.
47. Rosenstrom T, Torvik FA, Ystrom E, et al. Prediction of alcohol use disorder using personality disorder traits: a twin study. *Addiction.* 2017.
48. Levy S, Katusic SK, Colligan RC, et al. Childhood ADHD and risk for substance dependence in adulthood: a longitudinal, population-based study. *PloS one.* 2014;9(8):e105640.
49. Sihvola E, Rose RJ, Dick DM, et al. Prospective relationships of ADHD symptoms with developing substance use in a population-derived sample. *Psychol Med.* 2011;41(12):2615-2623.

50. O'Dell LE, Nazarian A. Enhanced vulnerability to tobacco use in persons with diabetes: A behavioral and neurobiological framework. *Prog Neuropsychopharmacol Biol Psychiatry*. 2016;65:288-296.
51. Caravaggio F, Borlido C, Hahn M, et al. Reduced insulin sensitivity is related to less endogenous dopamine at D2/3 receptors in the ventral striatum of healthy nonobese humans. *Int J Neuropsychopharmacol.* 2015;18(7):pyv014.
52. Low Income Cut Offs for 2005 and Low Income measures for 2004. Statistics Canada; 2005.

GeneNetwork
University of Tennessee: www.genenetwork.org

Genes co-expressed with the insulin receptor gene in the striatum and PFC in mice[28]

BRAINSPAN
ATLAS OF THE DEVELOPING HUMAN BRAIN

Human homologous genes overexpressed in fetal/childhood periods by comparison to adulthood[29]

Common genes

Selection of all SNPs

NCBI

Selection of SNPs shared with ADHD GWAS[31]

Clumping using the study sample to create final list of SNPs[30]

Calculate ePRS-IR[32]

Study Sample

id	rs1	rs2	rs3	...
id1	0	2	1	
id2	1	1	2	
id3	2	0	1	
...				

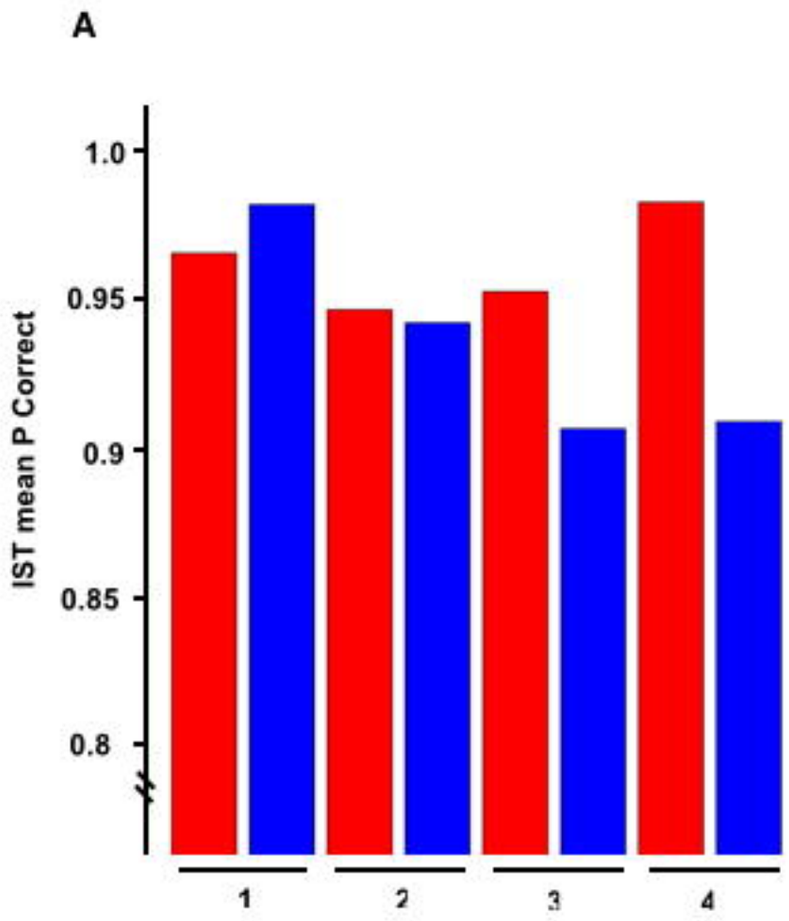
ADHD GWAS[31]

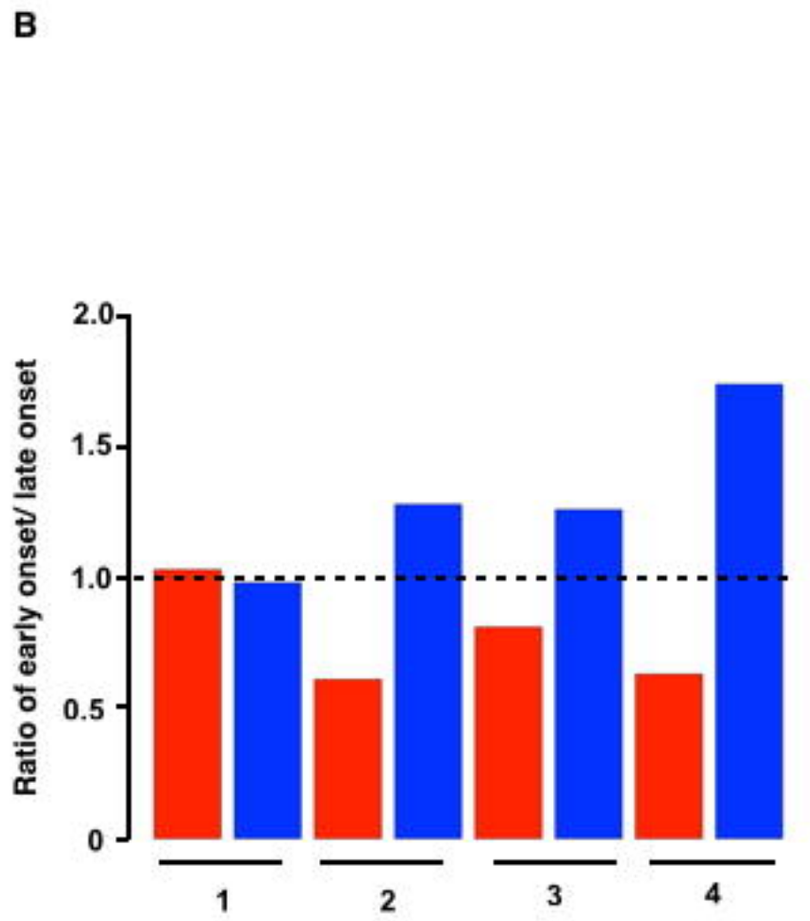
rs1	rs2	rs3	...
ES=x	ES=y	ES=z	

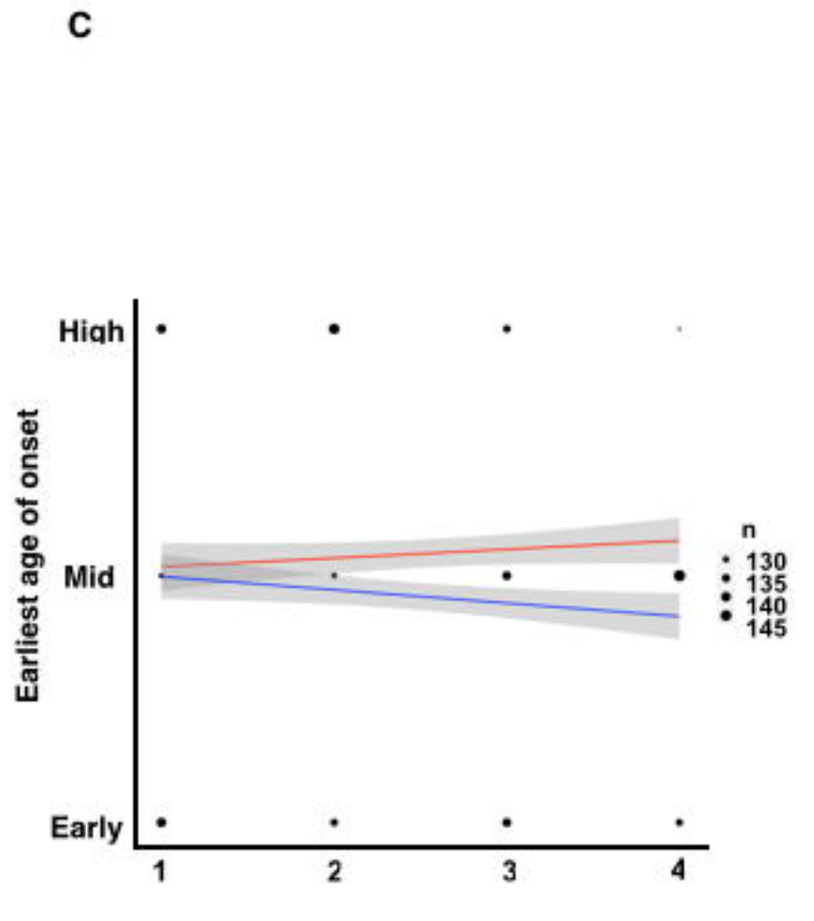
id	rs1	rs2	rs3	...
id1	0x	2y	1z	
id2	1x	1y	2z	
id3	2x	0y	1z	
...				

∑ rs1, rs2, rs3...

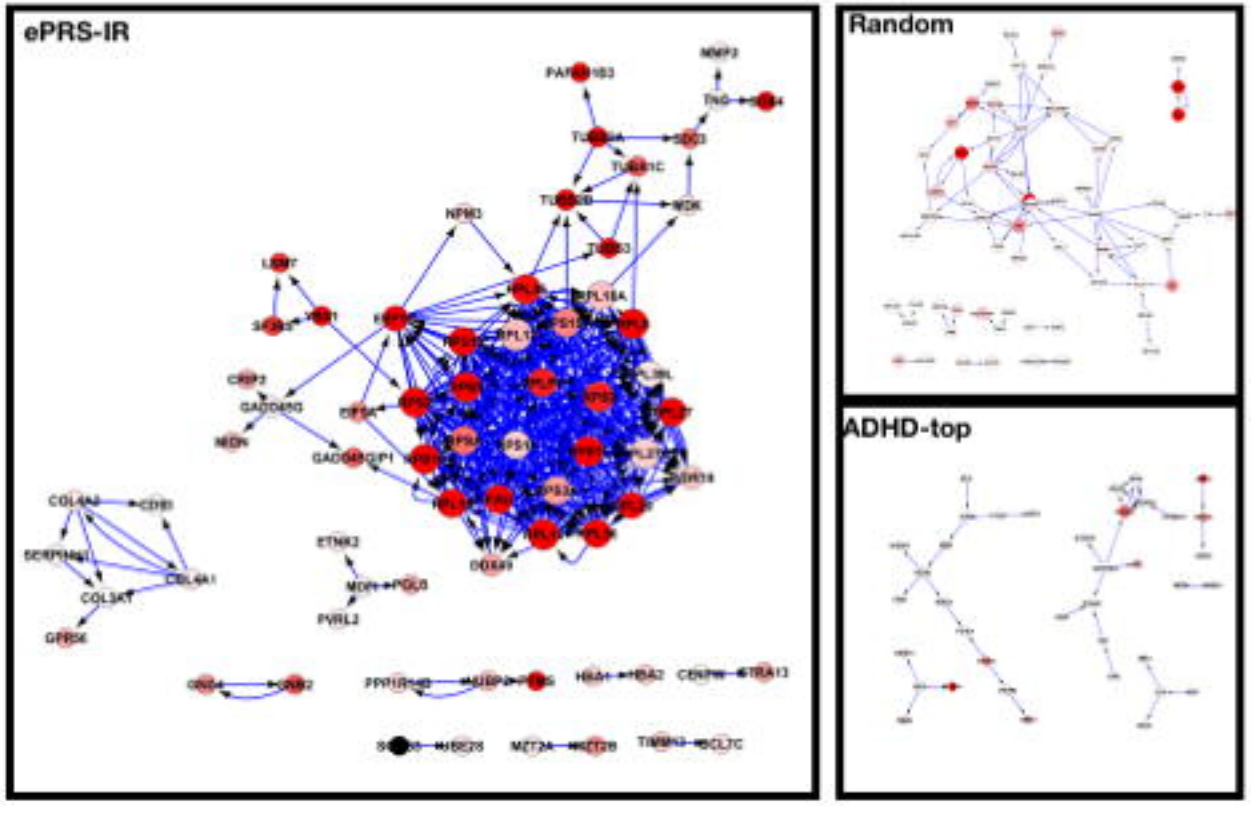
id	ePRS-IR
id1	score 1
id2	score 2
id3	score 3
...	



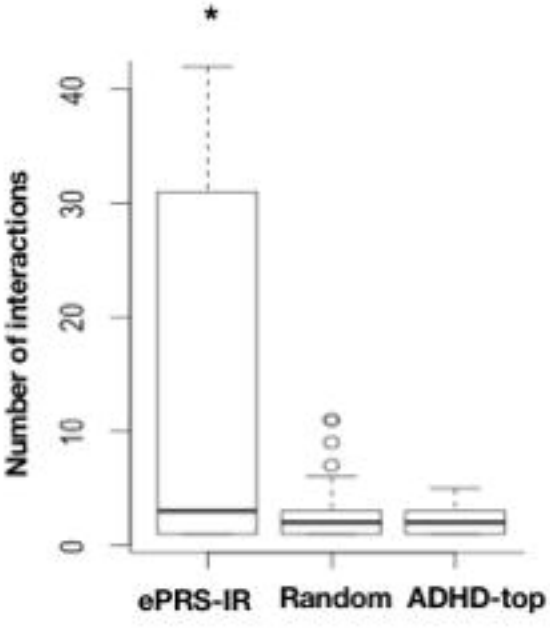




A


ePRS-IR
Random
ADHD-top

B


Number of interactions
0
10
20
30
40
*
ePRS-IR
Random
ADHD-top