

# Genetic overlap between educational attainment, schizophrenia and autism

Varun Warri<sup>1</sup>, Richard AI Bethlehem<sup>1</sup>, Daniel H Geschwind<sup>2</sup>,  
and Simon Baron-Cohen<sup>1,3</sup>

1. Autism Research Centre, Department of Psychiatry, University of Cambridge 18B  
Trumpington Road, Douglas House, Cambridge CB2 8AH, UK

2. Program in Neurogenetics, Department of Neurology, Center for Autism Research and  
Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los  
Angeles, Los Angeles, CA 90095, USA

3. CLASS Clinic, Cambridgeshire and Peterborough NHS Foundation Trust (CPFT),  
Cambridgeshire, UK

Corresponding author: Varun Warri ([vw260@medschl.cam.ac.uk](mailto:vw260@medschl.cam.ac.uk)) and Simon Baron-Cohen  
([sb205@cam.ac.uk](mailto:sb205@cam.ac.uk))

**Acknowledgments:** VW was funded by the Cambridge Trusts and St. John's College, Cambridge. RB was funded by The Pinsent Darwin Trust and the MRC fellowship. This work was funded by the Autism Research Trust and the Templeton World Charity Foundation, Inc. The research was carried out in association with the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care East of England at Cambridgeshire and Peterborough NHS Foundation Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

## Key Points

**Question:** How do genes for educational attainment interact with risk genes for autism and schizophrenia?

**Findings:** We show that genes for educational attainment (edu genes) are significantly likely to be mutated in autism and intellectual disability. We further show that edu genes also interact with co-expression modules that are associated with autism or schizophrenia and are enriched for differentially expressed genes in autism or schizophrenia. Finally, we identify that the enrichment between risk genes for autism and schizophrenia and human accelerated regions are driven, in part, by their overlap with edu genes.

**Meaning:** Edu genes interact with schizophrenia and autism risk genes in specific pathways, contributing to both cognitive deficits and talents.

## Abstract

**Importance:** The genetic relationship between cognition, autism, and schizophrenia is complex. It is unclear how genes that contribute to cognition also contribute to risk for autism and schizophrenia.

**Objective:** To investigate the interaction between genes related to cognition (measured via proxy through educational attainment, which we call ‘edu genes’) and genes/biological pathways that are atypical in autism and schizophrenia.

**Design:** Genetic correlation and enrichment analysis were conducted to identify the interaction between edu genes and risk genes and biological pathways for autism or schizophrenia.

**Results:** First, edu genes are enriched in a specific developmental co-expression module that is also enriched for high confidence autism risk genes. Second, modules enriched for genes that are dysregulated in autism and schizophrenia are also enriched for edu genes. Finally, genes that overlap between the two above modules and educational attainment are significantly enriched for genes that flank human accelerated regions, suggesting increased positive selection for the overlapping gene sets.

**Conclusion:** Our results identify distinct co-expression modules where risk genes for the two psychiatric conditions interact with edu genes. This suggests specific pathways that contribute to both cognitive deficits and cognitive talents, in individuals with schizophrenia or autism.

## Introduction

Autism and schizophrenia are complex psychiatric conditions with considerable phenotypic heterogeneity and significant heritability. Both conditions are genetically highly heterogeneous, with common and rare variants implicated in their aetiology. Approximately 40% of individuals with autism have intellectual disability defined as  $IQ < 70$ <sup>1</sup> and as a group, individuals with autism show deficits in theory of mind<sup>2</sup> and central coherence<sup>3</sup>. In contrast, some studies show superior functioning in domain specific aspects of cognition in individuals with autism, including superior attention to detail<sup>4</sup>, stronger systemizing<sup>5,6</sup>, hypercalculia<sup>7,8</sup> and savantism<sup>9</sup>. Furthermore, elevated rates of autism and autistic traits have been reported in those working in science, technology, engineering and maths (STEM) fields<sup>10,11</sup>. The relationship between IQ/cognitive ability and schizophrenia is also complex, but in a different way. Cognitive difficulties are also observed in individuals with schizophrenia<sup>12</sup>. Other studies have identified a link between schizophrenia and creativity<sup>13</sup>, with polygenic risk scores for schizophrenia predicting creativity. Clearly, both autism and schizophrenia have a complex genetic link with cognition and IQ. This study aims to understand this set of genetic relationships.

There is an established positive genetic correlation between autism and different measures of cognitive attainment such as years of education<sup>14</sup> and a derived measure of general cognition in children<sup>15,16</sup>. Polygenic risk score analysis has identified significant risk between general cognitive ability and autism<sup>17</sup> and creativity and schizophrenia<sup>13</sup>. In addition, conditional False Discovery Rate based analysis using genome-wide association study (GWAS) for educational attainment has identified common variants associated with schizophrenia<sup>18</sup>. On the other hand, individuals with autism who are co-morbid for intellectual disability (ID) have a significantly higher rate of de novo variants than individuals with autism who have average or above average IQ<sup>19</sup>.

It is unclear, at a genetic and transcriptional level, how exactly genes for educational attainment interact with genes that contribute to risk in autism and schizophrenia. Two broad, non-mutually exclusive hypotheses might elucidate the complex relationship between cognition/educational attainment and the two psychiatric conditions at a biological level. First, mutations in genes related to educational attainment (what we call 'edu genes') might lead to autism and schizophrenia co-morbid with intellectual disability and other cognitive deficits. In other words, there may be an upstream convergence in shared genetic mutations.

Second, edu genes might interact with risk pathways for schizophrenia and autism in specific transcriptional co-expression modules, leading to an interaction in common downstream biological pathways. This could lead to increased risk for the psychiatric conditions with preserved or above average cognitive ability.

A few studies have identified a role for human accelerated regions (HARs) in autism<sup>20</sup>, schizophrenia<sup>21</sup>, and cognition<sup>20</sup>. HARs are conserved regions of the genome that show human-specific divergence at an accelerated pace, thought to indicate positive selections in humans. Genes in these regions may thus have contributed to human-specific traits. Indeed, mutations in genes that lie in HARs affect cognition and contribute to autism<sup>20</sup>. Furthermore, schizophrenia-associated loci are enriched in genes that lie in or close to HARs<sup>21</sup>. It is unclear why HARs are enriched for genes that contribute to both these conditions. One explanation is that genes that lie in or near HAR contribute pleiotropically to both cognition and risk for schizophrenia or autism. This might explain the positive selection in these regions in spite of their association with autism and schizophrenia, two conditions with reduced fecundity<sup>22</sup>. Thus, we hypothesize that genes flanking or interacting with HARs may contribute to the genetic and phenotypic overlap seen between cognition and the two psychiatric conditions.

In this study, we examine the relationship between edu genes and genes and transcriptional co-expression modules implicated in autism or schizophrenia. We identify a specific gene co-expression module in the developmental transcriptome of the human cortex that is enriched for edu genes and genes that are frequently mutated in autism. We further identify co-expression networks that are enriched for differentially expressed genes in schizophrenia and autism and are also enriched for edu genes. Finally, we also demonstrate that the overlapping genes between edu attainment and the co-expression modules enriched for differentially expressed genes for autism or schizophrenia are also enriched for HARs. A schematic diagram of the study design is provided in Figure 1.

*Insert Figure 1 here*

## Methods

### Genes

We used a list of genes previously identified using DEPICT<sup>23</sup> for educational attainment<sup>24</sup> (edu genes). These genes were identified using independent GWAS loci with  $P < 1 \times 10^{-5}$ . SNPs were defined as being independent if they had an  $r^2 < 0.1$ , and if they were 500kb apart. The genes were identified using a sample of 405,072 participants as previously described by Okbay et al., (2016)<sup>24</sup>. As a control, we also used genes previously identified using DEPICT for height (641 genes)<sup>25</sup> identified using independent GWAS loci with  $P < 5 \times 10^{-8}$  (height genes). DEPICT uses data gathered from 77,870 microarray experiments expressing levels of 19,997 genes covered by the microarray platforms. To investigate enrichment in autism and intellectual disability, a list of genes implicated in autism (henceforth, high confidence autism genes) was obtained from Sanders et al (2015)<sup>26</sup> and a list of genes implicated in intellectual disability (henceforth, ID genes) from Vissers et al, (2016)<sup>27</sup>.

### *Brain-expressed genes*

We identified a list of genes expressed in the brain using expression data from the Gene Tissue Expression portal (GTEx v6)<sup>28</sup>. We downloaded the list of median gene expression per tissue and identified a list of genes expressed in any of the 11 brain tissues (amygdala, frontal cortex, cortex, cerebellar cortex, cerebellum, anterior cingulate cortex, caudate, hippocampus, hypothalamus, nucleus accumbens and putamen). Genes expressed with a median RPKM  $> 1$  in any of the 11 tissues were considered brain-expressed genes. We identified 15,395 brain-expressed genes.

### *Human accelerated region genes*

We identified a pre-compiled list of genes found in the human accelerated regions (HAR) using data in multiple different publications<sup>20</sup>. We used the list of genes closest to HARs (HAR genes), and a separate non-independent list of genes that interact with HARs, identified through HiC and Chia-PET experiments (HAR interacting genes)<sup>20</sup>. There is a considerable overlap between the two gene lists. This is not surprising as, according to Doan et al., (2016)<sup>20</sup>, more than 42% of the HARs interact with the flanking promoters. While HAR interacting genes provide evidence for the genes that are influenced by HAR when

compared to HAR genes, we consider this list to be preliminary as interaction between HARs have not been investigated in multiple tissues and across multiple time points.

## Co-expression modules

### *Cognition*

To test the enrichment of edu genes in networks associated with general cognition, we used two previously identified modules for cognition<sup>29</sup>: M1 (renamed M<sub>c</sub>1) and M3 (renamed M<sub>c</sub>3). Initial weighted gene co-expression networks were constructed using 122 frozen hippocampus samples that were surgically removed from patients with temporal lobe epilepsy. Significant modules within these weighted networks were identified (empirical  $P \leq 0.002$ ) after ensuring that co-expression patterns were unrelated to epilepsy, co-expression modules were preserved in rodent hippocampus and preserved across different human neural tissues by comparing the average topological overlaps of the of the genes in each module between different tissues.

### *Autism, schizophrenia, and the developmental transcriptome*

We used previously identified gene co-expression modules for both autism<sup>30,31</sup> and schizophrenia<sup>32</sup>. Due to the relatively small sample sizes of existing transcriptome projects, we used all nominally significant differentially co-expressed modules for both autism and schizophrenia compared to controls. For autism, we used modules mod1 (renamed M<sub>A</sub>1), mod5 (renamed M<sub>A</sub>5), mod6 (renamed M<sub>A</sub>6) from Gupta et al., (2014)<sup>30</sup>. Briefly, RNA sequencing was performed for multiple cortical tissues obtained from the Autism Tissue Programme. In total, there were 47 (32 unique individuals) autism samples, and 57 (40 unique individuals) control samples. Weighted gene co-expression network analyses (WGCNA) were performed for 13,443 genes combining both cases and control samples. Modules differentially co-expressed for autism were identified using a linear mixed regression framework for case-control status. As a validation, we also used three gene co-expression modules that are associated with autism and are enriched for significantly downregulated genes in the autism post-mortem cortex, identified using a larger dataset<sup>31</sup>. These modules are: M4 (renamed CTX-M<sub>A</sub>4), M10 (renamed CTX-M<sub>A</sub>10), and M16 (renamed CTX-M<sub>A</sub>16). All three modules significantly overlap with the M<sub>A</sub>1 module<sup>31</sup>. Briefly, rRNA-depleted RNA sequencing and WGCNA was performed using 82 cases and 81

control cortex samples from 47 individuals with autism and 44 controls. Co-expression networks were constructed from a total of 16398 genes and lncRNAs<sup>31</sup>. We also used the list of differentially expressed genes in the autism cortex identified after FDR correction from Parikshak et al., (2016)<sup>31</sup>.

For schizophrenia we used data from Fromer et al., (2016)<sup>32</sup>. RNA was extracted and sequenced from the dorsolateral prefrontal cortex from 258 schizophrenia cases and 279 control subjects from the University of Pittsburgh, University of Pennsylvania, and Mount Sinai Hospital brain banks. Co-expression networks were constructed for the control samples separately with a total of 16,423 genes using WGCNA. Modules implicated in schizophrenia were defined as those that harboured an excess of differentially expressed genes for schizophrenia. Four control modules were significantly enriched for differentially expressed schizophrenia genes: M2C (renamed M<sub>S</sub>2), M9C (renamed M<sub>S</sub>9), M11C (renamed M<sub>S</sub>11), and M13C (renamed M<sub>S</sub>13). Three additional control modules were nominally significant for enrichment of differentially expressed schizophrenia genes: M14C (renamed M<sub>S</sub>14), M27C (renamed M<sub>S</sub>27) and M30C (renamed M<sub>S</sub>13). We also used a list of differentially expressed genes in the schizophrenia cortex identified after FDR correction from Fromer et al., (2016)<sup>32</sup>.

We tested all three modules in the autism discovery and validation datasets and all seven modules in the schizophrenia dataset for enrichment of genes associated with educational attainment. In addition, we also utilized co-expressed modules identified using gene-expression datasets of the human developmental transcriptome from Parikshak et al., (2013)<sup>33</sup>. These modules were identified using spatio-temporal expression data from the BrainSpan dataset<sup>34</sup> and provide insight into the temporal window of this potential modular gene enrichment. We did not include Module 7 in the analyses, as the genes in the module were not highly co-expressed with each other<sup>33</sup>.

## Statistical analyses

Genetic correlation was performed using LDSC<sup>14,35</sup>. We used summary GWAS data for educational attainment<sup>24</sup> (n = 328,917) downloaded from the Social Sciences Genetic Association Consortium website (<http://www.thessgac.org/#!data/kuzq8>). GWAS data for



schizophrenia ( $n = 79,845$ )<sup>36</sup> and autism ( $n=10,263$ ) were downloaded from Psychiatric Genomics Consortium website (<https://www.med.unc.edu/pgc/results-and-downloads>). As a control dataset GWAS data for height ( $n = 253,288$ )<sup>25</sup> was downloaded from the GIANT website ([http://www.broadinstitute.org/collaboration/giant/index.php/GIANT\\_consortium\\_data\\_files](http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files)). Probability of Loss of Function Intolerance score (pLI) was identified for each gene from the Exome Aggregation Consortium web browser (<http://exac.broadinstitute.org/>)<sup>37</sup>.

Enrichment analyses were performed using hypergeometric tests. For each enrichment pair, the background gene list was identified as the set of total common genes between the two pairs. For example, DEPICT uses expression data from 19,997 genes from 77,840 microarray experiments as the initial input to identify enriched genes and gene-sets<sup>23</sup>. A total of 13,262 genes were used to construct a weighted co-expression network for autism genes<sup>30</sup>, of which 11,963 genes were common between the DEPICT gene list and the autism co-expression gene list. A total of 16,423 genes were used for network construction for schizophrenia<sup>32</sup>, and, in total, there were 14,112 genes common between the DEPICT gene list and the schizophrenia co-expression gene list. For the enrichment with high confidence autism genes and ID, we used a list of all genes in DEPICT. For enrichment with modules identified with cognition, we used a list of genes common to DEPICT and all genes used in network construction using WGCNA, which was a total of 10,727 genes. Further, for enrichment using only brain expressed genes, a total of 12,625 genes were common between the DEPICT gene list and genes expressed in the brain using the GTEx dataset.

All enrichment analyses were performed using Ensembl gene IDs. A full list of genes used for enrichment analysis is provided in the Supplementary File. We report significant enrichments of all hypergeometric tests with FDR adjusted P-values below 0.05. All hypergeometric tests report P-values for probabilities of drawing  $x$  or more (or less in case of an underlap) successes, where  $x$  is the number of overlapping genes between the two gene sets tested. Hypergeometric test P-values and fold difference (or fold enrichment/fold change, defined as the ratio of observed to expected overlap) were calculated using: [http://nemates.org/MA/progs/overlap\\_stats.html](http://nemates.org/MA/progs/overlap_stats.html). To test the specificity of all significant enrichments between educational attainment and the trait of interest, we used height as a negative control.

To identify multi-set enrichment we used the R package ‘SuperExactTest’<sup>38</sup>. We computed 1-tailed P-values with the hypothesis that we expect to find a significant overlap between the different sets. For each of the multi-set enrichment analyses, we identified background genes as those that were investigated across all different sets. For example, for the 3-way enrichment between the M<sub>A</sub>1 autism co-expression module, edu genes and HAR genes, the background list of genes was the intersection of all genes identified in the transcriptional analyses of the autism post-mortem tissues, and all the genes covered in DEPICT. Further, we restricted our HAR genes to only those that were covered in the DEPICT and identified in the autism transcriptome.

## Results

### Genetic correlation analyses

Previous studies have identified genetic correlations between educational attainment and autism, but not with schizophrenia. Here, we used the recent, larger educational attainment dataset to conduct genetic correlations between both educational attainment and both schizophrenia and autism. There was a modest but significant positive genetic correlation between autism and educational attainment ( $r_g = 0.28 \pm 0.04$ ;  $P=2.4 \times 10^{-11}$ ) (Figure 2). This correlation is considerably lower than the correlation between childhood IQ and autism ( $r_g = 0.5 \pm 0.11$ ), but similar to the previously calculated correlation between autism and college years ( $r_g = 0.3 \pm 0.08$ ;  $P = 1.5 \times 10^{-4}$ )<sup>14,15</sup>. In addition, there was a small but significant positive correlation between schizophrenia and educational attainment ( $r_g = 0.096 \pm 0.02$ ;  $P = 5.39 \times 10^{-6}$ ) (Figure 2a). Furthermore, we did not identify significant genetic correlations between height and schizophrenia ( $0.0011 \pm 0.01$ ;  $P = 0.952$ ) or height and autism ( $r_g = -0.02 \pm 0.004$ ;  $P = 0.47$ ), suggesting height to be an appropriate negative control.

### Genes for education (edu genes) are highly resistant to loss of function

Using data from the Exome Aggregation Consortium (ExAC), we calculated the probability of loss of function intolerance (pLI) for all 146 genes identified for educational attainment. pLI scores were available for 141 of the 146 genes. A pLI  $\Rightarrow$  0.9 is considered to be highly intolerant to loss of function, and is likely to be critical to foetal survival and

development. Of the remaining 141 genes, 89 had a pLI  $\Rightarrow$  0.9 (Supplementary Data) that is significant ( $P = 0.0012$ ; one-sided binomial sign test). In contrast, there were only 19 LoF tolerant genes, suggesting a significant depletion of LoF tolerant genes to educational attainment ( $P < 0.0001$ , one-sided binomial sign test). However, brain genes are likely to be more intolerant to other genes. We thus compared the distribution of pLI scores between genes for edu attainment with pLI scores for all brain genes. Edu genes in the brain had higher mean and median pLI scores compared to all brain expressed genes (edu genes vs brain genes; mean: 0.36 and 0.74 respectively, median: 0.08 and 0.98 respectively). However, analysis of the frequency histograms of the pLI scores for both gene sets revealed a non-normal distribution with heavy tails. Wilcoxon-rank test identified significantly higher pLI scores for the brain edu genes compared to non-edu brain genes ( $P < 2.2 \times 10^{-16}$ ) (Figure 2b). These results suggest that the edu genes, more specifically than simply being brain expressed genes, are unlikely to harbour severe mutations and are likely to be embryonically lethal.

### **Edu genes are enriched in a module enriched for general cognitive function**

Educational attainment and general cognitive function share a significant, albeit modest, correlation (Proxy correlation: Years of education and Childhood IQ:  $rg = 0.731 \pm 0.08$ ;  $P = 2.18 \times 10^{-17}$ )<sup>15</sup>. Genetic correlation between educational attainment and specific aspects of cognitive ability ranged from  $0.73 \pm 0.04$  (for numerical reasoning) to  $0.22 \pm 0.05$  (for reaction times)<sup>39</sup>. A recent study identified two neural modules ( $M_c1$  and  $M_c3$ ) that were enriched for genes implicated in various measures of general cognition<sup>29</sup>. We tested if the edu genes are enriched in the two cognition modules. While we did not identify a significant enrichment on  $M_c1$ , we did identify a significant enrichment in  $M_c3$  (Fold difference = 3.2;  $P_{FDRadjusted} = 0.024$ ). We did not identify a significant enrichment for height genes in either of the modules (Supplementary Data).  $M_c1$  is highly expressed across the cortex, and has a high expression in fetal brain starting at early mid-fetal development ( $16 \leq PCW \leq 19$ ), and increases at birth with the expression persisting postnatally.

### **Edu genes are enriched in modules implicated in schizophrenia and autism**

We subsequently investigated if these genes have a more downstream effect on schizophrenia and autism, affecting common biological pathways and systems. To investigate

this, we used previously identified co-expression modules from adult brain tissues that are enriched for genes that are differentially expressed in autism (three modules:  $M_{A1}$ ,  $M_{A5}$ , and  $M_{A6}$ ) and schizophrenia (seven modules;  $M_{S2}$ ,  $M_{S9}$ ,  $M_{S11}$ ,  $M_{S13}$ ,  $M_{S14}$ ,  $M_{S27}$ ,  $M_{S30}$ ) (see methods). We identified a significant enrichment for educational attainment genes in  $M_{S2}$  (fold difference = 2.9;  $P_{FDRadjusted} = 8.6 \times 10^{-8}$ ). In total, there were 34 edu genes that overlapped with the  $M_{S2}$  module. In contrast, for the same module, there was no enrichment for genes associated with height (fold difference = 1.1;  $P_{FDRadjusted} = 0.31$ ). Notably, the  $M_{S2}$  module comprises 119 genes that are differentially expressed in schizophrenic post-mortem brain tissue compared to controls. Of this, 117 genes are upregulated in schizophrenia compared to controls. The  $M_{S2}$  module is most significant enriched for the gene ontology (GO) term ‘synaptic transmission’. We also identified a modest enrichment for edu genes in the autism  $M_{A1}$  module (25 overlapping genes, fold difference = 1.5;  $P_{FDRadjusted} = 0.049$ ). The  $M_{A1}$  module is also enriched for synaptic genes, but is downregulated in autism compared to controls. Again, this enrichment was not significant for height. The  $M_{A1}$  module is enriched for genes that contain the GO term ‘synaptic transmission’. To confirm the analysis, we investigated if there is enrichment with Edu genes in three autism co-expression cortical modules identified in a different dataset but that overlaps with the  $M_{A1}$  modules (CTX- $M_{A4}$ , CTX- $M_{A10}$ , and CTX- $M_{A16}$ ). All three modules are enriched for genes that are significantly downregulated in autism. We identified a similar modest enrichment of the edu genes in the CTX- $M_{A4}$  (fold difference = 2.61;  $P_{FDRadjusted} = 0.033$ ) and CTX- $M_{A10}$  modules (fold difference = 2.49;  $P_{FDRadjusted} = 0.033$ ) and the CTX- $M_{A16}$  (fold difference = 2.57;  $P_{FDRadjusted} = 0.033$ ) modules. The CTX- $M_{A4}$  and the CTX- $M_{A16}$  module are also enriched for the GO term ‘synaptic transmission’. Taken together, these results suggest that edu genes are enriched in downregulated autism modules that are primarily involved in synaptic transmission.

### **Edu genes are enriched for genes implicated in autism and intellectual disability**

We next investigated if high confidence autism genes are enriched in educational attainment. We used a list of 65 genes that have been identified using exome and whole genome-sequencing at  $FDR < 0.1^{26}$ , but restricted it to genes that are both expressed in the brain and are captured in the DEPICT gene list. Enrichment analyses identified 5 common genes between the two datasets, which was higher than expected (fold difference = 8.3;

$P_{\text{FDRadjusted}} = 3.2 \times 10^{-4}$ ; Supplementary Data). As a negative control, we tested for enrichment in genes identified for height using DEPICT. We did not identify a significant enrichment (fold difference = 1.7,  $P = 0.22$ ; Figure 3d). We also investigated if edu genes also contribute intellectual disability, by testing for enrichment using ID. Again, we restricted our analyses only to brain-expressed genes and genes covered in the DEPICT gene list. We identified a significant overlap between intellectual disability genes and edu genes (fold difference = 2;  $P = 0.013$ ). Interestingly, there was a significant overlap between height and ID genes, though the enrichment was smaller (fold difference = 1.3;  $P = 0.047$ ). Several genes that contribute to intellectual disability also cause significant changes in height. For example, syndromes like Fragile X and Prader Willi influence both height and intellectual ability<sup>40</sup>.

*Insert Figure 3 here*

### **Edu genes are enriched in one module in the developmental transcriptome**

We next asked if edu genes are significantly enriched in developmental modules. We utilized modules identified by Parikshak and colleagues (2013)<sup>33</sup> using the developmental transcriptome data from BrainSpan. We identified a significant enrichment for edu genes in the M2 module (Fold difference = 2.5;  $P_{\text{FDRadjusted}} = 0.01$ ). Interestingly, this module has been reported to be enriched for rare de novo variants in autism, and genes that are implicated in intellectual disability. However, since the Parikshak et al., (2013) paper was published, additional genes have been identified for both autism and ID. We further tested if the M2 module overlapped with high confidence autism genes and ID genes. Indeed this module had significant enrichment for autism genes, (Fold difference = 4.2;  $P_{\text{FDRadjusted}} = 9.36 \times 10^{-5}$ ), but only nominally significant for the expanded ID genes (Fold difference = 1.3;  $P_{\text{FDRadjusted}} = 0.057$ ), supporting the predictive power of these ASD-related gene networks. There was a significant three-way enrichment between genes in the M2 module, high confidence autism genes and edu genes (Fold difference = 122.8;  $P = 2.2 \times 10^{-6}$ ). Three of the 5 overlapping genes between edu genes and the 65 high confidence autism genes lie in the M2 module. This module was not significantly enriched for genes that are differentially expressed in autism brain<sup>31</sup> (Fold difference = 0.37,  $P = 1$ ) or genes that are differentially expressed in schizophrenia<sup>32</sup> (Fold difference = 0.6,  $P_{\text{FDRadjusted}} = 0.02$ ).

## Human accelerated region genes are enriched in overlapping gene sets

We reasoned that the genes that overlap between educational attainment and the significant co-expression modules for the two psychiatric conditions might be under stronger selection pressure in humans. To investigate this, we used lists of genes that flank and interact with human accelerated regions (HARs)<sup>20</sup>. HARs are regions of the genes that are conserved in other mammals but are different in humans after they diverged from chimpanzees. These regions are thought to contribute to human specific traits including higher-order cognitive function. We identified a list of genes that interact closely with HARs. To avoid potential confounds, we focused only on brain-expressed genes for all datasets. We conducted a three way-enrichment analysis between brain-expressed Edu genes, brain-expressed schizophrenia M<sub>S</sub>2 genes, and brain expressed HAR genes. We identified a significant overlap between the three sets of genes (Fold difference = 12.31; P = 1.20x10<sup>-11</sup>). This was not significant for height (Fold difference = 1.9; P = 0.05). The overlap between Edu genes and HAR genes was higher (Fold difference = 3.01; P = 8.9x10<sup>-11</sup>) than between M<sub>S</sub>2 and HAR genes (Fold difference = 1.8; P = 1x10<sup>-18</sup>). In addition, the genes overlapping between educational attainment and the M<sub>S</sub>2 module were significantly more enriched for HAR genes than the non-overlapping M<sub>S</sub>2 genes (OR = 5.4; P = 1.3x10<sup>-7</sup>). However, we did not identify a significant enrichment for HAR genes in the overlapping gene set compared to the non-overlapping edu gene set (OR = 2.1; P = 0.07).

We next investigated the three-way overlap between the M<sub>A</sub>1 module for autism, edu genes and HAR genes. Using only genes expressed in the brain in all three datasets, we identified a significant enrichment between HAR genes, genes in the M<sub>A</sub>1 module, and edu genes (Fold difference = 6.3, P = 4.5x10<sup>-6</sup>). Again, there was no significant enrichment for height (Fold difference = 0.85; P = 0.69). We were able to validate this using the CTX-M<sub>A</sub>4 module (Fold difference = 13.2, P<sub>FDRadjusted</sub> = 4.89x10<sup>-3</sup>) and the CTX-M<sub>A</sub>10 module (Fold difference = 8.5, P<sub>FDRadjusted</sub> = 0.036). However, we did not identify a significant enrichment for the CTX-M<sub>A</sub>16 module (Fold difference = 3.69, P<sub>FDRadjusted</sub> = 0.23). The enrichment between HAR genes and edu genes was higher (Fold difference = 3.0, P = 2.5x10<sup>-10</sup>) than between M<sub>A</sub>1 and HAR genes (Fold difference = 1.1; P = 0.01). Again, the overlapping genes

between edu genes and  $M_{A1}$  were significantly more enriched for HAR genes than non-overlapping  $M_{A1}$  genes (OR = 6.59;  $P = 2.4 \times 10^{-7}$ ), but the enrichment was non-significant when comparing the overlapping genes with the non-overlapping edu genes (OR = 2,  $P = 0.12$ ). We also investigated if the overlaps would remain significant if considering only HAR-interacting genes. We did not observe a significant overlap for either of the datasets (Edu- $M_{S2}$ -HAR\_interacting: Fold difference = 3.9;  $P = 0.09$ ; Edu- $M_{A1}$ -HAR\_interacting: Fold difference = 1.2;  $P = 0.54$ ).

*Insert Figure 3 here*

## Discussion

Here, we provide evidence for the interaction between edu genes and risk genes and co-expression networks for autism and schizophrenia. We further show that genes that overlap between the co-expression networks encompassing disease risk and genetic association for educational attainment are significantly enriched for genes flanking human accelerated regions<sup>20</sup>, suggesting that the regulation of these overlapping genes is under positive selection in humans. This overlap with educational attainment likely also explains the previous observations that schizophrenia risk loci are also enriched near HARs<sup>21</sup>. We show that, for autism, both genes that are frequently mutated in autism, and genes that have differential expression in post mortem autism brain, interact with edu genes, but in distinct co-expression modules that represent different biological pathways.

We built on previous research that identified a positive genetic correlation between common variants educational attainment and common variation in two psychiatric conditions – autism and schizophrenia<sup>14,15</sup>. We replicated this correlation using the latest educational attainment GWAS, which has more than twice the number of participants from the previous GWAS on educational attainment. To identify points of convergence between these seemingly disparate phenotypes, we tested two broad hypotheses. The first is that the same genes that are implicated in educational attainment are also more frequently mutated in autism and learning difficulties or intellectual disability (ID) (*the upstream hypothesis*). The second is that educational attainment genes converge in more downstream cellular and molecular pathways (*the downstream hypothesis*). Our results provide preliminary evidence for the involvement of edu genes in both mechanisms.

The significant overlap between edu genes and high confidence autism risk genes is consistent with the observation that approximately 40% of individuals with autism have co-morbid ID<sup>1</sup>. This represents the *upstream hypothesis*, where mutations in specific genes are likely to have large effects. Hence, deleterious mutations in edu genes cause more severe phenotypes with marked cognitive deficits such as ID and autism. This is supported by the high PLI scores for edu genes compared to all brain expressed genes. Both edu and high confidence autism risk genes converge on common neural pathway, including synapse development and maintenance. In addition, both sets of genes are active in the fetal brain<sup>24,33</sup>. Our results also show that this enrichment is not specific to autism alone, but can be extended to ID, using genes identified in multiple cohorts.

We further identify a co-expression module (M2) in the developmental transcriptome of the cortex that is enriched for both high confidence, large effect size, autism risk genes<sup>33</sup> and edu genes, which is supported by a highly significant three-way overlap between edu genes, genes in the M2 co-expression module, and high confidence autism risk genes<sup>33</sup>. Since the M2 module was originally identified with a relatively small number of ASD risk genes<sup>33</sup>, its greater than 4-fold enrichment with subsequently identified risk genes validates the original prediction that such network modules are indeed enriched in ASD risk and can be used to prioritize candidate genes<sup>33</sup>. Notably, this module was not significantly enriched for genes that are differentially expressed in schizophrenia or autism underscoring previous reports that show that differentially expressed genes and high confidence, large effect size, autism risk genes may lie in different co-expression networks<sup>30,33</sup>.

The M2 module is highly preserved in utero and in early infancy and is enriched for genes involved in the GO processes chromatin binding and regulation, and thus is thought to be involved in neural differentiation and axon guidance. Mutations in these early processes likely contribute to global changes in neural architecture leading to more severe phenotypes, and thus, it is not surprising that high confidence autism genes and edu genes converge in this co-expression module. While variants with small effects in these genes contribute to higher educational attainment, more deleterious mutations can have the opposite effect and significantly reduce cognitive function leading to autism co-morbid with ID. To our knowledge, far fewer major risk genes for schizophrenia have been identified with similar statistical support as for ASD<sup>41</sup>, and as such we do not have sufficient power to test for overlap. One previous study has investigated the enrichment of mutated genes in autism,



schizophrenia, and ID with educational attainment<sup>24</sup>. However, here we focussed only on the subset of high confidence genes for autism and included an expanded ID gene set.

We also identify an enrichment for edu genes in the schizophrenia M<sub>S</sub>2 and the autism M<sub>A</sub>1, CTX-M<sub>A</sub>4, CTX-M<sub>A</sub>10, and CTX-M<sub>A</sub>16 modules using the largest yet cortical transcriptome dataset for schizophrenia and autism that is publically available. We note that the autism transcriptome datasets are considerably smaller than the schizophrenia dataset, and as such, affords lower power for subsequent analyses. All the modules represent more *downstream processes* involved in the two conditions, and can be thought to reflect both genetic and environmental processes that contribute to the two conditions. The M<sub>S</sub>2 module comprises genes that are upregulated in schizophrenia when compared to controls, whereas the autism modules comprise genes that are downregulated in autism. Notably, the M<sub>S</sub>2, M<sub>A</sub>1, CTX-M<sub>A</sub>4 and CTX-M<sub>A</sub>16 modules are also enriched for genes involved in synaptic transmission, suggesting that they capture similar underlying biological processes. The M<sub>S</sub>2 module was also enriched for common variants associated with schizophrenia<sup>32</sup>. Further, both the overlapping gene sets were significantly enriched for genes flanking HARs, suggesting that they may be under positive selection, maintaining the variants in the general population. Previous research has showed that genetic loci associated with schizophrenia, a condition with reduced fecundity, are enriched near HARs, which are under positive selection. Here we show that this association may be driven by the pleiotropic overlap with educational attainment. We found that, at the level of transcriptional co-expression modules, risk genes for schizophrenia interact closely with genes that contribute to educational attainment. We show that genes that overlap between schizophrenia and edu attainment have higher enrichment for HAR genes than edu genes alone, or the M<sub>S</sub>2 genes alone. We also identify a similar mechanism of enrichment with the autism co-expression modules. This pleiotropy at the level of transcriptional networks may also explain why some individuals with autism or schizophrenia have considerable talent in certain cognitive domains, and the positive genetic correlation at the level of common variants between schizophrenia, autism and educational attainment.

We used genes identified for educational attainment as a proxy for intelligence and cognition. While there is not a perfect correlation between educational attainment and IQ, the educational attainment GWAS is the largest GWAS to date of a cognitive phenotype<sup>24</sup>. As such, it affords substantial power to robustly identify genes associated with cognition and

intelligence. We reasoned that some cellular and molecular pathways will be preserved between educational attainment and general cognition. Indeed, we identified a modest enrichment for edu genes in the M<sub>c</sub>3 module that is also enriched for genes identified for cognition<sup>29</sup>. This module is conserved across different brain regions in humans and rodents. The M<sub>c</sub>3 module has been previously shown to be significantly enriched for genes associated with delayed recall and general fluid cognitive ability<sup>29</sup>. Both M<sub>c</sub>1 and M<sub>c</sub>3 are enriched for pathways involved in synaptic processes, in particular, the post-synaptic density and Reelin signalling pathway<sup>29</sup>. To underscore the specificity of this enrichment, we did not identify any enrichment for genes identified for height, a phenotype with large sample sizes and relatively well powered GWAS, but that has limited neural basis.

Our results also provide preliminary evidence that general cognition and educational attainment converge in common co-expression modules derived from normal brain and disease tissue supporting previous genetic correlation analyses that show a high to modest correlations between educational attainment and different aspects of cognition<sup>39</sup>. However, while educational attainment is often regarded as a proxy phenotype for cognition, it can also reflect contributions from other factors such as ‘grit’<sup>42</sup> and socio-economic status<sup>43</sup>. In addition, while general cognitive aptitude may be enriched across modules conserved across different species of mammals, educational attainment can be thought of as a much more complex phenotype involving several uniquely human traits. Further, while some aspects of cognitive aptitude may primarily involve the hippocampus (such as working memory), educational attainment may involve many cortical areas. As such, it is possible that edu genes may be enriched in additional cortical networks that are not necessarily conserved across species.

It is difficult to tease apart the directionality in gene regulation from genome-wide association studies. Indeed, it is likely that these genes exhibit complex regulatory patterns with changes in specific developmental time point in specific tissues contributing to higher cognitive aptitude. Understanding how the expression edu genes changes in relation to educational aptitude will help understand if their co-expression with schizophrenia and autism risk genes lead to cognitive deficits, or conversely to talent. In this study, we have used the largest available datasets to identify both edu genes and modules associated with autism and schizophrenia, though, the autism co-expression modules were constructed using only about a third of the samples used for schizophrenia. Replication in a larger dataset will

help confirm the validity of these findings and may aid in identifying both common pathways and directionality between the three phenotypes.

## References

1. Lai M-C, Lombardo M V, Baron-Cohen S. Autism. *Lancet*. 2013. doi:10.1016/S0140-6736(13)61539-1.
2. Baron-Cohen S. *Mindblindness*: An Essay on Autism and Theory of Mind. MIT Press; 1997.
3. Frith U. *Autism*: Explaining the Enigma. Blackwell Pub; 2003.
4. Baron-Cohen S. Autism: research into causes and intervention. *Pediatr Rehabil*. 7(2):73-78.
5. Wheelwright S, Baron-Cohen S, Goldenfeld N, et al. Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain Res*. 2006;1079(1):47-56.
6. Baron-Cohen S. Autism: The Empathizing-Systemizing (E-S) Theory. *Ann N Y Acad Sci*. 2009;1156(1):68-80.
7. Wei X, Christiano ERA, Yu JW, Wagner M, Spiker D. Reading and math achievement profiles and longitudinal growth trajectories of children with an autism spectrum disorder. *Autism*. 2015;19(2):200-210.
8. Iuculano T, Rosenberg-Lee M, Supekar K, et al. Brain organization underlying superior mathematical abilities in children with autism. *Biol Psychiatry*. 2014;75(3):223-230.
9. Howlin P, Goode S, Hutton J, Rutter M. Savant skills in autism: psychometric approaches and parental reports. *Philos Trans R Soc Lond B Biol Sci*. 2009;364(1522):1359-1367.
10. Baron-Cohen S. Does autism occur more often in families of physicists, engineers, and mathematicians? *Autism*. 1998;2(3):296-301.

11. Ruzich E, Allison C, Chakrabarti B, et al. Sex and STEM occupation predict Autism-Spectrum Quotient (AQ) scores in half a million people. *PLoS One*. 2015;10(10):e0141229.
12. O'Carroll R. Cognitive impairment in schizophrenia. *Adv Psychiatr Treat*. 2000;6(3):161-168.
13. Power RA, Steinberg S, Bjornsdottir G, et al. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat Neurosci*. 2015;18(7):953-955.
14. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236-1241.
15. Zheng J, Erzurumluoglu M, Elsworth B, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 2016. [Epub ahead of print]
16. Benyamin B, Pourcain Bs, Davis OS, et al. Childhood intelligence is heritable, highly polygenic and associated with FBNP1L. *Mol Psychiatry*. 2014;19(2):253-258.
17. Clarke T-K, Lupton MK, Fernandez-Pujals AM, et al. Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Mol Psychiatry*. 2016; 21(3):419-25.
18. Le Hellard S, Wang Y, Witoelar A, et al. Identification of gene loci that overlap between schizophrenia and educational attainment. *Schizophr Bull*. 2016:sbw085.
19. Robinson EB, Samocha KE, Kosmicki JA, et al. Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proc Natl Acad Sci*. 2014;111(42):15161-15165.
20. Doan RN, Bae B-I, Cubelos B, et al. Mutations in human accelerated regions disrupt cognition and social behavior. *Cell*. 2016;167(2):341-354.e12.

21. Xu K, Schadt EE, Pollard KS, Roussos P, Dudley JT. Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. *Mol Biol Evol.* 2015;32(5):1148-1160.
22. Power RA, Kyaga S, Uher R, et al. Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry.* 2013;70(1):22.
23. Pers TH, Karjalainen JM, Chan Y, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun.* 2015;6:5890.
24. Okbay A, Beauchamp JP, Fontana MA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature.* 2016;533(7604):539-542.
25. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet.* 2014;46(11):1173-1186.
26. Sanders SJ, He X, Willsey AJ, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron.* 2015;87(6):1215-1233.
27. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet.* 2015;17(1):9-18.
28. Ardlie KG, Deluca DS, Segre A V., et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80- ).* 2015;348(6235):648-660.
29. Johnson MR, Shkura K, Langley SR, et al. Systems genetics identifies a convergent gene network for cognition and neurodevelopmental disease. *Nat Neurosci.* 2015;19(2):223-232.
30. Gupta S, Ellis SE, Ashar FN, et al. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat Commun.* 2014;5:5748.

31. Parikshak NN, Swarup V, Belgard TG, et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*. 2016. [Epub ahead of print]
32. Fromer M, Roussos P, Sieberts SK, et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci*. 2016;19(11):1442-1453.
33. Parikshak NN, Luo R, Zhang A, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*. 2013;155(5):1008-1021.
34. Kang HJ, Kawasawa YI, Cheng F, et al. Spatio-temporal transcriptome of the human brain. *Nature*. 2011;478(7370):483-489.
35. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295.
36. Ripke S, Neale BM, Corvin A, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421-427.
37. Lek M, Karczewski K, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2015;536(7616):285-291
38. Wang M, Zhao Y, Zhang B, et al. Efficient test and visualization of multi-set intersections. *Sci Rep*. 2015;5:16923.
39. Davies G, Marioni RE, Liewald DC, et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112,329). *Mol Psychiatry*. 2016;21(6):758-767.
40. Sabin MA, Werther GA, Kiess W. Genetics of obesity and overgrowth syndromes. *Best Pract Res Clin Endocrinol Metab*. 2011;25(1):207-220.
41. Singh T, Kurki MI, Curtis D, et al. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci*. 2016;19(4):571-577.

42. Duckworth AL, Peterson C, Matthews MD, Kelly DR. Grit: Perseverance and passion for long-term goals. *J Pers Soc Psychol.* 2007;92(6):1087-1101.

44. Paterson L. Socio-economic status and educational attainment: A multi-dimensional and multi-level study. *Eval Res Educ.* 1991;5(3):97-121.

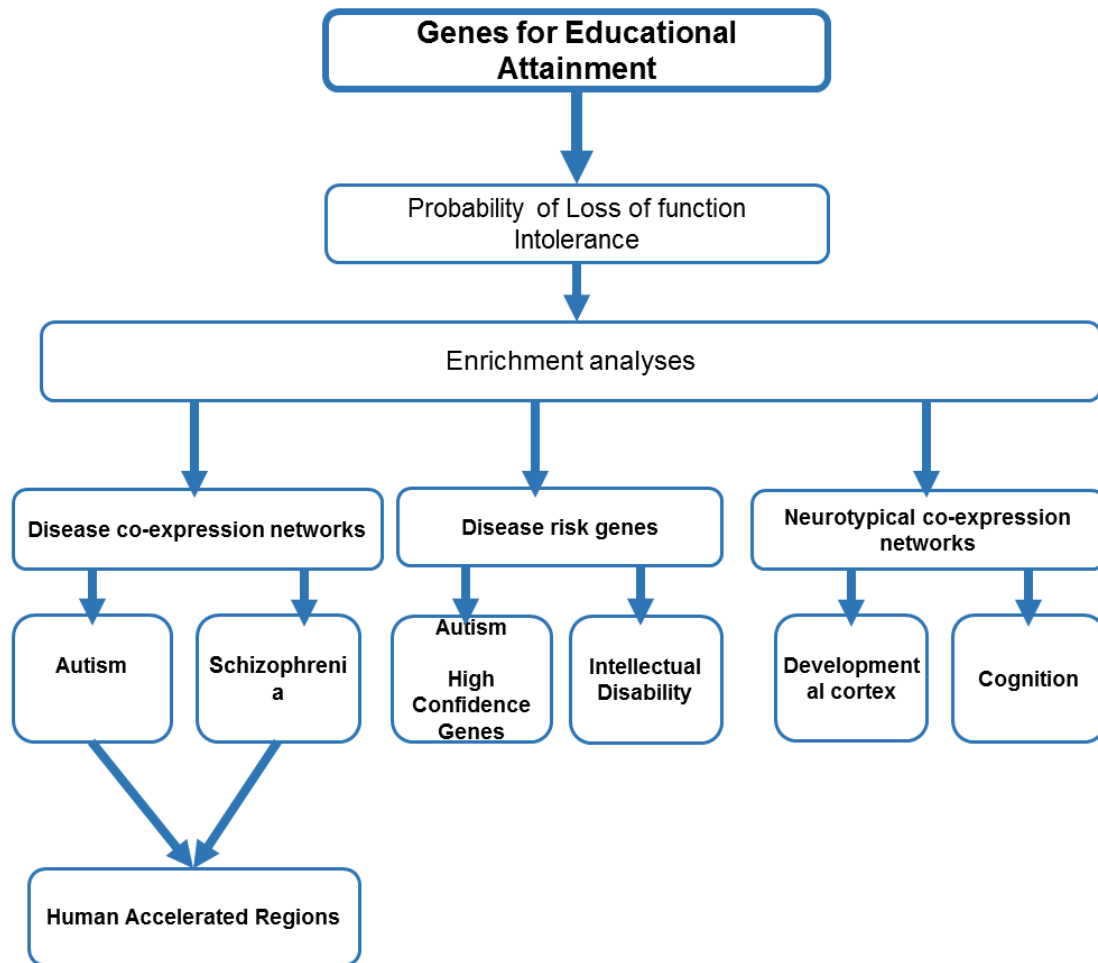


## **Author contributions**

All authors designed the study. VW and RAIB analysed the data. VW and RAIB wrote the manuscript that was then edited by DG and SBC.

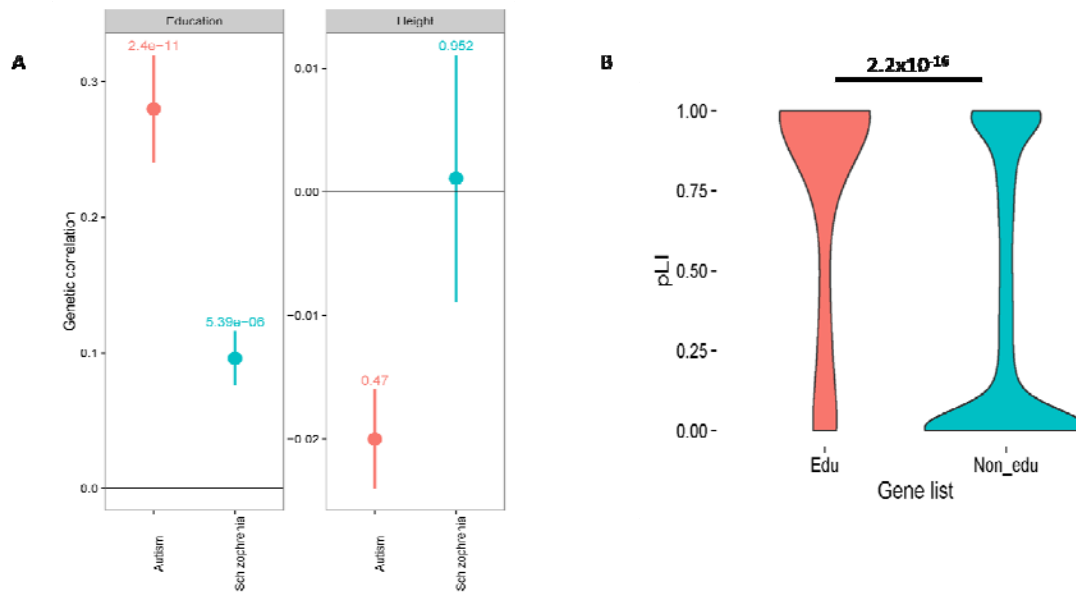
## Figures

**Figure 1: Schematic diagram of the study protocol**

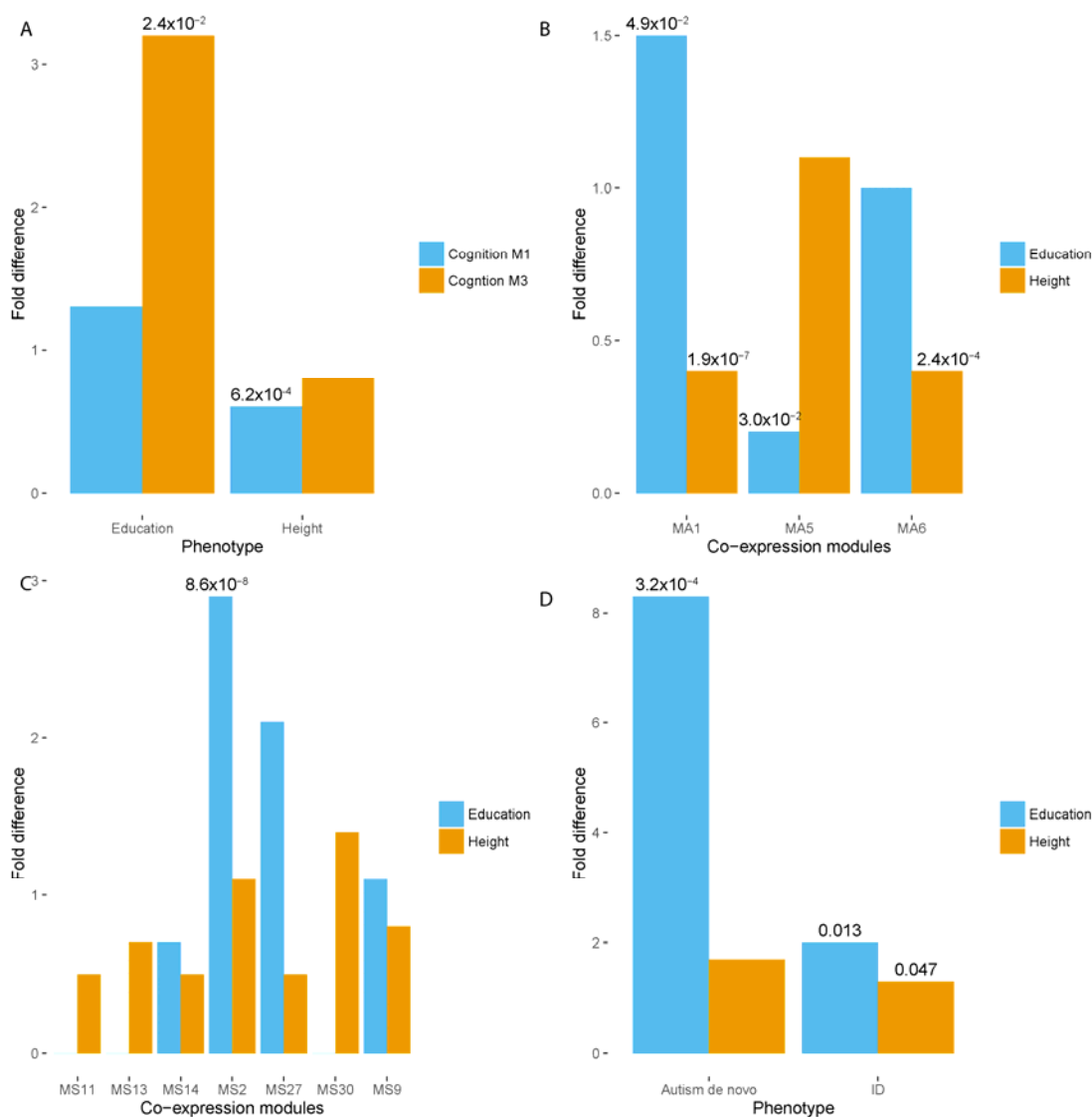


*Schematic diagram of the study protocol. We first tested for probability of loss of function intolerance for the edu genes and compared with all brain expressed genes. We next tested the enrichment of edu genes in various previously identified weighted gene co-expression modules for autism, schizophrenia, cognition and the developmental cortical transcriptome. We also checked for overlap between edu genes and high-confidence autism genes and intellectual genes. We finally investigated if the enrichment in the autism and schizophrenia co-expression module was guided by genes flanking human accelerated regions.*

**Figure 2: Genetic correlation analyses and distribution of pLI scores for edu genes**



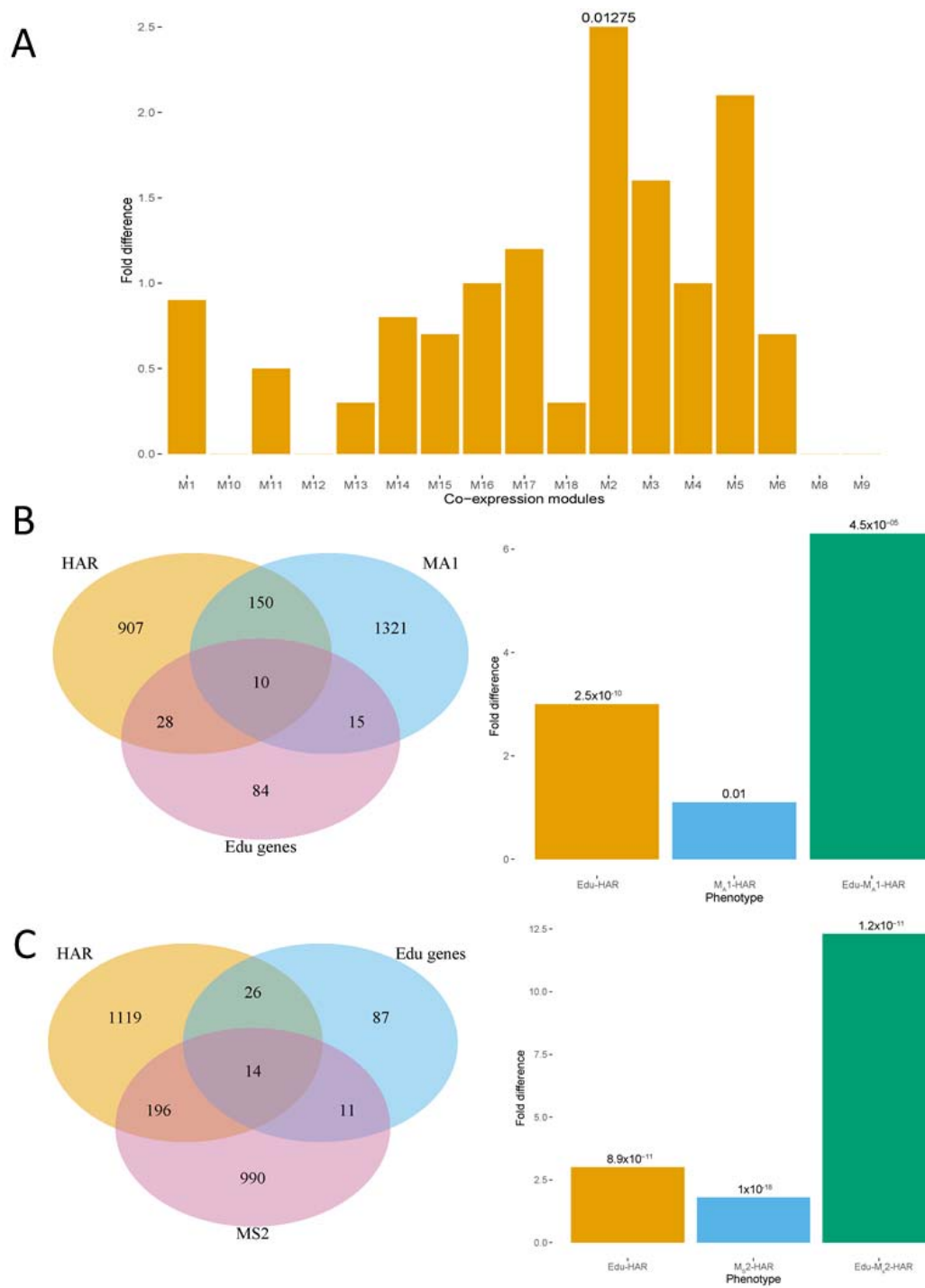
### Figure 3: Enrichment analyses



For all the bar charts, y-axes indicate fold difference. A fold difference  $> 1$  is indicative of enrichment, and a fold difference  $< 1$  is indicative of depletion. Significant P-values after FDR correction are provided on top of the bar charts. 3A: Enrichment of edu genes and height genes in the two cognition modules ( $M_{C1}$  and  $M_{C3}$ ). 3B: Enrichment of edu genes and height genes in three autism co-expression modules ( $M_{A1}$ ,  $M_{A5}$  and  $M_{A6}$ ). 3C: Enrichment of edu genes and height genes in the schizophrenia co-expression modules ( $M_{S11}$ ,  $M_{S13}$ ,  $M_{S14}$ ,

*M<sub>S</sub>2, M<sub>S</sub>27, M<sub>S</sub>30, and M<sub>S</sub>9). 3D: Enrichment of edu genes and height genes in genes implicated through sequencing in autism and intellectual disability (ID).*

**Figure 4: Enrichment in the developmental cortical transcriptome modules and overlap with genes flanking human accelerated regions**



*4A: Bar charts showing enrichment of Edu genes in developmental cortical co-expression modules. Y-axis provides fold difference. Significant P-values after FDR correction provided on top of the graphs. 4B: Venn diagram (left) and bar chart showing overlap and enrichment between Edu genes, autism  $M_{A1}$  module and HAR flanking genes. P-values provided on top of bar charts. 4C: Venn diagram (left) and bar chart showing overlap and enrichment between Edu genes, Schizophrenia  $M_{S2}$  module and HAR flanking genes. P-values provided on top of bar charts.*