Corticolimbic DCC gene co-expression networks as predictors of impulsivity in children

Jose M. Restrepo-Lozano1,2, Irina Pokhvisneva2,3, Zihan Wang2,3, Sachin Patel2,3, Michael J. Meaney2,3,4,5,6, Patricia P. Silveira2,3,5,6, and Cecilia Flores1,2,4,5

© The Author(s) 2022

INTRODUCTION

Several psychiatric disorders of developmental origin are characterized by deficits in cognitive control – a compromised ability to voluntarily choose a context-appropriate goal-directed response. Altered connections and communication between prefrontal and striatal regions appear to be at the core of this behavioral trait [1], but the underlying neurobiological processes, as well as early markers of vulnerability, are yet to be discovered [2–4]. The cognitive capacity to control and override impulsive behaviors improves gradually from childhood to early adulthood, mirroring the protracted developmental trajectory of the prefrontal cortex (PFC) [5–9], and its gradual quantitative and qualitative changes in dopamine innervation [10–12]. While dopamine axons establish local connections in the nucleus accumbens (NAcc) in adolescence, mesocortical dopamine axons are still growing from the NAcc to the PFC across this period [13–18]. The extent and organization of mesocortical dopamine axon growth in adolescence determines the organization of local PFC circuitry and cognitive function in adulthood [19–21].

The developmental trajectories of mesocortical and mesolimbic dopamine inputs are temporally different but have a reciprocal functional connection [21–24]. Transient postnatal developmental overexpression of dopamine D2 receptor in the striatum leads to adult mesocortical dopamine PFC dysfunction and cognitive deficits, indicating that striatal dopamine maturational events interact with those controlling mesocortical dopamine axon growth [25]. Changes in PFC dopamine neurotransmission are associated with opposite changes in NAcc dopamine function [23, 26], and alterations in D1- or D2-expressing NAcc pathways impact gene expression in the PFC [27]. Clearly, PFC and cognitive control development involve the recruitment of corticostriatal neuronal networks [28, 29].

A rapidly increasing number of studies show a strong association between genetic variability within the Netrin-1/DCC signaling pathway and its developmental onset, most notably those emerging in adolescence. These disorders are characterized by PFC and NAcc dysfunction and deficits in impulse control [30–33]. Early postnatal expression of the DCC gene network in the PFC associates with total brain volume in children [34], emphasizing that the Netrin-1/DCC guidance cue system is tightly linked to overall early neurodevelopment. In adolescent rodents, DCC-mediated Netrin-1 signaling organizes...
the maturation of dopamine networks by promoting mesolimbic dopamine axon targeting in the NAcc and controlling the growth of dopamine axons to the PFC [12, 17]. Changes in DCC receptor levels in adolescent mice lead to mis-targeting of mesolimbic dopamine axons in the NAcc and to their ectopic growth to PFC, altering PFC function and cognitive control in adulthood [17, 35]. Similar anatomical and behavioral changes occur in humans that are DCC mutation carriers [36, 37], indicating that the Netrin-1/DCC pathway is part of a gene network closely involved in corticolimbic development.

To date, most human genetic studies have focused on associations between genetic variants and phenotypes, and the estimated effects of a given number of variants can be aggregated into a score that represents individual genetic risk (called polygenic risk score; PRS). This association between genetic variation and behavior/disease ultimately results in relatively few genome-wide significant variants (e.g. [38]), most of which belong to noncoding portions of the genome and whose effect is diminished by the polygenic nature of complex phenotypes [39, 40]. Several of these non-coding variants are regulatory in nature, likely affecting the expression of nearby genes [40], ultimately placing gene expression as an intermediate molecular phenotype between genetics and disease [41]. Our approach exploits the facts that genes operate within complex networks, code with remarkable tissue-specificity for precise biological functions, and the likelihood of identifying relevant biological markers increases by relying on genotype-gene expression rather than genotype-disease associations (see [42, 43]).

We use a systems biology approach to create a genetic score based on genes co-expressed with a gene of interest in a specific brain region. We gather all SNPs from the co-expressed genes and assign for each SNP the effect size estimated by the Genome-Tissue Expression (GTeX) project [44], which quantifies the influence of variants on tissue-specific gene expression. We aggregate genotypes weighted by the GTEx across all SNPs within the co-expression network into an expression-based polygenic score (ePRS), according to the individual’s genotype [42, 43].

The relationship between genes and behavior is highly indirect, regardless of how strong the relationship may be. Here, we forgo direct genotype-disease associations to construct an ePRS based on genes co-expressed with DCC in the PFC and the NAcc. Our goal is to create a marker that captures individual variation in the processes involved in the maturation of corticolimbic substrates supporting inhibitory control. By modifying the approach to genomic profiling, we generated a biological marker that can help identify early vulnerability for impulsivity-related phenotypes. We tested the association of the biologically-informed genetic score with measurements of impulsivity in three ethnically different community samples of 6- and 10-year-old children.

### Identification of corticolimbic DCC gene co-expression networks and ePRS calculation

Figure 1 shows the steps involved in the identification of the gene networks and the ePRS score. The ePRS was calculated considering genes co-expressed with DCC in the PFC and NAcc. We aimed to capture DCC co-expression networks within each brain region, with the final ePRS being a joint representation of the functional co-expression networks in these two corticolimbic regions. As described previously [42, 43, 49], the score was created using the data from: (1) GeneNetwork (http://genenetwork.org), (2) BrainSpan (http://www.brainspan.org), (3) the NCBI Variation Viewer, U.S. National Library of Medicine, (NCBI) [50], (4) the GTex project [44], and (5) genotype data in the three cohorts. We used GeneNetwork to generate a list of genes co-expressed with DCC in the PFC and in the NAcc in mice (absolute value of co-expression correlation greater or equal to 0.5). We used gene expression datasets from mice (see Supplementary Data file) because our study is guided by our previous findings in rodents linking variations in Dcc expression to changes in neural control and in mesocorticolimbic dopamine axon targeting [13, 17, 51, 52]. To retain genes that are more active when the brain is still undergoing core maturational processes in humans, we used BrainSpan to select autosomal transcripts expressed at least 1.5-fold more during the early postnatal development (0–18 months after birth) than in adulthood (20–40 years of age), with the final networks consisting of 154 genes in the PFC (see Table S4) and 72 genes in the NAcc (see Table S5). For annotations, we used GRCi37.p13 assembly of the NCBI to source chromosome and start/ end positions for the co-expressed genes, which allowed us to gather all the gene-SNP pairs from the GTEx dataset in human PFC and NAcc (PFC: 41,053 SNPs; NAcc: 66,428 SNPs). These lists were merged with the genotyping data in each of the three cohorts, keeping only the common SNPs and subjecting the final genotyping data sets to linkage disequilibrium clumping ($r^2 < 0.2$) to eliminate highly correlated SNPs.

To calculate the ePRS, number of effect alleles at a given cis-SNP were weighted by the estimated brain-region-specific effect of the genotype on gene expression from the GTEx data. The ePRS was obtained by adding the weighted SNPs, accounting for the sign of the correlation between each gene’s expression and DCC gene expression. The sum of the estimated effects resulted in ePRS scores for the DCC co-expression networks in the PFC and NAcc, which were then aggregated (by summation of the two scores) into a single global genetic score termed “corticolimbic DCC-ePRS”.

Finally, an enrichment analysis was conducted to characterize the functional and biological properties of the gene networks that comprise the corticolimbic DCC-ePRS score. A description of the tools and datasets used throughout the study can be found in the Supplementary Material, Table S7.

### Behavioral outcomes

We tested whether the ePRS associates with two aspects of impulsivity: (i) impulsive choice, reflecting proneness to make risky choices, as measured by the Information Sampling Task; (ii) impulsive action, reflecting the ability to inhibit motor responses, as measured by the Stop-Signal Task. In both cases, the ability to self-regulate behavior is required for interrupting or inhibiting competing inputs or actions in order to accomplish a specific goal-directed response [1]. A description of the behavioral data obtained from each cohort is described in the Supplementary Material.

### Statistical analysis

Data were analyzed using R v3.6 [53] and Python v3.7 (https://www.python.org/), and polygenic scores were generated using the PRSoS pipeline (https://github.com/MeaneyLab/PRSoS). We considered two-tailed hypothesis tests, and significance levels for all tests were set at $\alpha = 0.05$. For each cohort we categorized the ePRS into high- and low-ePRS groups using a median split of the genetic score. Analysis of baseline characteristics was performed using Student’s t-test for continuous data (in case of unequal variances, Welch’s t test was used) and $X^2$ for categorical variables. Linear regression analysis was used to examine the association of the ePRS with the behavioral outcomes, adjusting for sex and population stratification. Adjustment for multiple comparisons was applied using Bonferroni method, independently for each behavioral construct/cohorts. All data were inspected to ensure that the assumptions for the tests and the linear regression analyses were met. Power analysis was conducted for linear multiple regression, considering the effect of the ePRS on the different outcomes: for $\alpha = 0.05$, sample size of 202, 398, and 4392, and a small effect size $f^2 = 0.02$, the achieved power will be 0.64, 0.88, and $>0.95$ in MAVAN, GUSTO, and ALSPAC, respectively.
RESULTS
We found no differences in baseline characteristics between ePRS groups in MAVAN, GUSTO, and ALSPAC cohorts (Table 1).

Behavioral outcomes
Lower Corticolimbic DCC-ePRS Scores Associate with Higher Measurements of Impulsivity in Children. Information Sampling Task: In MAVAN, the ePRS was associated with meanP-correct values (Fig. 2A: $\beta = -0.04, p = 0.045$); the low-ePRS group had lower meanP-correct values (less information sampled, lowering the probability of making a correct choice at the point of decision) indicating higher levels of impulsive choice in comparison to the high-ePRS group.

Stop-Signal Task: In GUSTO, the ePRS was associated with the proportion of successfully inhibited responses (Fig. 2B: $\beta = -0.03, p = 0.010$; Estimated SSRT: $\beta = -10.07, p = 0.57$; see Table S2 for complete results). After adjustments for multiple comparisons, the association between the ePRS and the proportion of successful stops in GUSTO cohort is no longer significant ($p = 0.054$). These results show that an ePRS score reflecting variability in the expression of corticolimbic DCC gene co-expression networks is associated with the levels of inhibitory control in children from ethnically diverse backgrounds.

Corticolimbic DCC gene co-expression networks: enrichment analysis
Protein–protein interaction (PPI). We used STRING [54] and Cytoscape [55] to visualize catalogued PPIs in protein products of genes within each co-expression network (networks were analyzed separately; only proteins with interactions are depicted in Fig. 3A). PFC: This network contains 152 nodes (one for each protein with at least 1 connection with another protein in the network) and 151 edges, corresponding to the mapped interactions among the nodes. The PPI enrichment ($p = 0.004$) indicates that this network contains more interactions than expected, compared to a network of equal size composed of a random set of proteins, and that the proteins are involved in common biological functions. NAcc: Contains 74 nodes and 50 edges, and the PPI enrichment ($p = 5.1e-11$) also suggests a strong biological connection among the proteins (see the corresponding gene networks, created using GeneMANIA [56], in Fig. S5).

Tissue-specific gene expression. We used FUMA [57] to visualize the expression levels of the genes from the co-expression networks across the 54 tissue-types included in GTEx. The PFC (comprising BA24 and BA9) and the NAcc are the 1st and 4th most enriched tissues for the gene networks’ expression (Fig. 3B).
Functional ontologies. Using Metacore™, we explored the biological context in which the gene networks operate, by mapping the genes from each network onto functional categories (Fig. 3C, Table S1 for more detailed results). The networks are enriched in synaptic components, predominantly in cell junction and plasma membrane regions (strongest enrichment for cell junction in PFC: \( p < 3.3 \times 10^{-8} \), and for synapse in NAcc: \( p < 1.3 \times 10^{-19} \)). Enrichment for biological processes showed the involvement of the network in neurodevelopmental processes including neuronal differentiation and development.
(PFC: \(p < 2.1 \times 10^{-4}\); NAcc: \(p < 1 \times 10^{-19}\)), neuron projection guidance (PFC: \(p < 1 \times 10^{-3}\); NAcc: \(p < 1 \times 10^{-16}\)) and regulation of trans-synaptic signaling (PFC: \(p < 2.9 \times 10^{-5}\); NAcc: \(p < 3.09 \times 10^{-17}\)). Enrichment for molecular functions showed a role of the networks in protein binding and cell adhesion (PFC: \(p < 2.72 \times 10^{-8}\); NAcc: \(p < 5.6 \times 10^{-6}\)). These functions are fundamental to the establishment of brain connectivity, mainly via axon guidance (e.g. [58]) and synaptogenesis (e.g. [59]).

**Developmental gene expression.** We assessed the enrichment of gene expression for each network across brain regions and developmental periods in humans, using the CSEA tool [60, 61].
Both networks are enriched across the brain during perinatal periods. Notably, the expression of the NAcc network in the PFC is enriched again during adolescence (Fig. 3D; \( p = 3.679 \times 10^{-4} \)), in line with previous descriptions of the developmental trajectory mediating adolescent corticolimbic maturation [17, 25].

To explore the ability of the PFC and NAcc networks, and ultimately the ePRS itself, to capture transcriptionally co-regulated biological processes, we analyzed the networks’ co-expression patterns in their corresponding brain regions during childhood and adulthood using the Brainspan dataset ([62]; Fig. 4). PFC. In the heatmap representing correlation of gene expression during childhood, there are 3 main clusters of high co-expression, but only 1 cluster is maintained in adulthood. Finding that correlations of expression of genes in the PFC in childhood are not maintained in adulthood is in line with the marked developmental changes in the PFC transcriptome landscape previously described in humans and mice [63]. NAcc. The heatmap for childhood gene co-expression shows a large main cluster, containing several highly correlated smaller clusters. Many of the smaller NAcc clusters perdure into adulthood, indicating that the NAcc network is more stable than the PFC network.

Comparison between polygenic scores

We compared our ePRS to a traditional PRS for ADHD on the capacity to predict the same behavioral outcomes. For that, we selected the top 4515 most significant SNPs identified in the latest ADHD GWAS [64], which corresponds to the GWAS \( p \)-value threshold 4.91e-5, and created a score comparable in size to the ePRS in terms of number of SNPs. There was no association between the PRS and the main outcomes for MAVAN (mean-corrected; \( \beta = -0.01, p = 0.48 \)), GUSTO (proportion of successful inhibitions: \( \beta = -0.009, p = 0.52 \); Fig. S3) or ALSPAC (mean reaction time – incorrect stop trials; \( \beta = -3.248, p = 0.46 \); Fig. S4) cohorts. We also performed an enrichment analysis to characterize the functional/biological properties of the PRS genes and found that they are upregulated across the brain, but not as selectively and to a lesser extent than the genes from the ePRS. Finally, results from the CSEA show no selective spatiotemporal enrichment in the human brain (Fig. S6).

DISCUSSION

Impulse control deficits are a common trait of numerous neurodevelopmental psychiatric disorders. Discovering their neurobiological underpinnings and early biomarkers will help identifying at risk individuals and improving/implementing early prevention and intervention strategies. Here, we generated an expression-based polygenic score (ePRS) consisting of SNPs within genes co-expressed with the axon guidance cue receptor gene, DCC, in the PFC and NAcc, to create a functional and corticolimbic-specific marker of vulnerability to heightened impulsivity. Our results show that the ePRS is significantly associated with different measures of impulsive behaviors in children from three ethnically diverse independent birth cohorts. Across all cohorts, the low-ePRS groups show higher impulsivity-related phenotypes. Detailed characterization of the gene networks comprising the corticolimbic DCC–ePRS show significant functional interactions, contribution to core neurodevelopmental processes, and enriched expression in cortical neurons, particularly from embryonic life to adolescence.

Most PRSs are characterized by a limited generalizability due to a marked disparity in prediction accuracy across different populations [65, 66]. This limitation, partially explained by the biased ancestry representation in most well-powered discovery GWASs, does not affect the ability of the ePRS to predict impulsive phenotypes across 3 independent birth cohorts from Canada, Singapore, and UK. Other studies that have implemented a similar approach to polygenic risk analysis by using the ePRS methodology reported a high predictive value of their genetic scores applied across diverse populations [42, 67]. To understand how the ePRS compares to a traditionally derived PRS, in this study we constructed a score based on the latest GWAS for ADHD and found that, even though the PRS score predicts impulsive behavior in one cohort (Fig. S3), our ePRS predicts a larger number of outcomes, across all three cohorts. This is consistent with other studies that have observed a higher prediction accuracy of their ePRS compared to conventional PRSs [34, 42]. Since the ePRS methodology relies on identifying tissue-specific gene networks and their function, instead of identifying scattered genetic variants across the genome, we are able to create a more biologically meaningful score compared to conventional PRSs. Finally, the associated phenotype to weigh the SNPs in our ePRS is gene expression, which, given the current state of technology, is a highly quantitative trait measurable with high precision, across different tissues and conditions, by high-throughput sequencing, and thus yielding a score that globally represents transcriptionally co-regulated biological processes. These results suggest that our genetic profiling approach increases the likelihood of identifying trait-relevant biological markers.

We identified co-expression networks for the guidance cue receptor, DCC, specifically in the NAcc and the PFC, to create a biological marker related to neurodevelopmental processes occurring in these regions, that could predict levels of impulsivity in children. In addition to establishing the ePRS’ predictive power of impulsivity across 3 different cohorts, we found the co-expression networks to be highly enriched for protein–protein interactions, suggesting their involvement in common biological functions. Since DCC receptors are master organizers of neuronal circuits, and since variations in its expression in early life result not only in functional and anatomical alterations of neural pathways involved in inhibitory control, but also in alterations of inhibitory control itself [17, 36, 51], it is not surprising that DCC co-expression networks in these corticolimbic hubs associate...
with behavioral traits implicated in psychopathology. Indeed, proper establishment of neuronal circuits is essential to mental health [68]. The genes that make up the networks are highly upregulated in the PFC and the NAcc and are involved in a wide range of neurodevelopmental processes. This enrichment suggests a prominent role of the gene networks in the maturation of both PFC and NAcc circuits, validating the use of these networks as the basis for the ePRS calculation and their potential use as a functional biomarker to predict reflection and motor impulsivity in children.

Results from several studies in humans show that mutations in the DCC gene lead to dramatic neurodevelopmental changes, including agenesis of the corpus callosum [36, 69, 70], developmental split-brain syndrome [69], and congenital mirror movements [37, 70, 71]. Similar noticeable changes have been described in DCC homozygous or haploinsufficient mice [72], highlighting the core role of DCC in neurodevelopmental wiring. As DCC expression shifts from high to low in adolescence, its functional role also shifts from broad organization of developing neuronal networks to the refinement of neuronal architecture, synaptogenesis and synaptic plasticity of established matured circuits [30, 31, 73]. Recent human studies have also shown that many polymorphisms in DCC, as well as altered levels of gene expression, are related to numerous neuropsychiatric conditions of developmental onset, some of which are characterized by deficits in PFC function and impulse control [30, 31]. Individual genes do not operate in isolation and cannot explain the entire spectrum of mental disorders, as it has been well established by a wealth of data from recent GWAS studies showing massive polygenicity among neuropsychiatric disorders. Therefore, DCC receptors act as a master organizer of specific synaptic circuits, as a part of a gene network, and we have shown that a PFC gene network for DCC is associated with overall brain structure [34]. Our functional analyses of the corticolimbic DCC gene networks suggest their implication in the development of the neural substrates underlying inhibitory control behaviors. The genes that comprise the networks are co-expressed in crucial brain regions (see Fig. 3B, 4), suggesting their spatial convergence.

Furthermore, the expression of genes known to increase risk for neuropsychiatric disorders converge temporally, especially before and during the onset of the disorder [68]. Here we observed that gene expression for both networks is enriched during specific pre- and post-natal periods, including an enriched expression of the NAcc network in cortical neurons during late childhood and adolescence. As noted previously, the neurodevelopmental role that DCC plays changes as a function of developmental stage, and the fact that a DCC co-expression network is enriched again during late childhood and adolescence suggest that alterations in its function/expression can impact the adolescent development of synaptic connectivity and function in the PFC later in life.

We propose a novel type of marker for impulsivity-related phenotypes in children. Our biologically-informed approach to polygenic risk analysis aims to capture variation in the function/expression of gene networks predominantly associated with PFC and NAcc maturation, two regions subserving inhibitory control. Whether integrating relevant SNPs associated with other forms of gene expression regulation beyond cis (e.g., transcription factors, promoter regions, and chromatin modifications) in non-coding regions changes the performance of the scores, will be investigated in future studies. Exploring the association between the ePRS and inhibitory control behaviors later in life is needed in order to investigate the possible use of this genetic marker as a probabilistic risk score for vulnerability phenotypes linked to psychopathologies of adolescent onset. Our results are an example of the utility of understanding the molecular processes that govern the development of a neural circuit, and how this knowledge can be applied to predict genetic susceptibility to endophenotypes linked to psychiatric conditions.

**DATA AVAILABILITY**

Data from MAVAN can be made available via reasonable request to the corresponding authors. For GUSTO, visit [https://www.gusto.sg/](https://www.gusto.sg/). For ALSPAC, data can be purchased; the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool at [http://www.bristol.ac.uk/alspac/researchers/our-data/](http://www.bristol.ac.uk/alspac/researchers/our-data/).
REFERENCES

1. Somerville LH, Casey BJ. Developmental neurobiology of cognitive control and motivational systems. Curr Opin Neurobiol. 2010;20:236–41.

2. Jentsch JD, Ashenhurst JR, Cervantes MC, Gronman SM, James AS, Pennington JT. Dissecting impulsivity and its relationships to drug addictions. Ann N Y Acad Sci. 2014;1327:1–26.

3. Luna B, Sweeney JA. The emergence of collaborative brain function: fMRI studies of the development of response inhibition. Ann N Y Acad Sci. 2004;1021:296–309.

4. Petrovic P, Castellanos FX. Top-down dysregulation—from ADHD to emotional instability. Front Behav Neurosci. 2016;10:1–25.

5. Casey BJ, Getz S, Galvan A. The adolescent brain. Dev Rev. 2008;28:662–77.

6. Dalley JW, Everitt BJ, Robbins TW. Impulsivity, Compulsivity, and Top-Down Cognitive Control. Vol. 69. Neuron. Cell Press; 2011. p. 680–94.

7. Galvan A, Hare T, Voss H, Glover G, Casey BJ. Risk-taking and the adolescent brain: Who is at risk? Vol. 10. Developmental Science. John Wiley & Sons, Ltd; 2007. p. 8–14.

8. Luna B, Padmanabhan A, O’Hearn K. What has fMRI told us about the Development of Cognitive Control through Adolescence? Brain Cogn. 2010;72:101–13.

9. Spear LP. The adolescent brain and age-related behavioral manifestations. Neurosci Biobehav Rev. 2000;24:47–63.

10. Caballero A, Granberg R, Tseng KY. Mechanisms contributing to prefrontal cortex maturation during adolescence. Neurosci Biobehav Rev. 2016;70:4–12.

11. Reynolds LM, Pokinko M, Torres-Berrío A, Cuesta S, Lambert LC, Del Cid Pellitero E, et al. DCC receptors drive prefrontal cortex maturation by determining netrin-1 homeostasis: role of mesocorticillobic connectivity and volumetric alterations in dcc mutation carriers. J Neurosci. 2018;38:4655–65.

12. Hoops D, Flores C. Making dopamine connections in adolescence. Trends Neurosci. 2015;38:603–10.

13. Soh S-E, Tint MT, Gluckman PD, Godfrey KM, Rifkin-Graboi A, Chan YH, et al. Interpreting psychiatric genome-wide association studies with multispecies Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science (80-) 2015;348:648–60.

14. Hari Dass SA, McCracken K, Pokhvisneva I, Chen LM, Garg E, Nguyen TTT, et al. A biologically-informed polygenic score identifies endophenotypes and clinical conditions associated with the insulin receptor function on specific brain regions. ElBioMedicine. 2021;89:20–31.

15. Silveira PP, Pokhvisneva I, Parent C, Cai S, Rema ASS, Broekman BFP, et al. A biologically-informed polygenic score identifies endophenotypes and clinical conditions associated with the insulin receptor function on specific brain regions. ElBioMedicine. 2021;89:20–31.

16. J.M. Restrepo-Lozano et al.
Cuesta S, Restrepo-Lozano JM, Popescu C, He S, Reynolds LM, Israel S, et al. DCC-related developmental effects of abused- versus therapeutic-like amphetamine doses in adolescence. Addict Biol. 2020;25:1–11.

Urbanek S, Bibiko H-J, Stefano ML. R: a language and environment for statistical computing. The R Foundation for Statistical Computing. 2014.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47:D607–13.

Demchak B, Hull T, Reich M, Liefeld T, Smoot M, Ideker T, et al. Cytoscape: the network visualization tool for GenomeSpace workflows. F1000Res. 2014;2014:1–12.

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: W214–20.

Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.

Ebens A, Brose K, Leonardo ED, Hansson MG, Bladt F, Birchmeier C, et al. Heparan sulfate growth factor/scatter factor is an axonal chemotactant and a neurotrophic factor for spinal motor neurons. Neuron. 1996;17:1157–72.

Schmeisser MJ, Baumann B, Johannsen S, Vindedal GF, Jensen V, Hvalby OC, et al. Analytical approaches to RNA computing. The R Foundation for Statistical Computing. 2014.

Finger JH, Bronson RT, Harris B, Johnson K, Przyborski SA, Ackerman SL. The guidance of corticospinal tract axons. J Neurosci. 2002;22:10346–53.

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;51:606–12.

Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, et al. The online version contains supplementary material. This research was specifically funded by the National Institute on Drug Abuse (CF Grant number: R01DA037911), the Canadian Institutes of Health Research (CF Grant Number: MOP-74709; PPS Grant Number: PJT-166066), the Natural Science and Engineering Research Council of Canada (CF Grant Number: 29822268; PPS Grant Number: RGPN-2018-05063), a Doctoral Training fellowship from the Fonds de Recherche du Québec en Santé to JMRL, and a Healthy Brains for Healthy Lives Graduate Student Fellowship to JMRL. This publication is the work of the authors and Cecilia Flores and Patricia Peloza Silveira will serve as guarantors for the contents of this paper.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-022-01533-7.

Correspondence and requests for materials should be addressed to Patricia P. Silveira or Cecilia Flores.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ACKNOWLEDGEMENTS
We are grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole team for all cohorts: MAVAN, GUSTO, and ALSPAC, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and Wellcome (Grant ref. 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is available on the ALSPAC website: http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf.

AUTHOR CONTRIBUTIONS
JMRL, PPS and CF conceived and designed the project. MJM provided scientific advice and the high-performance computational infrastructure. JMRL and SP calculated the polygenic scores. JMRL, ZW and IP performed the linear regression analyses. JMRL performed the enrichment analyses. JMRL, PPS, and CF wrote the manuscript. JMRL, IP, PPS, and CF discussed the results, and reviewed and edited the final manuscript. Figures were created using BioRender.com.

FUNDING
This research was specifically funded by the National Institute on Drug Abuse (CF Grant number: R01DA037911), the Canadian Institutes of Health Research (CF Grant Number: MOP-74709; PPS Grant Number: PJT-166066), the Natural Science and Engineering Research Council of Canada (CF Grant Number: 29822268; PPS Grant Number: RGPN-2018-05063), a Doctoral Training fellowship from the Fonds de Recherche du Québec en Santé to JMRL, and a Healthy Brains for Healthy Lives Graduate Student Fellowship to JMRL. This publication is the work of the authors and Cecilia Flores and Patricia Peloza Silveira will serve as guarantors for the contents of this paper.

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022