



# A biologically-informed polygenic score identifies endophenotypes and clinical conditions associated with the insulin receptor function on specific brain regions

Shantala A. Hari Dass<sup>c</sup>, Kathryn McCracken<sup>f,g</sup>, Irina Pokhvisneva<sup>c</sup>, Lawrence M. Chen<sup>c</sup>, Elika Garg<sup>c</sup>, Thao T.T. Nguyen<sup>c</sup>, Zihan Wang<sup>c</sup>, Barbara Barth<sup>g</sup>, Moein Yaqubi<sup>g</sup>, Lisa M. McEwen<sup>e</sup>, Julie L. MacIsaac<sup>e</sup>, Josie Diorio<sup>c</sup>, Michael S. Kobor<sup>e</sup>, Kieran J. O'Donnell<sup>a,b,c</sup>, Michael J. Meaney<sup>a,b,c,d</sup>, Patricia P. Silveira<sup>a,b,c,\*</sup>

<sup>a</sup> Department of Psychiatry, Faculty of Medicine, McGill University, 6875 Boulevard LaSalle, Verdun, QC H4H 1R3, Canada

<sup>b</sup> Ludmer Centre for Neuroinformatics and Mental Health, Douglas Mental Health University Institute, McGill University, 6875 Boulevard LaSalle, Verdun, QC H4H 1R3, Canada

<sup>c</sup> Sackler Institute for Epigenetics & Psychobiology, McGill University, 6875 Boulevard LaSalle, Verdun, QC H4H 1R3, Canada

<sup>d</sup> Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research (A\*STAR), Brenner Centre for Molecular Medicine, 30 Medical Drive, 117609, Singapore

<sup>e</sup> Centre for Molecular Medicine and Therapeutics, BC Children's Hospital Research Institute, Department of Medical Genetics, The University of British Columbia, 938 West 28th Avenue, Vancouver, BC V5Z 4H4, Canada

<sup>f</sup> John Abbott College, Sainte-Anne-de-Bellevue, 6875 Boulevard LaSalle, Verdun, QC H4H 1R3, Canada

<sup>g</sup> McGill University, 6875 Boulevard LaSalle, Verdun, QC H4H 1R3, Canada

## ARTICLE INFO

### Article history:

Received 16 January 2019

Received in revised form 18 March 2019

Accepted 19 March 2019

Available online 26 March 2019

### Keywords:

Polygenic score

Insulin

Dopamine

Gene networks

Impulsivity

Addiction

Alzheimer's disease

## ABSTRACT

**Background:** Activation of brain insulin receptors modulates reward sensitivity, inhibitory control and memory. Variations in the functioning of this mechanism likely associate with individual differences in the risk for related mental disorders (attention deficit hyperactivity disorder or ADHD, addiction, dementia), in agreement with the high co-morbidity between insulin resistance and psychopathology. These neurobiological mechanisms can be explored using genetic studies. We propose a novel, biologically informed genetic score reflecting the mesocorticolimbic and hippocampal insulin receptor-related gene networks, and investigate if it predicts endophenotypes (impulsivity, cognitive ability) in community samples of children, and psychopathology (addiction, dementia) in adults.

**Methods:** Lists of genes co-expressed with the insulin receptor in the mesocorticolimbic system or hippocampus were created. SNPs from these genes (post-clumping) were compiled in a polygenic score using the association betas described in a conventional GWAS (ADHD in the mesocorticolimbic score and Alzheimer in the hippocampal score). Across multiple samples ( $n = 4502$ ), the biologically informed, mesocorticolimbic or hippocampal specific insulin receptor polygenic scores were calculated, and their ability to predict impulsivity, risk for addiction, cognitive performance and presence of Alzheimer's disease was investigated.

**Findings:** The biologically-informed ePRS-IR score showed better prediction of child impulsivity and cognitive performance, as well as risk for addiction and Alzheimer's disease in comparison to conventional polygenic scores for ADHD, addiction and dementia.

**Interpretation:** This novel, biologically-informed approach enables the use of genomic datasets to probe relevant biological processes involved in neural function and disorders.

**Fund:** Toxic Stress Research network of the JPB Foundation, Jacobs Foundation (Switzerland), Sackler Foundation. © 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The co-morbidity between metabolic and neuropsychiatric disorders is well-established, but poorly understood. The high co-occurrence of several psychiatric conditions (e.g. major depression, bipolar disorder, dementia) with insulin resistance suggests a common underlying mechanism [1]. Metformin, a medication that corrects insulin resistance, has beneficial psychotropic effects in psychiatric conditions [2–4].

\* Corresponding author at: Department of Psychiatry, Faculty of Medicine, McGill University, Douglas Hospital Research Centre, 6875 Boulevard LaSalle, Montreal, QC H4H 1R3, Canada.

E-mail address: [patricia.silveira@mcgill.ca](mailto:patricia.silveira@mcgill.ca) (P.P. Silveira).

## Research in context

### *Evidence before this study*

Employing large sample sizes and specific statistical methods, genome-wide association studies (GWAS) have permitted large scale analyses of common markers for specific disorders. The use of methods of genomic risk profiling is consistent with the idea that the genetic contribution to a certain condition is derived from a combination of small effects from many genetic variants. To take into account the effects of many SNPs, the concept of polygenic risk score (PRS) was introduced. PRS summarizes an individual's genetic risk for a specific condition. A polygenic risk score is calculated for each subject in the target sample as a sum of the risk alleles count, weighted by the effect size described in a discovery GWAS. One problem of the GWAS technology 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.

### *Added value of this study*

We created a novel approach to genomic profiling, informed by biological function, and characterizing gene networks based on the levels of co-expression with a determined gene in a specific tissue. In our study, the relevant biological unit of influence is a gene network, and not a single gene. A gene network involves a number of genes that are co-expressed within a specific tissue or brain region and exert a concerted effect on a target biological process. Hence, following our interest in deciphering the effects of variations in the function of the insulin receptor on the mesocorticolimbic pathway, we created a genetic score based on genes co-expressed with the insulin receptor on the striatum and prefrontal cortex or in the hippocampus, that we called ePRS-IR. The ePRS-IR represents a cohesive gene network and predicts impulsivity endophenotypes and cognitive abilities in children, as well as risk for addiction and dementia in adults.

### *Implications of all the available evidence*

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 neuropsychiatric conditions in community-based samples, highlighting possible targets for drug development.

Insulin receptors (IR) are expressed throughout the brain [5] (striatum, prefrontal cortex [6,7] and hippocampus [8,9]). Insulin is actively transported across the blood-brain barrier, and its action on mesocorticolimbic receptors modulates synaptic plasticity in dopaminergic neurons, affecting dopamine-related behaviors such as response to reward [10], impulsivity [11], mood [12,13], cognition [14] and decision-making [15]. Impulsivity associated with affected dopamine signaling is also a core feature of attention deficit hyperactivity disorder (ADHD), which is co-morbid with obesity [16]. IR location on hippocampal glutamatergic synapses suggests a role of insulin in neurotransmission, synaptic plasticity and modulation of learning and memory, while its inhibition is described in Alzheimer's disease and related animal models [17].

Evidence suggests a developmental trajectory for the establishment of these co-morbidities over the life-course [18–24]. For instance, variations in early growth trajectories during development – essentially an insulin-dependent process [25] – are associated with differences in childhood impulsivity [26]. It is suggested that both psychopathology and metabolic dysfunction are preceded by a long “silent” phase of several decades, that might start very early in life; however, it is likely that subtle neuropsychological dysfunction or small endophenotypic differences may be apparent already in childhood, when neuroadaptations related to individual variation in brain insulin function influence the behavioral phenotype [11,12,27].

The neurobiological processes involved in the co-morbidity between metabolic and neuropsychiatric conditions can be explored using genetic studies. Genome-wide association studies (GWAS) provide the basis for cumulative variants that associate with health outcomes and reflect genetic predispositions to common disorders where individual variants carry small effects. The cumulative polygenic risk of the individual can be used to estimate risk through polygenic risk scores (PRS), that can be calculated as the sum of the count of risk alleles weighted by the effect size of the association between a particular genotype and the outcome from the relevant GWAS study [28,29]. PRSs are therefore derived using existing GWAS data sets for specific disorders.

GWAS and PRS methodologies are focused on statistically significant candidate associations between scattered loci and a certain condition or trait, not accounting for the fact that genes operate in networks, and code for precise biological functions in specific tissues. Since the PRSs are conventionally based on statistical thresholds, while they might associate with the final disease state, they are not informative of the various clinical manifestations or endophenotypes that may precede the disease development. This makes the biological underpinnings of the PRS complicated to disentangle, restricting their clinical scope/implications. A score that captures the genetic background associated with an endophenotype – potentially an early onset manifestation – would be able to mold subsequent preventive and therapeutic endeavors.

With the reduction in cost of genotyping strategies we are poised to expand the implications of a multitude of GWA studies into healthcare. Hence, we aimed to address this need by developing a novel genomics approach that provides a biologically-informed genetic score, based on genes co-expressed with the IR in specific brain regions. Our method is based on the assumption that the relevant biological unit of influence during development is a gene network, and not isolated statistically significant SNPs. A gene network involves a number of genes that are co-expressed within a specific tissue or brain region, and exert a concerted effect on a target biological process. Our new methodology enables a hypothesis-driven, neurobiological analysis of phenotypes linked to brain disorders across developmental stages.

## 2. Materials and methods

### 2.1. Samples

**Main Cohort:** We used data from the prospective Maternal Adversity, Vulnerability and Neurodevelopment (MAVAN) birth cohort [30] that followed children at different time points in the first years of life in Montreal (Quebec) and Hamilton (Ontario), Canada. The sample size included in these analyses were 218 children at 48 months of age, and 204 children at 72 months of age. Approval for the MAVAN project was obtained from obstetricians performing deliveries at 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.

**Other Cohorts:** (A) Study of Addiction, Genetics and Environment (SAGE) repository [31–37], acquired from dbGaP (<https://www.ncbi.nlm.nih.gov/gap>, Accession number: phs000092.v1.p). The SAGE

dataset was compiled from three studies: the Collaborative Study on the Genetics of Alcoholism (COGA); the Family Study of Cocaine Dependence (FSCD); and the Collaborative Genetic Study of Nicotine Dependence (COGEND). The SAGE dataset contains genotyping and clinical phenotypes related to substance dependence for adult subjects. The sample size for these analysis was 2719 individuals. 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. (B) GEN-ADA [38,39] is a multi-site study conducted at GlaxoSmithKline Inc. and nine medical centres. The study was designed to capture genetic information from Alzheimer's disease patients and matched controls. The sample size for this analysis was 1565 individuals. We received access to the GenADA 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.

## 2.2. Behavioral Phenotyping

**Reflection impulsivity in children:** The Information Sampling Task (IST) from the CANTAB battery was designed to measure reflection impulsivity and decision making [40], and was applied at 72 months in MAVAN. Despite the fact that we have applied other tasks in MAVAN, we specifically analyzed the Information Sampling Task with regards to the ePRS-IR, considering the theoretical background that insulin modulates dopaminergic neurotransmission and, hence, impulsivity and inhibitory control. 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  $5 \times 5$  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 colour 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. 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 colour 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 [41,42], therefore in this study we calculated and used the new mean P Correct [43]. Moreover, as we have reported differences in impulsivity levels between boys and girls in MAVAN [44], and impulsivity is highly influenced by sex [45,46], we analyzed the effect of sex as well as sex by genetic score interactions on IST.

**Number Knowledge Task (NK):** The Number Knowledge task is part of the School Readiness Battery, applied at 48 months in MAVAN, assessing school readiness which may be defined as the minimum developmental level allowing the child to respond adequately to school demands [47]. This task is a well validated diagnostic screening test of school readiness [48].

**Addiction risk:** The clinical assessment of substance dependency was based on a Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA II) [49] and adapted versions of the SSAGAII, which assesses the physical, psychological and social manifestations of substance dependence. For each type of substance (alcohol, cocaine, marijuana, nicotine, opiates, other drugs than marijuana/cocaine/opiates), dependence was characterized as a maladaptive pattern of substance use, leading to clinically significant impairment or distress, as manifested

by three or more of the following occurring at any time in the same 12-month period: (a) tolerance, (b) withdrawal, (c) the substance is often taken in larger amounts or over a longer period than intended, (d) there is a persistent desire or unsuccessful efforts to cut down or control substance use, (e) a great deal of time is spent in activities necessary to obtain the substance, use the substance, or recover from its effects, (f) important social, occupational, or recreational activities are given up or reduced because of substance use, g) the substance use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance. The number of co-morbidities was calculated as the number of reported dependencies according to these criteria. Presence of alcohol dependence was also used as an outcome.

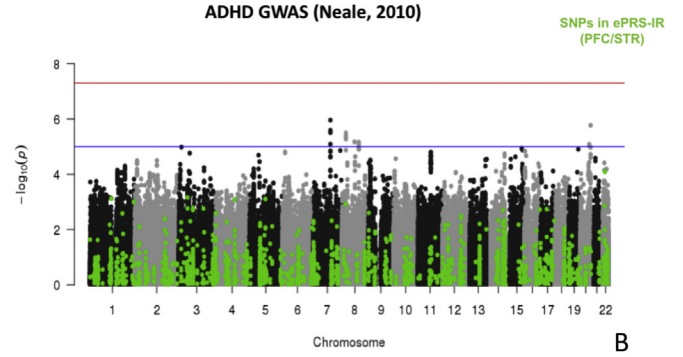
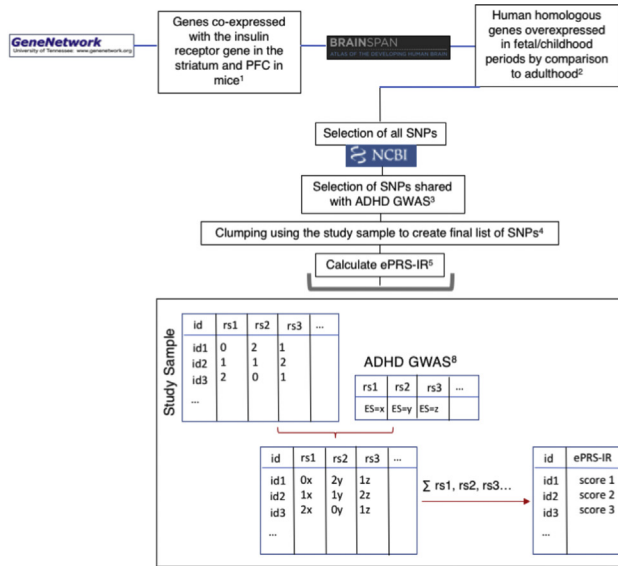
**Alzheimer's diagnosis:** The Alzheimer's disease status was diagnosed using the Diagnostic and Statistical manual of Mental Disorders Fourth Edition [50].

## 2.3. Genotyping

In MAVAN, we genotyped 242,211 autosomal SNPs using genome-wide platforms (PsychArray/PsychChip, Illumina) according to the manufacturer's guidelines with 200 ng 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 [51] resulting in 20,790,893 SNPs with an info score > 0.80 and posterior genotype probabilities >0.90. In SAGE, we used the imputed genotypes provided in the data repository. The imputation was performed using BEAGLE. Post imputation, SNPs were filtered to only include those with an  $r^2 > 0.3$  (total = 4,566,998). In GenADA, genotyping data was provided for 402,708 SNPs, 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%). Afterwards, we performed imputation using the Sanger Imputation Service [51], resulting in 7,267,018 SNPs with an info score > 0.80 and posterior genotype probabilities >0.90.

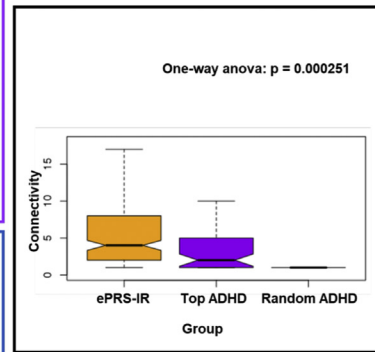
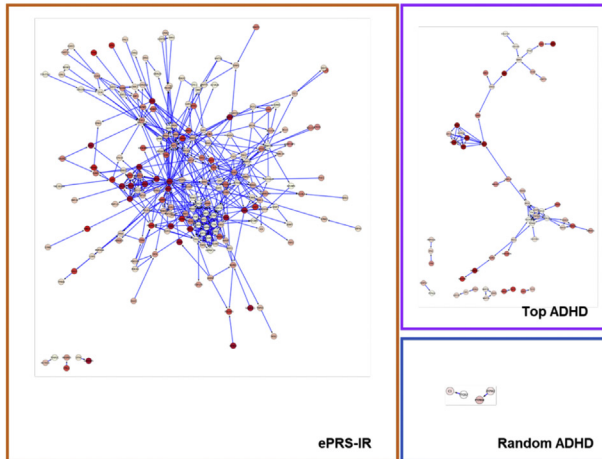
## 2.4. ePRS-IR calculation

The steps for the development of the IR expression-based polygenic risk score (ePRS-IR) are depicted in Fig. 1 (Fig. 1A). 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>) [52], (2) BrainSpan (<http://brainspan.org>) [53], (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), find their homologous human genes (BrainSpan), and identify SNPs for these genes in humans (NCBI Variation Viewer). The PRS was constructed as follows: (1) we used GeneNetwork to generate co-expression matrices with IR in the (i) ventral striatum, (ii) PFC in mice (absolute value of the co-expression correlation  $r \geq |0.5|$ ). The two matrices generated two lists of genes co-expressed with IR in these brain regions. These two gene lists were aggregated into a single list of genes; (2) we then used BrainSpan to identify human homologous transcripts from this list; (3) we selected autosomal transcripts differentially expressed in these brain regions at  $\geq 1.5$  fold during child and fetal development as compared to adulthood [53]; the final list included 281 genes; (4) 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 = 11,068); 5) we 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



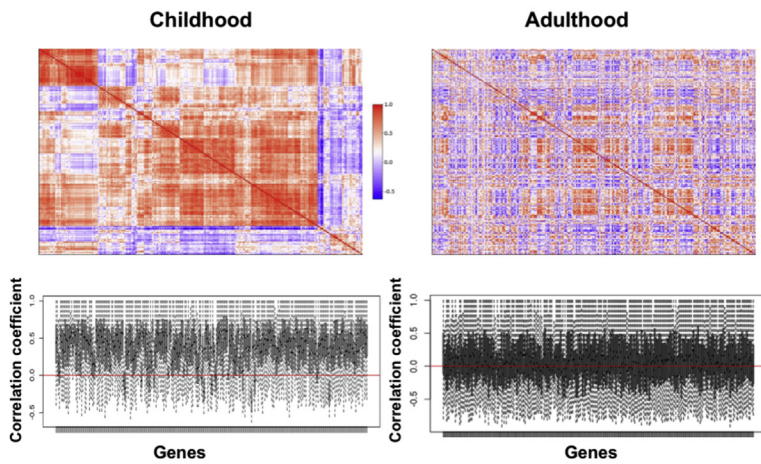
A

B

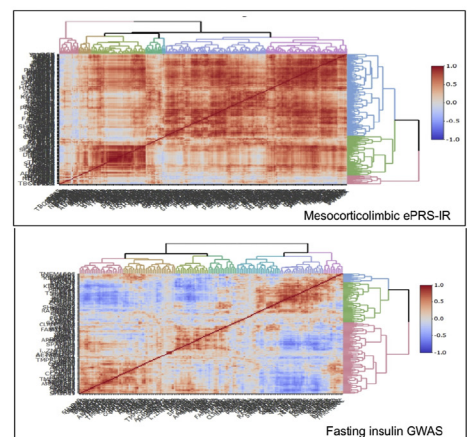


C

D



E



F

correlated SNPs ( $r^2 > 0.2$ ) across 500 kb regions [28], resulting in 1897 independent functional SNPs based on the children's genotype data from MAVAN; (6) 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 [54]. We compared the distributions of the score in the two cohorts (MAVAN and SAGE) using Kolmogorov-Smirnov test and there were no significant differences between the scores ( $p = .996$ ). All SNPs were subjected to linkage disequilibrium clumping ( $r^2 > 0.2$  across 500 kb) so only independent SNPs that are most associated to ADHD, based on the association  $p$ -values in the ADHD GWAS, comprised the ePRS-IR (Table S1).

To further validate the brain-specificity of the biologically-informed polygenic score, we followed the same approach depicted in Fig. 1A and created the hippocampal ePRS-IR by changing the brain region of interest in the process, and compiling the SNPs using the association betas described in the Alzheimer's GWAS [55] (Table S2). The final list included 544 genes (530 excluding genes on X and Y chromosomes) and 363,412 SNPs; this list was subjected to linkage disequilibrium clumping, resulting in 6594 independent functional SNPs based on the children's genotype data from MAVAN. We compared the distributions of the score in the two cohorts (MAVAN and GenADA) using Kolmogorov-Smirnov test and there were no significant differences between the scores ( $p = .668$ ). 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 Alzheimer's disease [55].

## 2.5. Gene network analysis

We extracted a list of the 281 genes from the SNPs with the lowest  $p$ -values based on the post clumped results of the ADHD GWAS [54]. RNA-sequencing data was downloaded from BrainSpan [53], including samples from 8 postconceptional weeks to 11 years old within prefrontal cortex (dorsolateral, ventrolateral, anterior cingulate cortex and orbitofrontal cortex) and striatum, for three gene lists: (a) ePRS-IR; (b) Random gene list (as detailed below in 2.7) and (c) ADHD top 281 genes. A mean expression value was computed across the mentioned brain regions for each subject from the BrainSpan dataset. The protein-protein interaction data were retrieved from STRING (<https://string-db.org/>) [56] and GeneMANIA (<https://genemania.org/>) [57] databases and the protein-protein interaction networks were constructed and visualized in the Cytoscape software [58]. One-way ANOVA was used to compare these values across the three gene lists. The gene

network for the hippocampal ePRS-IR was built using the same methodology. Gene set enrichment and transcription factor analysis were performed using MetaCore™ (Clarivate Analytics).

## 2.6. Coexpression analysis

We used publicly available gene expression data from BrainSpan (<http://www.brainspan.org>) [53] to analyze the correlation between the expression levels of the genes included in the ePRS-IR in the human mesocorticolimbic system and hippocampus, comparing children (1 to 12 years;  $n = 26$ ) to adults (20 to 40 years;  $n = 22$ ), or including all subjects in the same matrix to compare with the same number of genes (281) associated with the most significant SNPs from the fasting insulin GWAS [70]. The analyses were carried out in R using the heatmaply package.

## 2.7. Other genetic scores – Analysis of validation

We generated other polygenic scores using our accelerated pipeline (<https://github.com/MeaneyLab/PRSoS>) [59], for each subject. The ePRS-IR was compared to many other polygenic risk scores for validation:

- 1) Traditional polygenic risk score for ADHD based on the GWAS published in 2010 [54]; the most significant 1897 SNPs from this GWAS were included in this score.
- 2) Traditional polygenic risk score for ADHD based on the GWAS published in 2019 [60]; the most significant 1897 SNPs from this GWAS were included in this score.
- 3) Traditional polygenic risk score for addiction based on the GWAS for tobacco smoking [61]; the most significant 1897 SNPs from this GWAS were included in this score.
- 4) A random list of 1854 SNPs from the clumped version of the ADHD GWAS [54] was included in the 'random' score. We created 30 random scores and repeated linear regression analyses for each of the random PRSs. The results of the regression analyses involving different random scores were pooled together to obtain an overall estimate and standard error of the effect of the random PRS interaction with sex on the IST performance at 72 months (MAVAN) or addiction risk (SAGE), using Rubin's set of rules implemented in a function pool [62] of the R package mice [63].

**Fig. 1.** The mesocorticolimbic ePRS-IR: (A) 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. (i) 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  $r \geq 0.5$ ); (ii) BrainSpan was then used to identify consensus human homologous transcripts from this list; (iii) BrainSpan was also used for selecting genes differentially expressed at  $\geq 1.5$  fold during child and fetal development as compared to adulthood within the same brain areas. The final list included 281 genes; (iv) 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 = 11,068) and subjected this list of SNPs to linkage disequilibrium clumping, to inform removal of highly correlated SNPs ( $r^2 > 0.2$ ) across 500 kb regions, resulting in 1897 independent functional SNPs based on the children's genotype data from MAVAN (Study Sample ids); (v) 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. References: 1. Mulligan, et al. 2017, 2. Miller et al. 2014, 3. Neale et al. 2010, 4. Wray et al. 2014, 5. Chen et al. 2018. (B) Single nucleotide polymorphisms (SNPs) included in the mesocorticolimbic ePRS-IR, shown in green, in relation to the Manhattan plot of the ADHD GWAS. The biologically-informed method for selecting SNPs is designed to capture signals associated with a functional gene network, and hence it can be expected that none of the individual SNPs included in the score are statistically significantly associated with the disease (in this case, ADHD). (C) Gene Network analysis of the mesocorticolimbic ePRS-IR 281 genes, ADHD 2010 (genes associated with most significant or "top" 1897 SNPs) and genes associated with a random selection of 1854 SNPs (bottom small panel). The picture demonstrates that the ePRS-IR represents a network with significantly higher connectivity than a list of genes associated with a random list of SNPs, and also in comparison to the top genes from the ADHD GWAS. (D) Comparison of the number of connections between the genes in each network (mesocorticolimbic ePRS-IR, random list and top genes in the ADHD GWAS). Genes included in the mesocorticolimbic ePRS-IR have significantly more interactions than the random list, and the top genes from the ADHD GWAS (One-Way ANOVA,  $p < .05$ ). This suggests a more cohesive gene network of biological relevance. (E) Coexpression of the genes included in the mesocorticolimbic ePRS-IR, in childhood and adulthood, combining gene expression data from PFC and striatal regions. Top panels: The heatmap of the genes' co-expression in childhood (left) shows several clusters of highly co-expressed genes. Although the clusters are not maintained in the retained order heatmap for adulthood (right), some other clusters of co-expressed genes are observed. Bottom panels: values of the gene expression correlation coefficients in childhood (left) and adulthood (right). Each vertical line represents correlation with a unique gene (the same genes are depicted in the lines vs. columns). The genes included are the ones that generate the ePRS-IR (see supplementary tables). The red line is drawn at a correlation score of 0. Especially in childhood, most genes included in the score have highly correlated gene expression values. Data for this analysis were extracted from BrainSpan. (F) Co-expression of the genes included in the mesocorticolimbic ePRS-IR (top panel) and a comparable number of genes associated with the most significant SNPs from the fasting insulin GWAS (bottom panel), from birth to 11 years of age in the PFC and striatal regions. The heatmaps demonstrate that while the mesocorticolimbic ePRS-IR includes genes that are highly co-expressed in striatum and PFC, genes from the fasting insulin GWAS are less consistently co-expressed in these brain regions. This suggests that the ePRS methodology captures gene networks that are cohesive in specific brain regions, which may not be represented when analyzing a GWAS based on peripheral insulin levels.

The traditional PRSs are cumulative summary scores computed as the sum of the allele count weighted by the effect size across SNPs at certain  $p$ -value thresholds ( $P_T$ ) based on the relevant GWAS [28]. In all the genetic scores above, the number of SNPs is comparable to the number of SNPs included in the mesocorticolimbic ePRS-IR.

## 2.8. Statistical analysis

### 2.8.1. Descriptive statistics for all cohorts

The statistical analysis of the baseline characteristics was performed using Spearman's correlations, One-Way ANOVA and Student's  $t$ -Test. We examined population structure within our data using a principal component analysis [64,65]. First, we pruned our dataset to common variants ( $MAF > 0.05$ ) that were not in linkage disequilibrium ( $r^2 < 0.20$ ) with a sliding window (50 kilobases) approach that examined linkage disequilibrium in increments of 5 SNPs using PLINK 1.9 [66]. We performed a principal component analysis using SMARTPCA on this pruned dataset and generated a scree plot (see Supplementary Fig. S1). 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 analyses.

### 2.8.2. Mesocorticolimbic ePRS-IR

A linear regression analysis was performed to explore if the sex by ePRS-IR interaction was associated with the IST outcomes, adjusting for population structure and birth weight in MAVAN [26]. Similarly, in SAGE, a linear regression analysis was performed to explore if the sex by ePRS interaction was associated with the number of addiction comorbidities. Logistic regression was applied to explore the association with the presence of alcohol dependence. In all cases we adjusted for the population structure and study source. Simple slopes analysis was performed to test the association between the ePRS-IR and the outcome in males and females when appropriate.

### 2.8.3. Hippocampal ePRS-IR

In MAVAN, we explored the association between the hippocampal ePRS-IR and number knowledge task performance at 48 months of age using linear regression analysis, adjusting by population stratification and sex. In ADA, we used logistic regression to investigate the association between the hippocampal ePRS-IR and the presence of Alzheimer's disease.

As a validation of the associations described, for the main analysis in MAVAN, SAGE and ADA, we applied a permutation test to assess the significance of the associations. After permuting the outcome 5000 times, we computed the sampling distribution of the test statistics of interest (mesocorticolimbic ePRS-IR \*sex interaction coefficient, or hippocampal ePRS-IR main effect) under the null hypothesis that there is no significant association, and calculated associated  $p$ -values as a proportion of number of times the test statistic was bigger than the real observed one.

Data were analyzed using the Statistical Package for the Social Sciences (SPSS) version 20.0 software (SPSS Inc., Chicago, IL, USA) and R [67]. We ran all the analysis in the full datasets and repeated it excluding influential observations. As the results were similar, we chose to present the more conservative statistical approach excluding influential observations. Significance levels for all measures were set at  $\alpha = 0.05$ .

## 3. Results

### 3.1. Establishment of the mesocorticolimbic biologically-informed genetic score based on genes co-expressed with the insulin receptor (ePRS-IR)

This biologically-informed method for selecting SNPs is designed to capture risk associated with a functional unit of a gene network, and hence it can be expected that none of the individual SNPs included in the score are statistically significantly associated with ADHD diagnosis in the GWAS (Fig. 1B), but instead, the cumulative ePRS-IR score

associates with the relevant endophenotypes. A gene network analysis was performed comparing the list of the 281 genes included in the ePRS-IR with (a) the genes associated with a random selection of 1854 SNPs from the clumped version of the ADHD GWAS [54] and (b) the genes associated with the 1897 most significant SNPs from the ADHD GWAS [54] (Fig. 1C). The analysis shows that the number of interactions between the genes that comprise the ePRS-IR is significantly higher than the control random list, or the genes from the ADHD GWAS (Fig. 1D), suggesting that the biologically-informed score represents a much more cohesive gene network. The reasons why the 'random' gene network seems small are: (a) not all random SNPs were located inside the genes and (b) most of the resulting genes had no connection at all with the other genes from the gene set, and hence are not depicted in the figure.

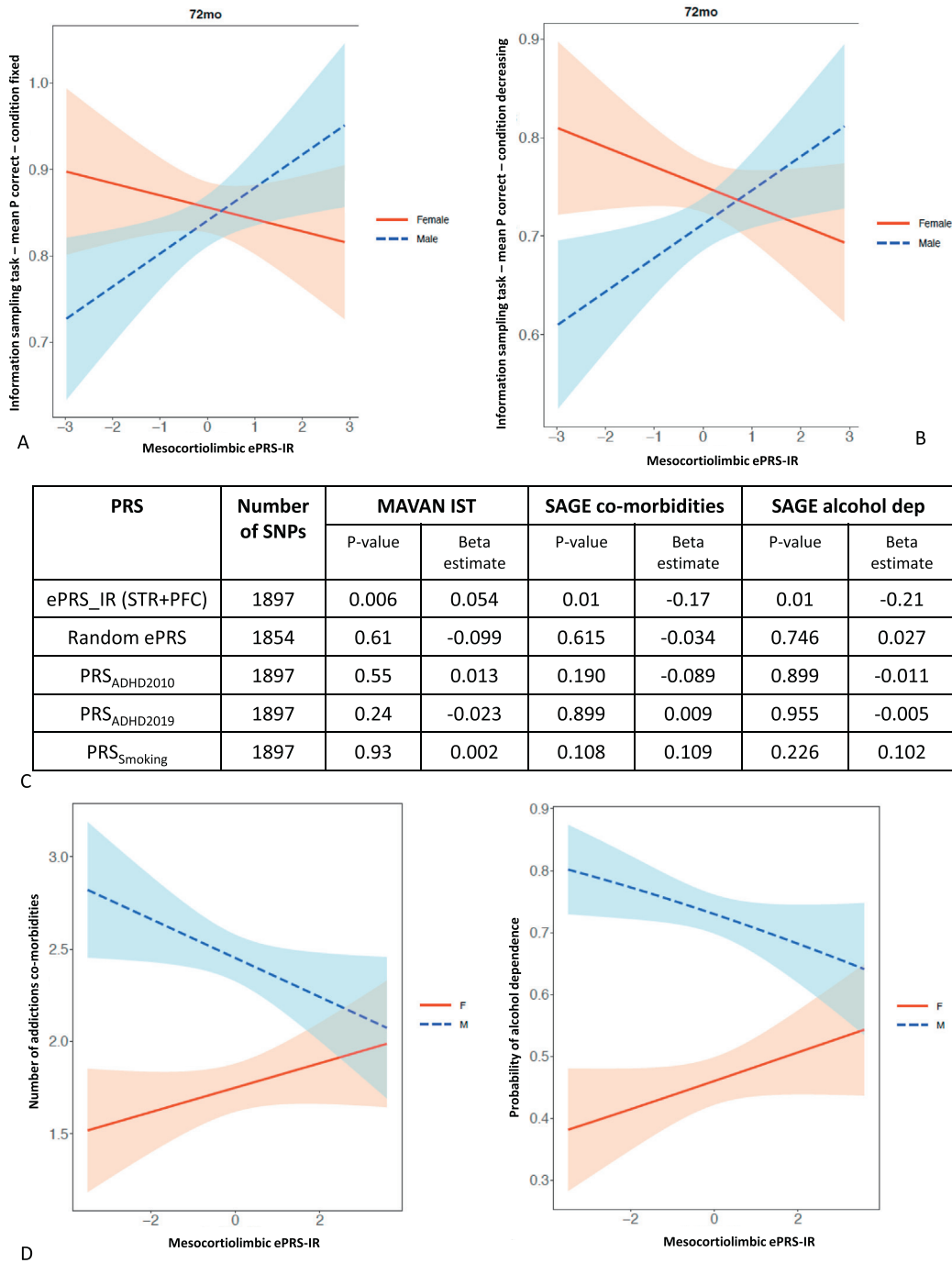
We examined the co-expression patterns of the genes included in the mesocorticolimbic ePRS-IR over the life-course using data from BrainSpan [53]. Several genes are highly co-expressed in childhood, confirming the co-expression matrix from mice that was used to select the genes included in the score. Although co-expression continues into adulthood, this is re-arranged in other, more discrete clusters of co-expression (Fig. 1E).

The brain specificity of our score is demonstrated when comparing the co-expression patterns of the genes included in our score and an equivalent number of genes associated with the most significant SNPs from the fasting insulin GWAS [70] in the PFC and striatal regions (Fig. 1F). While the mesocorticolimbic ePRS-IR has highly co-expressed genes in these brain areas, genes from the fasting insulin GWAS are less consistently co-expressed in these regions. This confirms the ability of the ePRS methodology to generate a genetic score that denotes tissue-specific gene networks, and highlights the meaningful difference between the information contained in the ePRS-IR score and peripheral levels of metabolic hormones such as insulin.

### 3.2. Endophenotypic differences predicted by the mesocorticolimbic ePRS-IR

We then analyzed whether the ePRS-IR would predict impulsivity, a highly sex-specific trait, in 6-year old children [30] tested using a computer-based, Information Sampling Task (IST). There was a significant interaction effect between ePRS-IR and sex on the Information Sampling Task (IST) (fixed condition  $\hat{\beta} = 0.052, p = .016$ ; decreasing condition  $\hat{\beta} = 0.054, p = .006$ ) applied at 72 months (Fig. 2A and B). While a simple slopes analysis showed no relationship between the ePRS-IR score and mean P Correct values in girls (fixed condition  $\hat{\beta} = -0.014, p = .36$ ; decreasing condition  $\hat{\beta} = -0.02, p = .15$ ), a lower ePRS-IR was significantly related to lower mean P Correct (less certainty when coming up to a decision or higher reflection impulsivity) in boys (fixed condition  $\hat{\beta} = 0.038, p = .01$ ; decreasing condition  $\hat{\beta} = 0.034, p = .01$ ). Results in MAVAN remain significant after adjustment for multiple comparisons. A large scale permutation analysis was significant for the interaction between the ePRS-IR and sex in both fixed ( $p = .008$ ) and decreasing conditions ( $p = .004$ ).

We next examined whether a "random" PRS or other conventional PRSs (for ADHD [54,60] or addiction [61]) of comparable size in terms of SNP number could also predict impulsivity in these children. We created 30 random ePRS and repeated the linear regression analyses for each of the random ePRSs. The pooled results showed a non-significant interaction term (fixed condition:  $\hat{\beta} = 0.009, p = .75$ ; decreasing condition:  $\hat{\beta} = 0.004, p = .90$ ). None of the control scores was associated with the interaction effect (sex vs. ePRS) on reflection impulsivity in this sample (Fig. 2C). Plots describing the main effects both of the mesocorticolimbic ePRS-IR and of sex on IST fixed and decreasing conditions can be found on the Supplementary material (Fig. S2).



**Fig. 2.** Phenotypic differences predicted by the mesocorticolimbic ePRS-IR: (A) and (B) Performance in the Information Sampling Task (IST, CANTAB) at 72 months according to sex and ePRS-IR in fixed (A) and decreasing (B) conditions. There is a significant interaction between the genetic score and sex on IST outcome, in which boys with lower ePRS-IR sample less information before taking the decision, being more impulsive than boys with higher ePRS-IR. Boys are depicted in blue and girls in red.  $N = 204$ . (C) Validation of the findings by using different polygenic risk scores to investigate the interaction with sex on the Information Sampling Task in children and in SAGE outcomes. PRS with a random SNP selection, a classic PRS from two different GWASes for ADHD or a PRS for addiction (i.e. smoking) were unable to predict the behavioral phenotypes, but the mesocorticolimbic ePRS-IR has an interaction effect with sex on IST outcomes in children and with addiction risk in adults. (D) Number of addiction co-morbidities according to sex and mesocorticolimbic ePRS-IR in the SAGE cohort. There is a significant interaction between the genetic score and sex on the number of co-morbidities, in which ePRS-IR has an effect on the number of co-morbidities in men but not women. (E) Probability of alcohol abuse in the SAGE cohort. A significant interaction between the genetic score and sex was found on the probability of alcohol dependence, in which ePRS-IR has an effect on the probability of alcohol dependence in men but not women ( $N = 2719$ ).

We then hypothesized that a lower ePRS-IR, associated with childhood impulsivity in boys as shown above, would predict risk for addiction in men, considering the clinical overlap between impulsive phenotypes and risk for addiction [72]. We therefore investigated if the ePRS-IR was associated with the risk for addiction phenotypes using data from the Study of Addiction: Genetics and Environment (SAGE) repository [31]. The analysis revealed a significant interaction

between the ePRS-IR score and sex for the number of addiction comorbidities (interaction effect;  $\hat{\beta} = -0.17$ ,  $p = .01$ ) and probability of alcohol dependence (interaction effect;  $\hat{\beta} = -0.21$ ,  $p = .01$ ). Further simple slopes analysis confirmed that a lower ePRS-IR was significantly associated with number of addiction comorbidities only in males (males,  $\hat{\beta} = -0.11$ ,  $p = .04$ ; females,  $\hat{\beta} = 0.07$ ,  $p = .14$ ); the simple slopes analysis

for alcohol dependence returned a marginally significant effect in males ( $\hat{\beta} = -0.12, p = .06, OR = 3.38$ ) with no effect in females ( $\hat{\beta} = 0.09, p = .10$ ) (Fig. 2D and E). A large scale permutation analysis of the interaction between ePRS-IR and sex was significant for alcohol dependency ( $p = .005$ ) and co-morbidities ( $p = .008$ ). The interaction results in SAGE remain significant after adjustment for multiple comparisons.

We also examined in SAGE whether a “random” PRS or other conventional PRSs (for ADHD [54,60] or addiction [61]) interact with sex predicting addiction co-morbidities or alcohol dependence in adults. The pooled results from the random PRS showed a non-significant interaction term (co-morbidities:  $\hat{\beta} = -0.016, p = .87$ ; alcohol dependence:  $\hat{\beta} = -0.005, p = .96$ ). None of the control scores had any interaction effect (sex vs. ePRS) on co-morbidities or alcohol intake in this sample (Fig. 2C).

### 3.3. Comparison between mesocorticolimbic ePRS-IR using two versions of the ADHD GWAS – 2010 and 2019

Considering the theoretical premise of the ePRS-IR (biological polygenic score based on gene co-expression), we hypothesized that the choice of the GWAS for providing effect sizes for weighing the SNPs would not have a major influence on the prediction ability of the expression based polygenic score. We then created simultaneously two mesocorticolimbic ePRS-IR scores at the point of the SNP selection, weighing one using the 2010 ADHD GWAS [54] and the second one using the recently published 2019 ADHD GWAS [60]. Confirming our hypothesis, both scores had a significant interaction effect with sex on impulsivity phenotype measured by the IST task in MAVAN (2010:  $\hat{\beta} = 0.02, p = .05$ ; 2019:  $\hat{\beta} = 0.02, p = .05$ ) (Fig. S3). This happens despite the large differences in the 2 studies' sample sizes (2010: 5415; 2019: 55,374), number of significant hits (2010: none; 2019: 12) and variance explained (2010 below 0.5; 2019: 5%). We highlight the fact that our approach is focused on the biological processes underlying endophenotypic differences in healthy/community samples, rather than in SNPs statistically associated with a disorder in clinical samples as identified by these two GWAS. As our score represents biological variability rather than presence or absence of disease, it is extremely important that our score predicts variation in endophenotypes rather than simply including statistically significant

SNPs (please refer to Fig. 1B). Several studies described associations between polygenic scores based on the 2010 ADHD GWAS and endophenotypes associated with attention-impulsivity problems in community samples (rather than ADHD cases vs. controls, similar to our main cohort MAVAN [73,74]). On the other hand, the new ADHD GWAS (2019), despite being adequately powered for detecting disease state, failed to demonstrate any associations with attention problems (= variation in endophenotype, assessed by the Child Behavior Checklist at 48 and 60 months) in our community-based sample MAVAN (data not shown). Therefore, in this study, we opted for our main analysis to be focused on the ePRS-IR weighted using the effect sizes described in the 2010 ADHD GWAS.

Table 1 describes the baseline characteristics of the samples (MAVAN and SAGE) in relation to the mesocorticolimbic ePRS-IR.

### 3.4. The Hippocampal ePRS-IR

To explore the ability of ePRS-IR to capture specific biological roles associated with IR in particular brain regions, we created the hippocampal ePRS-IR, considering that IRs are expressed in the hippocampus and are associated with cognitive processes. Similarly to the mesocorticolimbic ePRS-IR, the selected SNPs are not statistically significantly associated with Alzheimer's in the GWAS [55](Fig. 3A). The gene network analysis demonstrates that the hippocampal ePRS-IR also represents a highly interconnected network (Fig. 3B). Analysis of the co-expression pattern of the genes included in the hippocampal ePRS-IR revealed that the high co-expression clusters from childhood are not maintained in adult life (Fig. 3C). Informed by the role of the IR in the hippocampus on cognition, we explored this endophenotype in children by evaluating cognitive performance on the Number Knowledge Task applied at 48 months [30]. There was a main effect of the hippocampal ePRS-IR in predicting task performance (48 months:  $\hat{\beta} = -0.384, p = .03$ , permutation analysis  $p = .018$ ), in which children with higher ePRS-IR scores have lower performance on the cognitive test (Fig. 3D). We found no significant interaction effect between the hippocampal ePRS-IR and sex on the task at this age. The hippocampal ePRS-IR was also able to distinguish the presence of Alzheimer's disease diagnosis in the GenADA study [38] ( $\hat{\beta} = 0.32, p = 6.82e-09$ , permutation analysis  $p < 1e-6$ ), in which there are more cases in individuals with higher ePRS-IR scores (Fig. 3E). No significant interaction between ePRS and sex was found in the adult cohort.

Table 1 describes the baseline characteristics of the samples (MAVAN and GenADA) in relation to the hippocampal ePRS-IR.

### 3.5. Comparison Between Mesocorticolimbic and Hippocampal ePRS-IR Scores

To examine the isolated contribution of unique SNPs (as opposed to the global ePRS scores) to the above described effects, we performed a SNP by sex association analysis for the mesocorticolimbic ePRS-IR (investigating the interaction with sex) using linear regression models (each SNP included in the ePRS-IR by sex interaction models). No unique SNP included in the mesocorticolimbic ePRS-IR reached genome-wide significance level on the SNP by sex interaction on impulsivity in children, or risk for addiction in adults.

We also performed an association analysis for each SNP included in the hippocampal ePRS-IR (investigating the main effects of each SNP from the hippocampal ePRS-IR on the outcomes) using linear regressions. Similarly, no isolated SNP included in the hippocampal ePRS-IR was associated with poorer cognitive performance in children, or Alzheimer's disease at the genome-wide significance level, confirming that variations in the biological function of the gene networks represented in the ePRS-IR scores predict their respective outcomes as a global score, with small contributions from all included SNPs, but not by any isolated mutation (Fig. 4A).

**Table 1**

Baseline characteristics of the study samples in relation to the genetic scores.

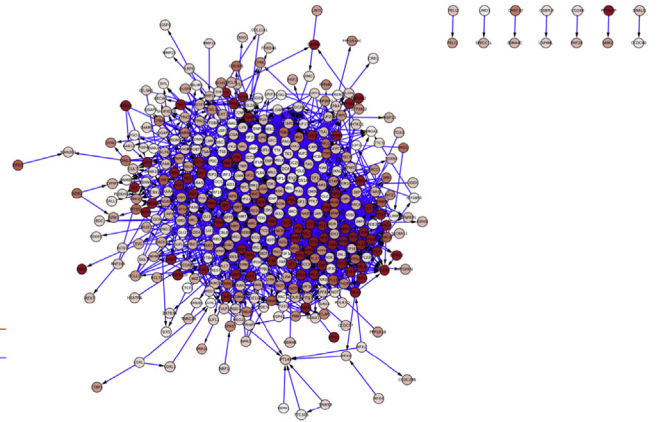
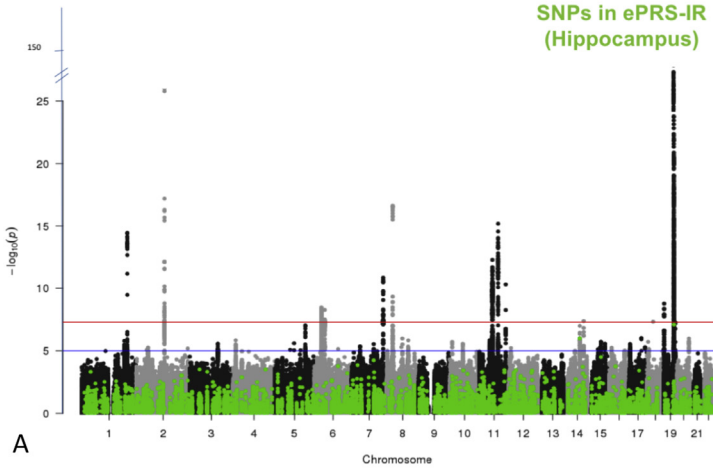
Sample descriptives	Spearman correlation coefficient	P-value
<b>Study participants' characteristics correlation with mesocorticolimbic ePRS-IR</b>		
<i>MAVAN cohort</i>		
Sex <sup>a</sup>	NA	0.785
Birth weight (grams) <sup>c</sup>	-0.05	0.459
Gestational age (weeks) <sup>c</sup>	0.02	0.817
Family income below Low Income Cut Off <sup>a</sup>	NA	0.945
<i>SAGE cohort</i>		
Sex <sup>a</sup>	NA	0.193
Family income below \$20 K <sup>a</sup>	NA	0.266
Age at interview <sup>c</sup>	0.03	0.159
Study source <sup>b</sup>	NA	0.033*
<b>Study participants' characteristics correlation with hippocampal ePRS-IR.</b>		
<i>MAVAN cohort</i>		
Sex <sup>a</sup>	NA	0.584
Birth weight (grams) <sup>c</sup>	-0.07	0.276
Gestational age (weeks) <sup>c</sup>	-0.04	0.585
Family income below Low Income Cut Off <sup>a</sup>	NA	0.063
<i>ADA cohort</i>		
Sex <sup>a</sup>	NA	0.558
Study site <sup>b</sup>	NA	0.180
Age at onset <sup>c</sup>	0.08	0.033*

Statistical analyses: <sup>a</sup> *t*-tests, <sup>b</sup> one-way ANOVA, <sup>c</sup> Spearman correlation. Low Income Cut Off according to Statistics Canada [96].



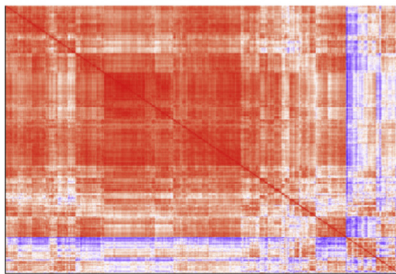
### Alzheimer's disease GWAS (Lambert et al., 2013)

SNPs in ePRS-IR  
(Hippocampus)

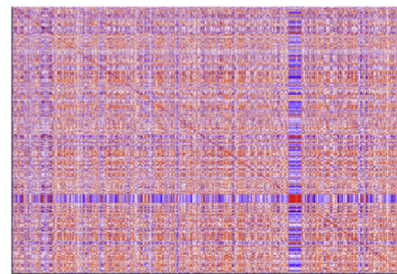


B

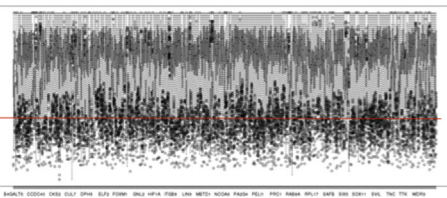
Childhood



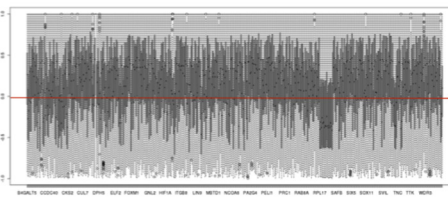
Adulthood



Correlation coefficient



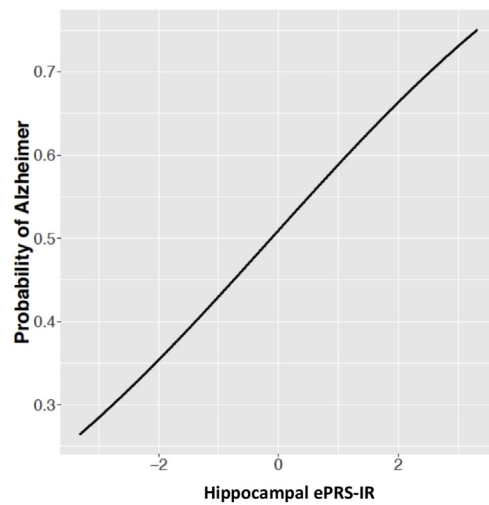
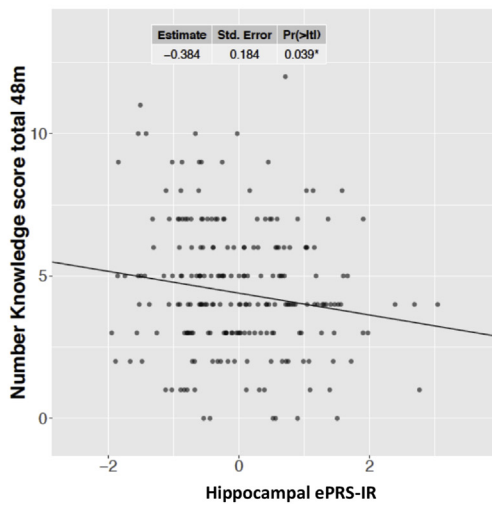
Correlation coefficient



C

Genes

Genes



To further validate the brain region specificity of the scores, we analyzed the distribution of children in each quartile of the mesocorticolimbic and hippocampal ePRS-IR in MAVAN (Fig. 4B) and verified that the scores are independent ( $r = 0.062$ ,  $p = .416$ ). In addition, the mesocorticolimbic ePRS-IR did not predict cognitive performance ( $\hat{\beta} = -0.27$ ,  $p = .08$  for main effect,  $\hat{\beta} = -0.15$ ,  $p = .62$  for the interaction with sex). Likewise, the hippocampal ePRS-IR was not associated with IST scores ( $\hat{\beta} = -0.006$ ,  $p = .62$  for main effect,  $\hat{\beta} = 0.020$ ,  $p = .29$ , for the interaction with sex). Gene ontology (pathways and processes networks) of the gene set comprising the mesocorticolimbic ePRS-IR using Metacore® (Thomson Reuters) showed statistically significant enrichment for pathways involved in cell cycle regulation (FDR  $q = 1.18 \times 10^{-2}$ ). Process networks were significant for neurogenesis/axonal guidance (FDR  $q = 6.667 \times 10^{-3}$ ), transcription and translation (FDR  $q = 1.39 \times 10^{-2}$ ). The gene set included in the hippocampal ePRS-IR is enriched for pathways involved in development (hedgehog signaling, FDR  $q = 1.19 \times 10^{-5}$ ) and transcription (FDR  $q = 1.42 \times 10^{-4}$ ). Process networks involve cell division (Fig. 4C). Both ePRS-IRs were significantly associated with similar transcription factors including CREB1, C-Myc, SP1 and ESF1 (Fig. 4D).

#### 4. Discussion

Recently, there has been a marked increase in the number of GWAS available. We now have access to association studies addressing a range of health related outcomes. Conventional polygenic risk scores derived from existing GWAS's are intended to reflect genetic vulnerability. However, the PRS approach is limited by the small explained variance in GWAS of common disorders and a lack of insight into the phenotypic constructs of the disease. We propose an alternative approach that integrates the power of GWAS datasets to create an ePRS informed by biological information in specific tissues. The resulting brain-region specific ePRS reflects an integrated gene network and molecular pathway that allows the use of genotyping data to test candidate biological mechanisms (Fig. 5).

The conventional polygenic profiling method is driven by a statistical approach. Additional levels of biological complexity, such as biomarkers and neuroanatomic loci of relevance – from experimental or theoretical databases – can be superimposed and inform the use of this technology, as we demonstrate here. The need to consider gene networks is further exemplified by the observation that the power to detect a causal SNP is drastically reduced when the trait/disorder is caused by two or more pathways [75]. We showed that a biologically-informed polygenic risk score based on genes co-expressed with the central IR represents highly cohesive and relevant gene networks. The mesocorticolimbic-specific ePRS-IR is more strongly associated with impulsivity and the risk for substance dependence in males than is a 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 [44,76]. Likewise the hippocampal-specific

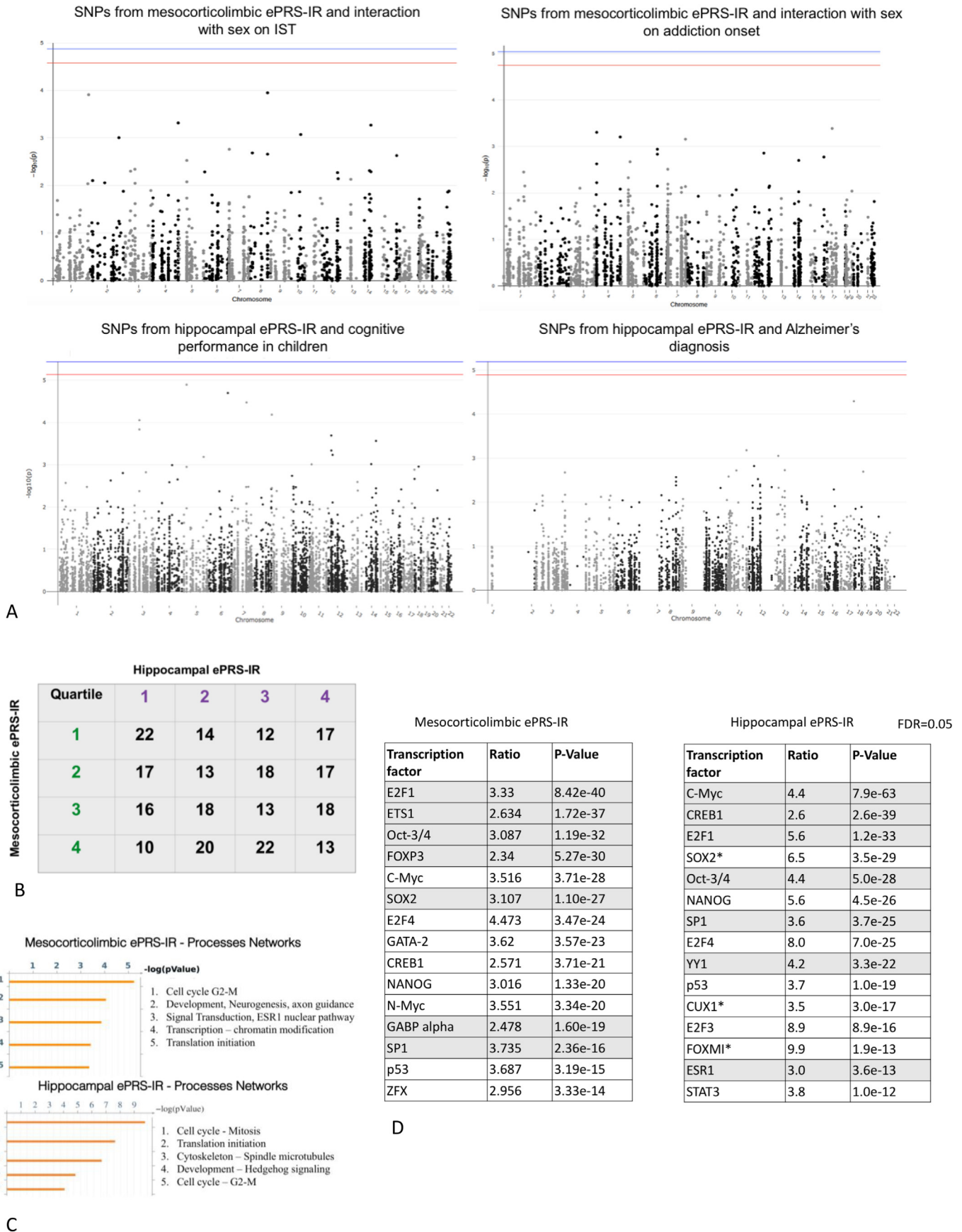
ePRS-IR was associated both with cognitive performance in children and with Alzheimer's diagnosis at later ages, demonstrating that the biologically-informed framework is brain region and endophenotype-specific.

We highlight the fact that our approach is focused on the biological processes underlying endophenotypic differences in healthy/community samples, rather than in SNPs statistically associated with a disorder in clinical samples as identified by the traditional GWAS and PRS methodologies. This is clear in Figs. 1B and 2A, in which we demonstrate that the SNPs selected to be part of the ePRS-IR scores are not statistically significant in the GWAS. However, they are biologically relevant for the insulin receptor gene network in the mesocorticolimbic system and hippocampus, which makes their statistical significance with regards to ADHD or Alzheimer's disease irrelevant for the purpose of the creation of the score. The betas extracted from the GWAS are used simply as a way to weight the SNPs and be able to aggregate them in a meaningful way. We could have used arbitrary weights as previously done in past multilocus approaches [77,78]. The comparison shown here (point 3.3), performing an interchange in SNP weights between 2010 and 2019 ADHD GWASes suggests that the weight attributed to each individual SNP contributes minimally to the prediction of the outcome. Fig. 4A also highlights the fact that no isolated SNP was associated with the outcomes in any of the analysis – which is consistent with the idea that the functional unit of analysis is the gene network (represented by the ePRS) rather than a single or only a few SNPs. Our approach is then more biologically plausible, considering that genes code for biological functions, and not for a disease state. Our methodology innovates in the SNP selection method – focused on the biological process rather than the statistical significance. Hence, as our scores represent biological variability rather than presence or absence of disease, our method is more likely linked to subtle variations in endophenotypes rather than the dichotomous presence or absence of a diagnosis. We are interested in predicting potential risk for disease later in life in healthy subjects, therefore the focus on children and on community samples.

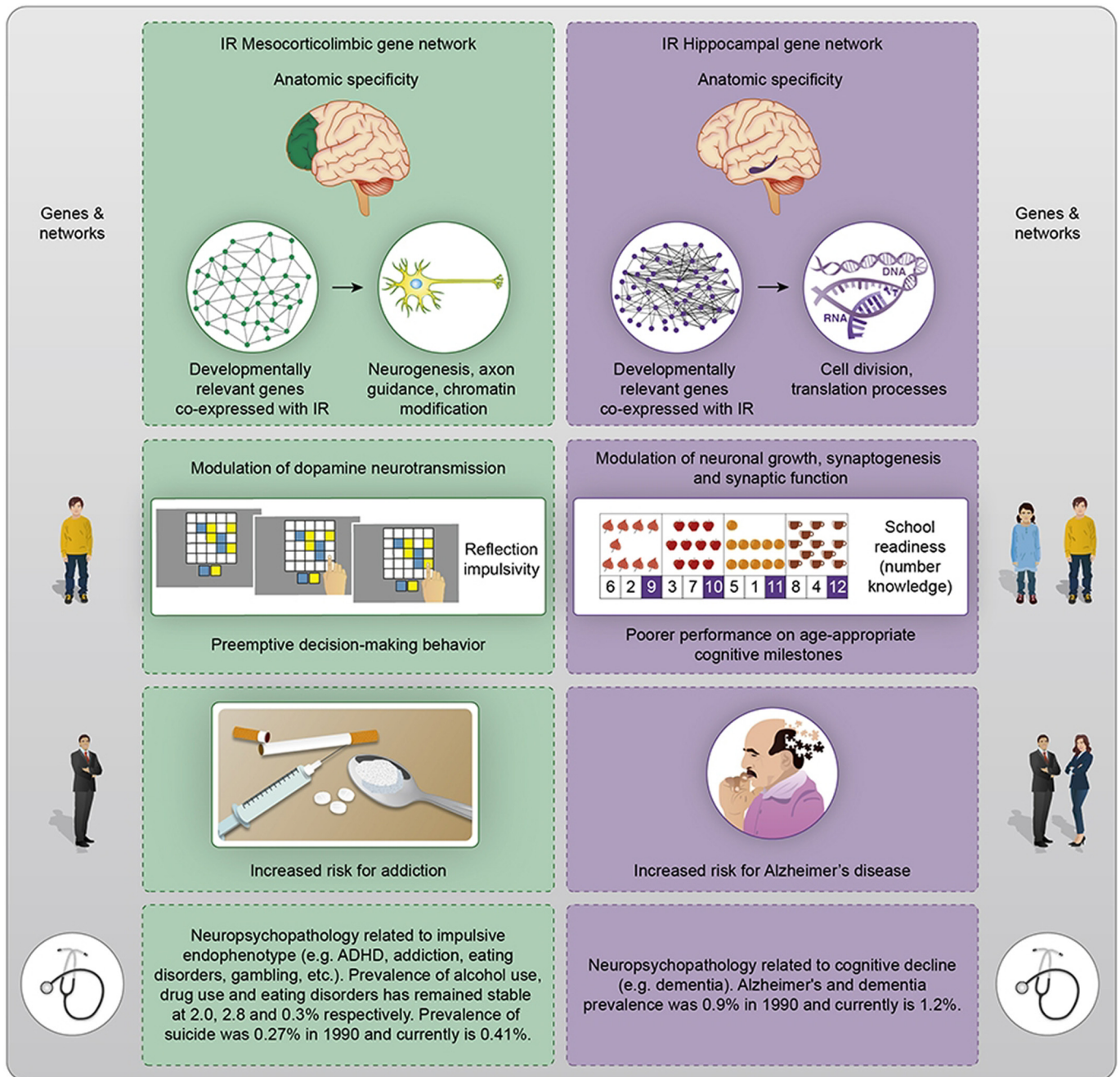
Our network analysis shows that the ePRS-IR represents a cohesive gene network with significantly more connections than the list of genes extracted from the GWAS for ADHD, or random lists of SNPs. This robust approach therefore goes beyond describing associations between single gene variants and the outcomes, but captures information about the whole gene network, and its function, in specific brain regions. One of the critiques of GWAS is that they are 'a static global' measure, unable to discern spatial or temporal differences, as compared to techniques such as RNA sequencing or whole genome bisulfite sequencing (WGBS), which are fundamental to the study of any trait/disorder. Our approach integrates a static global measure – which is easy to acquire in your dataset of interest – with a temporal and spatial information, which are, usually, not feasible to acquire in all human datasets – to compute a score that is adept at informing spatially relevant endophenotypes.

We filtered our gene set by genes overexpressed in early development by comparison to adulthood. According to [79], most of the

**Fig. 3.** The hippocampal ePRS-IR: (A) Single nucleotide polymorphisms (SNPs) included in the hippocampal ePRS-IR, shown in green, in relation to the Manhattan plot of the Alzheimer's GWAS. The SNPs captured in the score do not have genome-wide significance. (B) Gene network analysis in the hippocampal ePRS-IR 543 genes. The picture demonstrates that the hippocampal ePRS-IR also represents a highly cohesive gene network. (C) Coexpression of the genes included in the hippocampal ePRS-IR, in childhood and adulthood. Top panels: The heatmap of the genes' co-expression in childhood (left) shows several clusters of highly co-expressed genes. The clusters are not maintained in the retained order heatmap for adulthood (right) suggesting that these gene networks are relatedly connected in childhood, but not anymore in adulthood. Bottom panels: values of the correlation coefficients of the gene expression values in childhood (left) and adulthood (right). Each vertical line represents correlation with a unique gene. The red line is drawn at a correlation score of 0. Genes included in the score have highly correlated gene expression in childhood, but this is not seen in adult samples. (D) Main effect of hippocampal ePRS-IR on Number Knowledge Task performance, applied at 48 months, in which children with higher ePRS-IR scores have lower performance on the cognitive test. No significant interaction between ePRS and sex was found at this age ( $N = 218$ ). (E) Presence of Alzheimer's disease diagnosis according to hippocampal ePRS-IR scores in the GenADA study. There was a main effect of the hippocampal ePRS-IR in predicting Alzheimer's cases. There is an increased number of cases in individuals with higher ePRS-IR scores.  $N = 1565$ .



**Fig. 4.** ePRS-IR scores are brain-region specific: (A) Manhattan plots investigating the contribution of isolated SNPs into the reported findings. There was no unique SNP included in the mesocorticolimbic ePRS-IR responsible for the SNP by sex interaction on impulsivity in children (top left) or alcohol dependence in adults (top right). Similarly, no isolated SNP included in the hippocampal ePRS-IR was associated with poorer cognitive performance in children (bottom left), or Alzheimer's disease (bottom right). These findings confirm that variations in the biological function of the gene networks represented in the mesocorticolimbic ePRS-IR and hippocampal ePRS-IR can predict these outcomes as a global score, with small contributions from all included SNPs, but not by any isolated mutations. (B) Pearson's product-moment correlation demonstrating the distribution of children in each quartile of the mesocorticolimbic and hippocampal ePRS-IRs. There is no correlation between the quartiles of the two ePRSs ( $r = 0.062, p = .416$ ). This shows that a subject with a high mesocorticolimbic ePRS-IR would not necessarily have a high score in the hippocampal ePRS-IR, and vice-versa, demonstrating their brain specificity and independence. (C) Enrichment analysis of mesocorticolimbic (upper panel) and hippocampal (bottom panel) ePRS-IR. (D) Transcription factor analysis of the two brain-region specific ePRS-IRs.



**Fig. 5.** Theoretical framework and applications: The expression based, brain region specific ePRS-IR represents a developmentally relevant gene network with enrichment for specific processes like neurogenesis, chromatin modification and translation processes. Variation in this score predicts particular endophenotypes in childhood – for example, impulsivity and cognitive performance. These behaviors map onto risk for diseases later in life, such as risk for addiction and Alzheimer's disease, and the ePRS-IR is equally associated with these outcomes. Psychopathology related to poor inhibitory control or decision-making encompasses a wide range of conditions (ADHD, addiction, eating disorders, gambling, suicide, etc.). Similarly, there is an increasing incidence of Alzheimer's disease and other dementias. The neurobiological understanding of these conditions contributes to the development of tools for early identification and primary prevention.

differentially expressed genes during development have the highest expression in the fetal/early childhood period (81% of all differentially expressed genes during development, as opposed to only 5% in adulthood). The same author describes that these overexpressed genes in early life are enriched for gene ontology processes related to synaptic transmission and signaling [79], which are relevant for the current study. Genes that are overexpressed in adulthood vs. early life are enriched for cellular respiration [79]. The threshold that we used for selection was based on previous literature examining enrichment of genes associated with neuropsychiatric disorders in fetal versus adult brain [80].

Our mesocorticolimbic ePRS-IR predicted both impulsivity in male children as well as risk for addiction in adults. Conduct disorder and impulsivity are the foremost risk traits for alcohol use disorder among the 80 personality disorder criteria of DSM-IV [81]. There is a relationship between childhood ADHD and the risk for developing drug addiction later in life [82], especially considering the impulsivity component, rather than inattention [83], in agreement to the findings described here. Insulin function is associated with the risk for drug addiction [84]. Diminished insulin sensitivity is related to less endogenous dopamine at D2/3 receptors in the ventral striatum [85], reinforcing the idea that metabolic processes

are involved in dysfunctions of the mesocorticolimbic system, such as drug dependence. The enrichment analysis of our mesocorticolimbic ePRS-IR is consistent with the known role of insulin in cell division [86], neurogenesis [87], signal transduction [88], and axon guidance [89].

The hippocampal formation is the region with the higher IR gene expression in the brain [90]. Lower expression of IR and altered levels of different components of the insulin signaling pathways were described in hippocampi from Alzheimer's patients [91]. A down-regulation in insulin signaling causes memory impairment through inhibition of serine phosphorylation of the insulin receptor via activation of stress kinases like c-Jun-n-terminal kinase (JNK) [17]. JNK enhances glycogen synthase kinase 3 $\beta$  (GSK3 $\beta$ ) activity that leads to tau phosphorylation followed by accumulation of neurofibrillary tangles – a hallmark of Alzheimer's disease – in the brain [92]. Others have shown that impaired insulin sensitivity is linked to cognitive deficits in the elderly [93], and that measures of cognitive function in youth predict later risk for Alzheimer's disease [94,95]. Our study is based on this premise. That's why we (1) created the scores favoring genes that are overexpressed in childhood vs. adulthood; (2) started our analysis by verifying if the hippocampal ePRS-IR would predict cognitive function in children and (3) tested if the same genetic score could predict Alzheimer's diagnosis in a cohort of adults. It is interesting to note that, as opposed to the mesocorticolimbic ePRS-IR gene set, clusters of co-expression of the hippocampal ePRS-IR gene set are not maintained in adulthood. This indeed suggests that the function of this gene network early in life defines an endophenotype in childhood (cognitive ability), that persists as an increased risk for dementia at older ages.

It is important to note that the co-expression data is derived exclusively from mice in this work (see Fig. 1A). The co-expression was then confirmed in humans using the Brainspan dataset (see Figs. 1E, 3C). The use of co-expression data from mice allows the establishment of direct translational studies, integrating information from clinically-relevant experimental models into this bioinformatic approach. In this study, we used a publicly available dataset (GeneNetwork); in future studies, co-expression data from relevant animal models can be the source for the co-expression data.

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. The insulin receptor work shown here serves as a proof-of-concept for this methodology, that can be focused on other gene networks or other tissues, and be applicable to virtually all fields of research. As GWASes become more powerful and more specific for detecting disease states in clinical samples, they are also progressively less sensitive to capture the biological spectrum and endophenotypes – therefore less relevant for community samples and for large-scale applicability in preventative measures. The concept that genes code for biological processes rather than for disease, and work in networks rather than in isolation, is the basic idea of our method. Our ground-breaking approach applies the GWAS technology to the understanding of the adaptive and plasticity responses to developmental and environmental variation.

## Acknowledgments

This work was funded by the Toxic Stress Research network of the JPB Foundation and Sackler 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. Funders had no influence on study design, data collection, data analysis, interpretation, writing of the report.

## Contributors

All authors contributed substantively to this work. Conceptualization PPS, MJM, SHD; data curation KM, IP, LMC, TTTN, EG, LMM, JLM; formal analysis IP, SHD, PPS; funding acquisition MJM, MK; investigation PPS, KM, SHD; methodology PPS, SHD, KOD, MY; project administration JD; resources LMM, JLM, EG, TTTN; supervision MJM, PPS; validation SHD, LMC, MY; visualization SHD, KM, LMC; writing – original draft PPS, SHD, MJM; writing – review & editing all authors.

## Declaration of interests

Authors declare no competing interests.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ebiom.2019.03.051>.

## 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):1069–78.
- [2] Guo M, Mi J, Jiang QM, Xu JM, Tang YY, Tian G, et al. Metformin may produce antidepressant effects through improvement of cognitive function among depressed patients with diabetes mellitus. *Clin Exp Pharmacol Physiol* 2014;41(9):650–6.
- [3] Erensoy H, Niafar M, Ghafarzadeh S, Aghamohammadzadeh N, Nader ND. A pilot trial of metformin for insulin resistance and mood disturbances in adolescent and adult women with polycystic ovary syndrome. *Gynecol Endocrinol* 2018:1–4.
- [4] Wang CP, Lorenzo C, Habib SL, Jo B, Espinoza SE. Differential effects of metformin on age related comorbidities in older men with type 2 diabetes. *J Diabetes Complicat* 2017;31(4):679–86.
- [5] Havrankova J, Schmechel D, Roth J, Brownstein M. Identification of insulin in rat brain. *Proc Natl Acad Sci U S A* 1978;75(11):5737–41.
- [6] Figlewicz DP. Insulin, food intake, and reward. *Semin Clin Neuropsychiatry* 2003;8(2):82–93.
- [7] Sipols AJ, Bayer J, Bennett J, Figlewicz DP. Intraventricular insulin decreases kappa opioid-mediated sucrose intake in rats. *Peptides* 2002;23(12):2181–7.
- [8] Figlewicz DP. Expression of receptors for insulin and leptin in the ventral tegmental area/substantia nigra (VTA/SN) of the rat: historical perspective. *Brain Res* 1645;2016:68–70.
- [9] 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):1952–5.
- [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–R94.
- [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):1284–91.
- [12] Kleinriders A, Cai W, Cappellucci L, Ghazarian A, Collins WR, Vienberg SG, et al. Insulin resistance in brain alters dopamine turnover and causes behavioral disorders. *Proc Natl Acad Sci U S A* 2015;112(11):3463–8.
- [13] Kemp DE, Ismail-Beigi F, Ganocy SJ, Conroy C, Gao K, Obral S, 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 Disord* 2012;136(3):1164–73.
- [14] Stoeckel LE, Arvanitakis Z, Gandy S, Small D, Kahn CR, Pascual-Leone A, 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;77:165–76.
- [16] Cortese S, Tessari L. Attention-Deficit/Hyperactivity Disorder (ADHD) and obesity: update 2016. *Curr Psychiatry Rep* 2017;19(1).
- [17] Bomfim TR, Forny-Germano L, Sathler LB, Brito-Moreira J, Houzel JC, Decker H, et al. An anti-diabetes agent protects the mouse brain from defective insulin signaling caused by Alzheimer's disease-associated Abeta oligomers. *J Clin Invest* 2012;122(4):1339–53.
- [18] Moran C, Beare R, Phan TG, Bruce DG, Callisaya ML, Srikanth V. Type 2 diabetes mellitus and biomarkers of neurodegeneration. *Neurology* 2015;85:1123–30.

- [19] Arendt T, Stieler J, Ueberham U. Is sporadic Alzheimer's disease a developmental disorder? *J Neurochem* 2017;143(4):396–408.
- [20] Mustillo S, Worthman C, Erkanli A, Keeler G, Angold A, Costello EJ. Obesity and psychiatric disorder: developmental trajectories. *Pediatrics* 2003;111(4 Pt 1):851–9.
- [21] Chen Q, Hartman CA, Kuja-Halkola R, Faraone SV, Almqvist C, Larsson H. Attention-deficit/hyperactivity disorder and clinically diagnosed obesity in adolescence and young adulthood: a register-based study in Sweden. *Psychol Med* 2018:1–9.
- [22] Bowling A, Davison K, Haneuse S, Beardslee W, Miller DP. ADHD medication, dietary patterns, physical activity, and BMI in children: a longitudinal analysis of the ECLS-K study. *Obesity (Silver Spring)* 2017;25(10):1802–8.
- [23] O'Neill S, Rajendran K, Mahubani SM, Halperin JM. Preschool predictors of ADHD symptoms and impairment during childhood and adolescence. *Curr Psychiatry Rep* 2017;19(12):95.
- [24] Luo SX, Levin FR. Towards precision addiction treatment: new findings in co-morbid substance use and attention-deficit hyperactivity disorders. *Curr Psychiatry Rep* 2017;19(3):14.
- [25] Soto N, Bazaes RA, Pena V, Salazar T, Avila A, Iniguez G, et al. Insulin sensitivity and secretion are related to catch-up growth in small-for-gestational-age infants at age 1 year: results from a prospective cohort. *J Clin Endocrinol Metab* 2003;88(8):3645–50.
- [26] Silveira PP, Pokhvisneva I, Gaudreau H, Rifkin-Graboi A, Broekman BFP, Steiner M, et al. Birth weight and catch up growth are associated with childhood impulsivity in two independent cohorts. *Sci Rep* 2018;8(1):13705.
- [27] Kullmann S, Heni M, Hallschmid M, Fritsche A, Preissl H, Haring HU. Brain insulin resistance at the crossroads of metabolic and cognitive disorders in humans. *Physiol Rev* 2016;96(4):1169–209.
- [28] Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry* 2014;55(10):1068–87.
- [29] Wray NR, Goddard ME. Multi-locus models of genetic risk of disease. *Genome Med* 2010;2(2):10.
- [30] O'Donnell KA, Gaudreau H, Colalillo S, Steiner M, Atkinson L, Moss E, et al. The Maternal Adversity Vulnerability and Neurodevelopment (MAVAN) project: theory and methodology. *Can J Psychiatr* 2014;59(9):497–508.
- [31] Bierut LJ, Saccone NL, Rice JP, Goate A, Foroud T, Edenberg H, et al. Defining alcohol-related phenotypes in humans. The collaborative study on the genetics of alcoholism. *Alcohol Res Health* 2002;26(3):208–13.
- [32] Edenberg HJ. The collaborative study on the genetics of alcoholism: an update. *Alcohol Res Health* 2002;26(3):214–8.
- [33] Edenberg HJ, Xuei X, Chen HJ, Tian H, Wetherill LF, Dick DM, et al. Association of alcohol dehydrogenase genes with alcohol dependence: a comprehensive analysis. *Hum Mol Genet* 2006;15(9):1539–49.
- [34] Bierut LJ. Genetic variation that contributes to nicotine dependence. *Pharmacogenomics* 2007;8(8):881–3.
- [35] Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet* 2007;16(1):24–35.
- [36] Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA, 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.
- [37] 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.
- [38] Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* 2008;65(1):45–53.
- [39] Filippini N, Rao A, Wetten S, Gibson RA, Borrie M, Guzman D, et al. Anatomically-distinct genetic associations of APOE epsilon4 allele load with regional cortical atrophy in Alzheimer's disease. *Neuroimage* 2009;44(3):724–8.
- [40] 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. *Dev Neuropsychol* 2002;22(3):595–624.
- [41] Bennett D, Oldham S, Dawson A, Parkes L, Murawski C, Yücel M. Systematic Overestimation of Reflection Impulsivity in the Information Sampling Task. *Biol Psychiatry* 2017;82(4):e29–30.
- [42] Clark L, Robbins TW. Reply to: systematic overestimation of reflection impulsivity in the information sampling task. *Biol Psychiatry* 2017 Aug 15;82(4):e29.
- [43] Pokhvisneva I, Léger É, Meaney MJ, Silveira PP. Systematic overestimation of reflection impulsivity in the information sampling task: age dependency in children. *Biol Psychiatry* 2018;83(2):e33–e4.
- [44] Silveira PP, Agranonik M, Faras H, Portella AK, Meaney MJ, Levitan RD, et al. Preliminary evidence for an impulsivity-based thrifty eating phenotype. *Pediatr Res* 2012;71(3):293–8.
- [45] Murray AL, Booth T, Eisner M, Auyeung B, Murray G, Ribeaud D. Sex differences in ADHD trajectories across childhood and adolescence. *Dev Sci* 2019;22(1):e12721.
- [46] Fossati A, Gratz KL, Borroni S, Maffei C, Somma A, Carlotta D. The relationship between childhood history of ADHD symptoms and DSM-IV borderline personality disorder features among personality disordered outpatients: the moderating role of gender and the mediating roles of emotion dysregulation and impulsivity. *Compr Psychiatry* 2015;56:121–7.
- [47] Lemelin JP, Boivin M, Forget-Dubois N, Dionne G, Seguin JR, Brendgen M, et al. The genetic-environmental etiology of cognitive school readiness and later academic achievement in early childhood. *Child Dev* 2007;78(6):1855–69.
- [48] Okamoto Y, Case R. Exploring the microstructure of children's central conceptual structures in the domain of number. *Monogr Soc Res Child Dev* 1996;61(1–2):27–58.
- [49] Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger Jr JJ, 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):149–58.
- [50] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* Washington, D.C. ; 2000.
- [51] McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016;48(10):1279–83.
- [52] Mulligan MK, Mozhui K, Prins P, Williams RW. GeneNetwork: a toolbox for systems genetics. *Methods Mol Biol* 2017;1488:75–120.
- [53] Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. *Nature* 2014;508(7495):199–206.
- [54] Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch KP, et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010;49(9):884–97.
- [55] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013;45(12):1452–8.
- [56] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 2013;41(Database issue):D808–15.
- [57] Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 2010;38(Web Server issue):W214–20.
- [58] Saito R, Smoot ME, Ono K, Ruschinski J, Wang PL, Lotia S, et al. A travel guide to Cytoscape plugins. *Nat Methods* 2012;9(11):1069–76.
- [59] Chen LM, Yao N, Garg E, Zhu Y, Ngyuen TTT, Pokhvisneva I, et al. PRS-on-Spark (PRSos): a novel, efficient and flexible approach for generating polygenic risk scores. *BMC Bioinforma* 2018;19(1):295.
- [60] Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 2019;51(1):63–75.
- [61] Tobacco Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010;42(5):441–7.
- [62] Rubin DB. Multiple Imputation for Nonresponse in Surveys; 1987.
- [63] Sv Buuren, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011;45(3).
- [64] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38(8):904–9.
- [65] Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006;2(12):e190.
- [66] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;4:7.
- [67] R CT. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria; 2014.
- [70] Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 2012;44(6):659–69.
- [72] Fatseas M, Hurmic H, Serre F, Debrabant R, Daulouede JP, Denis C, et al. Addiction severity pattern associated with adult and childhood Attention Deficit Hyperactivity Disorder (ADHD) in patients with addictions. *Psychiatry Res* 2016;246:656–62.
- [73] Groen-Blokhuis MM, Middeldorp CM, Kan K-J, Abdellaoui A, van Beijsterveldt CEM, Ehli EA, et al. Attention-deficit/hyperactivity disorder polygenic risk scores predict attention problems in a population-based sample of children. *J Am Acad Child Adolesc Psychiatry* 2014;53(10):1123–9.e6.
- [74] Martin J, Hamsheer ML, Stergiakouli E, O'Donovan MC, Thapar A. Genetic risk for attention-deficit/hyperactivity disorder contributes to neurodevelopmental traits in the general population. *Biol Psychiatry* 2014;76(8):664–71.
- [75] Wray NR, Maier R. Genetic basis of complex genetic disease: the contribution of disease heterogeneity to missing heritability. *Curr Epidemiol Rep* 2014;1(4):220–7.
- [76] Vasiladias HM, Diallo FB, Rochette L, Smith M, Langille D, Lin E, 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 Psychiatr* 2017;706743717714468.
- [77] Stice E, Yokum S, Burger K, Epstein L, Smolen A. Multilocus genetic composite reflecting dopamine signaling capacity predicts reward circuitry responsivity. *J Neurosci* 2012;32(29):10093–100.
- [78] Silveira PP, Pokhvisneva I, Gaudreau H, Atkinson L, Fleming AS, Sokolowski MB, et al. Fetal growth interacts with multilocus genetic score reflecting dopamine signaling capacity to predict spontaneous sugar intake in children. *Appetite* 2018;120:596–601.
- [79] Jaffe AE, Shin J, Collado-Torres L, Leek JT, Tao R, Li C, et al. Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. *Nat Neurosci* 2015;18(1):154–61.
- [80] Birnbaum R, Jaffe AE, Hyde TM, Kleinman JE, Weinberger DR. Prenatal expression patterns of genes associated with neuropsychiatric disorders. *Am J Psychiatry* 2014;171(7):758–67.
- [81] Rosenström T, Torvik FA, Ystrom E, Czajkowski NO, Gillespie NA, Aggen SH, Krueger RF, Kendler KS, Reichborn-Kjennerud T, et al. Prediction of alcohol use disorder using personality disorder traits: a twin study. *Addiction* 2018;113(1):15–24.

- [82] Levy S, Katusic SK, Colligan RC, Weaver AL, Killian JM, Voigt RG, et al. Childhood ADHD and risk for substance dependence in adulthood: a longitudinal, population-based study. *PLoS One* 2014;9(8):e105640.
- [83] Sihvola E, Rose RJ, Dick DM, Korhonen T, Pulkkinen L, Raevuori A, et al. Prospective relationships of ADHD symptoms with developing substance use in a population-derived sample. *Psychol Med* 2011;41(12):2615–23.
- [84] O'Dell LE, Nazarian A. Enhanced vulnerability to tobacco use in persons with diabetes: a behavioral and neurobiological framework. *Prog Neuro-Psychopharmacol Biol Psychiatry* 2016;65:288–96.
- [85] Caravaggio F, Borlido C, Hahn M, Feng Z, Fervaha G, Gerretsen P, 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.
- [86] Perdureau D, Cailliau K, Browaeys-Poly E, Lescuyer A, Carre N, Benhamed F, et al. Insulin-induced cell division is controlled by the adaptor Grb14 in a Chfr-dependent manner. *Cell Signal* 2015;27(4):798–806.
- [87] Ronaghi A, Zibaii MI, Pandamooz S, Nourzei N, Motamedi F, Ahmadiani A, et al. Entorhinal cortex stimulation induces dentate gyrus neurogenesis through insulin receptor signaling. *Brain Res Bull* 2019;144:75–84.
- [88] Pearson-Leary J, Jahagirdar V, Sage J, McNay EC. Insulin modulates hippocampally-mediated spatial working memory via glucose transporter-4. *Behav Brain Res* 2018;338:32–9.
- [89] Gupta S, Yadav K, Mantri SS, Singhal NK, Ganesh S, Sandhir R. Evidence for compromised insulin signaling and neuronal vulnerability in experimental model of sporadic Alzheimer's disease. *Mol Neurobiol* 2018;55(12):8916–35.
- [90] De Felice FG, Benedict C. A key role of insulin receptors in memory. *Diabetes* 2015;64(11):3653–5.
- [91] Talbot K, Wang HY, Kazi H, Han LY, Bakshi KP, Stucky A, et al. Demonstrated brain insulin resistance in Alzheimer's disease patients is associated with IGF-1 resistance, IRS-1 dysregulation, and cognitive decline. *J Clin Invest* 2012;122(4):1316–38.
- [92] Bedse G, Di Domenico F, Serviddio G, Cassano T. Aberrant insulin signaling in Alzheimer's disease: current knowledge. *Front Neurosci* 2015;9:204.
- [93] Benedict C, Brooks SJ, Kullberg J, Burgos J, Kempton MJ, Nordenskjold R, et al. Impaired insulin sensitivity as indexed by the HOMA score is associated with deficits in verbal fluency and temporal lobe gray matter volume in the elderly. *Diabetes Care* 2012;35(3):488–94.
- [94] Snowdon DA, Greiner LH, Marksbery WR. Linguistic ability in early life and the neuropathology of Alzheimer's disease and cerebrovascular disease. Findings from the Nun study. *Ann N Y Acad Sci* 2000;903:34–8.
- [95] Bloss CS, Delis DC, Salmon DP, Bondi MW. Decreased cognition in children with risk factors for Alzheimer's disease. *Biol Psychiatry* 2008;64(10):904–6.
- [96] Statistics Canada. Low Income Cut Offs for 2005 and Low Income measures for 2004; 2005 (Catalogue no 75F0002MIE: Statistics Canada).