Genes influenced by MEF2C contribute to neurodevelopmental disease via gene expression changes that affect multiple types of cortical excitatory neurons

Donna Cosgrove[1a], Laura Whitton[1a], Laura Fahey[1,2a], Pilib Ó Broin[2], Gary Donohoe[1], Derek W. Morris[1]*

[1] Cognitive Genetics and Cognitive Therapy Group, Centre for Neuroimaging & Cognitive Genomics, School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Galway, Ireland.

[2] School of Mathematics, Statistics and Applied Mathematics, National University of Ireland Galway, Galway, Ireland.

[a] The authors wish it to be known that, in their opinion, the first 3 authors should be regarded as joint First Authors.

Corresponding author:

Dr Derek Morris, Room 106, Discipline of Biochemistry, National University of Ireland Galway, University Road, Galway, Ireland.

Tel: + 353 91 494439; Email: derek.morris@nuigalway.ie

Abstract

Myocyte enhancer factor 2 C (MEF2C) is an important transcription factor during neurodevelopment. Mutation or deletion of *MEF2C* causes intellectual disability (ID) and common variants within *MEF2C* are associated with cognitive function and schizophrenia risk. We investigated if genes influenced by MEF2C during neurodevelopment are enriched for genes associated with neurodevelopmental phenotypes, and if this can be leveraged to identify biological mechanisms and individual brain cell types affected. We used a set of 1,055 genes that were differentially expressed in the adult mouse brain following early embryonic deletion of *Mef2c* in excitatory cortical neurons. Using GWAS data, we found these differentially expressed genes (DEGs) to be enriched for genes associated with schizophrenia, intelligence and educational attainment but not autism spectrum disorder (ASD). For this gene-set, genes that overlap with target genes of the Fragile X mental retardation protein (FMRP) are a major driver of these enrichments. Using trios data, we found these DEGs to be enriched for genes containing *de novo* mutations reported in ASD and ID, but not schizophrenia. Using single cell RNA-seq data, we identified that a number of different excitatory glutamatergic neurons in the cortex were enriched for these DEGs including deep layer pyramidal cells and cells in the retrosplenial cortex, entorhinal cortex and subiculum, and these cell types are also enriched for FMRP target genes. The involvement of MEF2C and FMRP in synapse elimination suggests that disruption of this process in these cell types during neurodevelopment contributes to cognitive function and risk of neurodevelopmental disorders.

Introduction

Myocyte enhancer factor 2 C (MEF2C) is a transcription factor with a highly conserved DNA binding domain and is crucial for differentiation and development through potentiation of other regulatory mechanisms. MEF2C is required for neurogenesis (1), neuronal distribution, and electrical activity in the neocortex (2, 3). MEF2C is the earliest of four subclasses of MEF2 transcription factors to be expressed during embryonic brain development and maintains high levels of expression in adult neurons (4). The N-termini of MEF2 contain a highly conserved MADS-box, with an adjacent 'MEF2 domain' that mediates DNA binding and other interactions; and the C-terminal regions display patterns of alternative splicing. The activity of MEF2 proteins is also tightly regulated by class II histone deacetylases (HDACs) interacting directly with the MADS domain (5). This interaction with HDACs indicates a point of convergence of multiple epigenetic regulatory mechanisms, (2) with MEF2C playing a role in epigenetic alterations in chromatin configuration. This includes transmitting extracellular signals to the genome and activating genetic programmes that control cell differentiation, proliferation, survival and apoptosis.

Although *Mef2c* homozygous knockout mice do not survive the gestation period due to incomplete cardiac morphogenesis (2, 6), *Mef2c* heterozygous mice are viable, although only 44% survive until three weeks, compared to wild-type mice. *Mef2c* conditional knockout mice exhibit reduced spatial learning and memory function (1) and deficits in locomotor activity and motor coordination (7). *Mef2c* is a negative regulator of both embryonic and postnatal synaptogenesis (7, 8), with heterozygous knockout mice displaying neuronal and synaptic abnormalities (9). This includes a significant increase in dendritic spine density in the hippocampus, seen in conditional knockout mice that targeted *Mef2c* deletion selectively to the brain during embryogenesis and postnatally (7), and during CNS deletion of *Mef2c* (10). Adult hippocampal slices from Mef2c-null brains compared with WT showed characteristics of an immature neuronal network (1), and abnormal neuronal behaviour was observed in the neocortex during development. This indicates that MEF2C plays a pivotal role in early neuronal differentiation (11). MEF2C specific knockout impairs hippocampal-dependent learning and memory

by increasing synapse number and potentiating synaptic transmission. *Mef2c* expression is normally upregulated during neuronal differentiation and maturation (10, 12). However, deletion of *Mef2c* in the brain postnatally did not impact learning and memory, measurements of synaptic plasticity, or several behavioural measures suggested to recapitulate aspects of autism spectrum disorders in mice (7), implying a distinct role for Mef2c in the prenatal versus the postnatal brain.

Harrington et al. (13) sought to evaluate the role of MEF2C in neurodevelopment by generating conditional knockout (cKO) of *Mef2c* in forebrain excitatory neurons of mice, and investigating the downstream effects on cortical synaptic transmission. The effect observed was a small reduction in glutamatergic synaptic transmission and a large increase in inhibitory synaptic transmission, with Mef2c directly regulating the densities of both excitatory and inhibitory synapses. Data indicated that endogenous Mef2c functions predominantly as a transcriptional repressor to inhibit target genes that promote excitatory synapse elimination and inhibitory synapse formation and/or stability. Differentially expressed genes (DEGs) from the somatosensory cortex of *Mef2c* cKO vs control mice were enriched for genes involved in neuron differentiation and development (up-regulated genes) and synaptic transmission and ion transport (down-regulated genes). This set of DEGs was enriched for autism spectrum disorder (ASD) risk genes. Finally, analysis of *Mef2c* cKO mice identified impairments in multiple behavioural phenotypes, e.g. fear learning and memory, multiple social behaviours, socially-motivated ultrasonic vocalizations, reward-related behaviours and repetitive motor behaviours. These overlap with human neurodevelopmental disorders such as ASD, intellectual disability (ID) and schizophrenia (SZ) where cognitive function is impaired (13).

A *MEF2C* haploinsufficiency syndrome has been identified where affected individuals are frequently non-verbal (14), with severe intellectual disability, epilepsy, stereotypic abnormal movements, minor dysmorphisms, and brain abnormalities (15-17). The relevance of *MEF2C* to SZ and cognition is highlighted by a study showing that (a) MEF2C binding motifs are enriched within the pool of top genome-wide association studies (GWAS) SNPs for schizophrenia, (b) there is a specific over-

representation of MEF2C motifs among sequences with SZ-associated histone hypermethylation based on chromatin profiling in neuronal nucleosomes extracted from prefrontal cortex of cases and controls, and (c) Mef2c upregulation in mouse prefrontal projection neurons consistently resulted in enhanced cognitive performance in working memory and object recognition paradigms (18).

A number of very large GWAS have recently been published for ASD (19), SZ (20), cognitive ability / human intelligence (IQ) (21-23) and educational attainment (EA) (24, 25). These studies identify very strong association between *MEF2C* and IQ and EA in the general population, plus association for *MEF2C* with SZ but not with ASD. Based on *MEF2C* association with cognition and SZ risk, our hypothesis was that genes differentially expressed upon cKO of *Mef2c* during neurodevelopment would also be associated with these phenotypes. We investigated this using the set of DEGs reported by Harrington et al. (13). We tested and found that the DEGs are enriched for genes associated with IQ, EA and SZ using gene-set analysis of GWAS data and are enriched for genes containing *de novo* mutations reported in exome sequencing studies of ASD and ID trios. To gain further biological insights, we then leveraged these contributions of the MEF2C-influenced genes to neurodevelopment to investigate which individual cell types are enriched for these genes using single cell RNA-seq (scRNA-seq) data from the mouse brain and performed gene ontology analysis to help identify the molecular mechanisms involved.

Results

Comparison of association signals at MEF2C in GWAS data

We reviewed the largest published GWAS, which report that SNPs at the *MEF2C* locus are genome-wide significant for SZ, IQ and EA but not ASD (Supplementary Figure 1A-D). The top associated SNPs for IQ (rs34316) and EA (rs254781) are in high linkage disequilibrium (LD; r^2=0.98) and span the 3' end and downstream region of the gene. Another SNP, rs639725, which is in high LD with these IQ and EA associated SNPs (r^2>0.98), has been reported as an expression quantitative trait locus (eQTL) for MEF2C in the cerebellum of Alzheimer's disease patients and healthy controls (26) (identified via HaploReg v4.1), where analysis indicates that the alleles associated with reduced IQ and EA are associated with increased expression of *MEF2C*. We did not find eQTLs related to these GWAS SNPs in the other datasets investigated that included both adult and fetal samples. The top associated SNP for SZ is also in the downstream region but not in high LD with the IQ and EA associated SNPs indicating that there are different genetic variants influencing SZ risk and cognition. We found no evidence of an eQTL involving the SZ risk SNP. *Post mortem* gene expression analysis of cerebral cortex reports an increase in MEF2C expression in SZ cases compared to controls (27). The effect of the SZ risk SNP at MEF2C remains to be elucidated but these *post-mortem* gene expression results indicating that increased expression of MEF2C is observed in SZ is consistent with the eQTL data that points to increased expression of MEF2C being associated with reduced IQ and EA.

Analysis of MEF2C cKO gene-set in GWAS data

We used a set of 1,055 DEGs (listed in Supplementary Table 1) based on an RNA-seq study that captured the transcriptional changes in adult male mouse brain that result from the early embryonic deletion of *Mef2c* in cortical and hippocampal excitatory neurons (13). RNA-seq was performed on mRNA isolated from the somatosensory cortex of *Mef2c* cKO or control littermates (13). These genes that were differentially expressed upon cKO of MEF2C were enriched for genes associated with SZ (P=5.01×10^{-07}), IQ (P =9.91×10^{-08}) and EA (P=2.35×10^{-10}) but not ASD (P=0.247; Figure 1 and

Supplementary Table 2). Brain-expressed genes are a major contributor to these phenotypes. It is possible that the enrichment detected here could be due to the MEF2C gene-set representing a set of brain-expressed genes. However, the MEF2C enrichment was robust to the inclusion in the analyses of both 'brain-expressed' (n=14,243) and 'brain-elevated' (n=1,424) gene-sets as covariates (Supplementary Table 2). To examine if the enrichment we detect for SZ, IQ and EA is a property of polygenic phenotypes in general, we obtained GWAS summary statistics for ten phenotypes and tested the MEF2C gene-set for enrichment in each one. These were a child-onset psychiatric disorders, other brain-related disorders, non-brain related diseases, and height. No enrichment was detected for any of the ten phenotypes (Supplementary Table 3).

Gene ontology (GO) enrichment analysis of this gene-set had identified that the up-regulated genes were likely involved in neuron differentiation and development whereas the down-regulated genes were involved in synaptic transmission and ion transport (13). To investigate if the up- and down-regulated genes made different contributions to SZ, IQ and EA, we performed gene-set analysis on the up-regulated and down-regulated gene sets separately. Overall, the down-regulated gene-set showed slightly stronger enrichment for SZ, IQ and EA than the up-regulated gene-set (Figure 1 and Supplementary Table 2).

Harrington et al. (13) previously reported a significant overlap between the MEF2C DEGs and the target genes of the RNA-binding protein, Fragile X mental retardation protein (FMRP) (28). We investigated the contribution of this overlap to our gene-set enrichments by separating our MEF2C gene-set into those genes that are FMRP targets (MEF2C DEGs overlap FMRP; n=116) and those that are not (MEF2C DEGs minus FMRP; n=939). The MEF2C DEGs overlapping FMRP target genes are a major driver of the enrichments that we are observing in the GWAS data (EA $P=3.96\times10^{-06}$; IQ $P=1.46\times10^{-07}$; SZ $P=1.54\times10^{-08}$) but the set of MEF2C DEGs minus FMRP target genes still show significant enrichments (EA $P=5.45\times10^{-07}$; IQ $P=1.25\times10^{-04}$; SZ P=0.00044; Figure 1 and Supplementary Table 2). Additionally, overall results remain highly significant after analysis of the full MEF2C gene-set when

conditioning on the full set of FMRP targets (EA $P=1.94 \times 10^{-09}$; IQ $P=4.96 \times 10^{-07}$; SZ $P=2.27 \times 10^{-06}$; Supplementary Table 2), indicating that FMRP targets are important contributors to, but are not solely responsible for, the enrichment among MEF2C-regulated genes for genes contributing to cognition and SZ.

Analysis of MEF2C gene-set using data on *de novo* mutations

To investigate the contribution of rare variants in the MEF2C gene-set to SZ, ASD and ID, we tested if the gene-set was enriched for *de novo* mutations (DNMs) that have been reported in trios-based studies of these disorders. We detected strong enrichment for loss-of-function (LoF) DNMs reported in ASD patients in the MEF2C gene-set where we observed 45 DNMs in the set of DEGs when expecting just 24 DNMs (enrichment value of 1.88; $P=7.96 \times 10^{-05}$; Table 1). We also found enrichment for missense DNMs in the ASD trios but this did not survive multiple test correction. We detected strong enrichment for missense DNMs reported in ID patients in the MEF2C gene-set ($P=3.57 \times 10^{-06}$; Table 1) where the number of 25 observed DNMs was nearly 3-fold higher than expected. The sample size for ID (n=192 trios) was small for observing a significant enrichment of LoF DNMs. As the common variant enrichment signal was stronger in the down- compared to the up-regulated genes, for follow-up we tested for enrichment of LoF DNMs in ASD patients and missense DNMs in ID patients in these subsets of genes separately. For both disorders, we found the DNMs tested to be similarly significantly enriched in both down- and up-regulated genes (data not shown). The genes containing LoF DNMs in ASD cases and missense DNMs in ID cases are identified in Supplementary Table 1. By way of control, the MEF2C gene-set was not enriched for synonymous DNMs (unlikely to be functional) reported in ASD or ID trios. In addition, both control trios and trios including unaffected siblings of ASD patients showed no enrichment for DNMs in the MEF2C gene-set (Table 1).

Functional annotation analysis

Following the significant enrichments above, we investigated if the genes contributing to each of the neurodevelopmental phenotypes influence the same or different biological functions by generating five subsets of the MEF2C gene-set based; (i) genes associated with SZ (n=41), (ii) genes associated with IQ (n=37), (iii) genes associated with EA (n=109), (iv) genes containing LoF DNMs in ASD cases (n=45) and (v) genes containing missense DNMs in ID cases (n=25). GO analysis identified different terms to be significantly enriched for these gene subsets. These data are summarized in Table 2 and detailed in Supplementary Tables 4-8. Terms related to development of the nervous systems are enriched in analyses of all five subsets of genes. Terms related to synaptic signalling are strongly enriched for SZ and EA. The most significantly enriched terms for IQ map to pre- and post-synapse function. The most significantly enriched terms for EA map to the synaptic membrane and synaptic signalling. For ASD and ID genes, other significantly enriched terms map to cell morphogenesis.

Cell type enrichment analysis

By performing enrichment analysis on gene expression data from scRNA-seq of the mouse nervous system, we sought to detect individual cell types that are enriched for genes that are differentially expressed upon cKO of MEF2C. We used two scRNA-seq data-sets that included 265 cell types from the mouse nervous system (29) and 565 cell types from the mouse brain (30). After Bonferroni correcting for the 830 cell types tested, we identified 31 significantly enriched cell types in both data-sets (62 cell types in total with $p<6.03x10^{-5}$). It is possible that the enriched cell types detected here could be due to the MEF2C gene-set representing a set of genes that are expressed in the somatosensory cortex, which is their origin. To investigate this, we generated 1,000 randomized gene-sets derived from the total Harrington et al. (13) gene expression dataset of 7,977 genes from the somatosensory cortex, where each gene-set contained the same number of genes as our test MEF2C gene-set. Testing each random gene-set for cell type enrichment allowed us to calculate a simulated P value for each cell type. Fifty-six of the 62 cell types remained significant indicating that these cell type enrichments for the

MEF2C DEGs are not simply a function of analyses of cortically-expressed genes. These results are summarized in Table 3 and full details are provided in Supplementary Tables 9-10. Results were consistent between the two scRNA-seq data-sets where only neurons were enriched and these were predominantly glutamatergic excitatory neurons in the cortex, e.g. deep layer pyramidal cells and cell in the retrosplenial cortex, entorhinal cortex and subiculum. These are the specific cell types in the adult mouse brain that are likely to be most impacted by cKO of Mef2c during embryonic development. Although these DEGs are from analysis of the cortex, they are also enriched in other cell types: excitatory neurons in the hippocampus, e.g. cells in the cornu ammonis 1 (CA1) subfield, and GABAergic inhibitory neurons in hippocampus and striatum, e.g. dopamine D2 receptor expressing medium spiny neurons (MSNs; also called indirect spiny projection neurons; Table 3). Given the overlap between MEF2C DEGs and FMRP targets, we also performed a cell type enrichment analysis of the full set of FMRP targets (Supplementary Tables 9-10) and found that the cell type enrichments for MEF2C DEGs, measured as standard deviation (SD) from the mean, are highly correlated with the cell types enrichments for FMRP targets ($R^2 = 0.84$ for the Zeisel et al. dataset (29) and $R^2 = 0.86$ for the Saunders et al. dataset (30); Supplementary Figure 2A-B). Overall, our results are consistent with previous cell type enrichment analysis of GWAS data that highlights the relevant cell types for both SZ and IQ to be cortical somatosensory pyramidal cells, hippocampal CA1 pyramidal cells and striatal MSNs (22, 31). Thus, the cell types enriched for SZ- and IQ-associated genes are those cell types that are enriched for both genes that are differentially expressed as a result of cKO of *Mef2C* during neurodevelopment and genes that are targeted by FMRP.

Discussion

Across the MEF2C DEGs, there is a difference in terms of the type of genetic variant and the type of phenotype it associates with. The missense and LoF DNMs in ID and ASD cases are likely to disrupt or ablate one copy of a gene and are associated here with these severe early age-of-onset neurodevelopmental disorders. Common SNPs of weak effect underpin the enrichment of DEGs in genes associated with SZ, IQ and EA and these SNPs are likely to have subtle effects on the regulation and/or expression of these genes. Gene ontology analysis identifies both consistencies and differences between the biological processes that are enriched for DEGs contributing to the different phenotypes. Terms related to development of neurons are highly enriched for genes contributing to the five phenotypes. Terms related to synaptic function feature strongly in the analysis of SZ, IQ and EA but less so for ASD and not at all for ID. Terms related to cell morphogenesis feature strongly in the analysis of ASD and ID but do not rank as high for SZ, IQ and EA. Overall, it appears that the phenotypic effect of different genetic variants depends on (i) the effect of the variant on gene function, with larger effect mutations associating with early onset illness and (ii) the type of biological process impacted, with synaptic function being affected by genetic variants associating with SZ, IQ and EA. The latter finding above is consistent with previous analysis showing that the DEGs that are down-regulated upon cKO of MEF2C are involved in synaptic transmission (13) and our analysis showing that these down-regulated genes show greater enrichment for genes associated with SZ, IQ and EA than the up-regulated genes. A caveat to our analysis of GWAS data is that with a smaller sample size, the ASD GWAS may be under-powered to detect an enrichment of MEF2C DEGs.

Analysis of the *MEF2C* locus in GWAS indicates that although the signals for SZ and for IQ and EA co-localize to the 3' end and downstream region of the gene, the SNPs associated with SZ risk and cognitive function are not in high LD with each other. For the IQ/EA association signal, one eQTL study in *post-mortem* adult cerebellum shows that the alleles associated with reduced IQ and EA are associated with increased expression of *MEF2C*. This eQTL was not consistently detected across multiple eQTL analysis of different brain regions and is contrary to data that shows upregulation of

Mef2c in prefrontal projection neurons in mice resulted in enhanced cognitive performance (18). Ultimately, these are small effects and we do yet not know what effect these SZ- or IQ/EA-associated variants at *MEF2C* have on gene function in different brain regions and at different stages of development, and what the knock-on effects are on genes influenced by MEF2C.

scRNA-seq has delivered empirical taxonomies of brain cell types (30). Analysis of these data has allowed us to exploit the RNA-seq data to identify the individual cell types likely impacted by cKO of *Mef2c*. Genes that are differentially expressed upon cKO of *Mef2c* are specifically highly expressed across a large number of different cell types. As expected, based on the cell type targeted for cKO of *Mef2c* in the animal model, the majority of cell types affected are glutamatergic excitatory neurons. The advantage of these large scRNA-seq resources is that we can identify some of the specific regions and cell types affected. A number of the regional locations were identified in both analyses of scRNA-seq data and have previously been implicated in neurodevelopmental disorders. Transcriptomic analysis of dorsolateral prefrontal cortex layer 3 and layer 5 pyramidal cells have identified changes in gene expression between SZ cases and controls (32). LoF mutations in transcription factor T-brain-1 (*TBR1*) have been reported in ASD patients and conditional *Tbr1* deletion during late mouse gestation in cortical layer 6 neurons affects processes such as dendritic patterning and synaptogenesis (33). The retrosplenial cortex, a sub-region of the cingulate cortex, exhibited impaired function using a semantic memory task (34) and showed reduced grey matter volumes (35) in SZ patients. The common set of brain regions that underlies the default mode network, autobiographical memory, prospection, navigation, and theory of mind may include the retrosplenial cortex (36). Resting state connectivity of the retrosplenial cortex is altered in SZ patients, with connectivity differences observed to brain areas associated with language processing, suggesting a role for the retrosplenial cortex in the aetiology of hallucinations (37). In ASD, resting state overconnectivity between the anterior insula and the retrosplenial cortex is associated with internalizing symptoms (anxiety, depression, social withdrawal) (38)).

Two limitations of our study are: (i) The cKO of *Mef2c* only occurred in differentiated forebrain excitatory neurons during embryonic development and RNA-seq was only performed on the somatosensory cortex, thus we are not getting a whole-brain picture of what gene expression, biological pathways and cell types may be influenced if MEF2C was manipulated in all tissues. If KO of *Mef2c* results in altered expression of a similar set of genes in hippocampus and striatum, our scRNA-seq analysis suggests that glutamatergic excitatory neurons and GABAergic inhibitory neurons in the hippocampus, and GABAergic inhibitory D2 medium spiny neurons in the striatum would be affected. Such results would be consistent with the balance of excitatory and inhibitory synapses in the developing brain being an element of neurodevelopmental disease. (ii) RNA-seq was only performed on the adult mouse, thus we are not getting a picture of the effect cKO of *Mef2c* in the developing brain. We acknowledge that a set of DEGs from embryonic tissue would give us a better insight into the molecular mechanisms of neurodevelopmental phenotypes and how they are influenced by *Mef2c*. But what we are getting is insight into the molecular consequences of embryonic cKO of *Mef2c* on the adult brain, and importantly this cKO of *Mef2c* during neurodevelopment did cause the behavioural phenotypes in adult mice related to memory, social behaviour, reward-related behaviours and repetitive motor behaviours that are consistent with neurodevelopmental disorders (13).

What underpins these behavioural phenotypes in the mouse model is a reduction in cortical network activity, partly as a result of an increase in inhibitory and a decrease in excitatory synaptic transmission (13). We find that these MEF2C DEGs are enriched for genes associated with SZ, IQ and EA, based on analysis of common variants, and are enriched for genes that contribute to ASD and ID, based on analysis of rare mutations. The overlap between the MEF2C DEGs and the target genes of FMRP are a major driver of the genetic signals obtained for EA, IQ and SZ and both of these gene-sets are enriched in the same cell types. FMRP is a neuronal RNA-binding protein and its targets are polyribosomal mRNAs (28), which are thought to have their translation inhibited by FMRP binding. MEF2C and FMRP function together in a cell-autonomous mechanism to eliminate excitatory synapses (39). Therefore, it is reasonable to conclude that disruption of synapse elimination in the specific

glutamatergic excitatory neurons in the cortex, which we find to be enriched for both MEF2C DEGs and FMRP target genes, is fundamental to the pathophysiology of neurodevelopmental phenotypes and disorders.

Materials and Methods

Ethics statement

Data were directly downloaded from published studies and no additional ethics approval was needed. Each study is referenced and details on ethics approval are available in each manuscript.

MEF2C gene-set

There were 1076 DEGs reported in *Mef2c* cKO mice compared with controls with a log2 fold change >0.3 and at a FDR <0.05 (478 genes with increased expression level (up-regulated) in cKO mice and 598 genes with decreased expression level (down-regulated)) (13). This mouse DEG list was converted to human orthologue genes using Ensembl Biomart https://www.ensembl.org/biomart) or using with NCBI Gene database (https://www.ncbi.nlm.nih.gov/gene) if Biomart failed to find an orthologue. This resulted in a gene-set of 1,055 genes (465 up-regulated and 590 down-regulated; Supplementary Table 1). Sets of 'brain-expressed' genes (n=14,243) and 'brain-elevated' genes (n=1,424) were sourced from the Human Protein Atlas (https://www.proteinatlas.org/humanproteome/brain) and used as covariates in analyses. Brain-elevated genes are those that show an elevated expression in brain compared to other tissue types.

GWAS data

This MEF2C gene-set was tested for enrichment of genes associated with different neurodevelopmental phenotypes using GWAS summary statistics for SZ (40,675 cases and 64,643 controls) (20), ASD (18,381 cases and 27,969 controls) (19), IQ (269,867 individuals) (22) and EA (1,131,881 individuals) (24). For control purposes, we tested for enrichment using GWAS summary statistics for ten other phenotypes. These were attention deficit hyperactivity disorder (ADHD (40)), bipolar disorder (BPD (41)), major depressive disorder (MDD (42)), obsessive-compulsive disorder (OCD (43)), Alzheimer's

disease (AD (44)), Crohn's disease (CD (45)), height (46), stroke (47), type 2 diabetes (T2D (48)) and inflammatory bowel disease (IBD (49)).

eQTL data

Data from the following were used to identify eQTLs at *MEF2C*: PsychENCODE, adult human cortex (50) (http://resource.psychencode.org/); Genotype-Tissue Expression (GTEx), multiple brain regions (https://gtexportal.org/); BRAINEAC, multiple brain regions (http://www.braineac.org); BrainSeq, hippocampus and dorsolateral prefrontal cortex (http://eqtl.brainseq.org/phase2/); O'Brien et al. (51), human fetal brain. HaploReg v4.1 (52) (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php), links to eQTLs from multiple sources.

MAGMA gene-set analysis

A gene-set analysis (GSA) is a statistical method for simultaneously analysing multiple genetic markers in order to determine their joint effect. We performed GSA using MAGMA (53) and summary statistics from various GWAS. A MAGMA analysis involved three steps. First, in the annotation step we mapped SNPs with available GWAS results on to protein-coding genes (GRCh37/hg19 start-stop coordinates +/-20kb). Second, in the gene analysis step we computed gene P values for each GWAS dataset. This gene analysis is based on a multiple linear principal components regression model that accounts for LD between SNPs. The European panel of the 1000 Genomes data was used as a reference panel for LD. Third, a competitive GSA based on the gene P values, also using a regression structure, was used to test if the genes in a gene-set were more strongly associated with either phenotype than other genes in the genome. The MHC region is strongly associated in the SZ GWAS data. This region contains high LD and the association signal has been attributed to just a small number of independent variants (54). However, MAGMA still identifies a very large number of associated genes despite factoring in the LD

information. To avoid the excessive number of associated genes biasing the MAGMA GSA, we excluded all genes within the MHC region from our GSA of SZ. MAGMA was chosen because it corrects for LD, gene size and gene density (potential confounders) and has significantly more power than other GSA tools (55). The second step of MAGMA analysis generates a P value for each gene. By Bonferroni correcting for the number of genes tested in each GWAS, we used these data to identify individual genes that genome-wide associated with SZ (n=499), IQ (n=647), and EA (n=1,278).

Enrichment analysis for genes containing *de novo* mutations

Lists of genes harbouring DNMs identified in patients with SZ (n=1,024), ASD (n=3,985), ID (n=192) and in unaffected siblings (n=1,995) and controls (n=54) based on exome sequencing of trios were sourced from Genovese et al. (56). DNMs were categorized as silent, missense and loss-of-function (LoF). We tested for enrichment of our MEF2C gene-set in these gene lists using denovolyzeR (57). DenovolyzeR is an R package for the statistical analysis of DNMs. For each gene, denovolyzeR derives the expected number of DNMs in a given population based on the mutability of the gene and the number of trios sequenced. It then compares the observed number of DNMs against expectation using a Poisson framework to determine whether there is an excess of DNMs in a given gene or gene-set (57).

Enrichment analysis of single cell transcriptomic data from the mouse nervous system

The expression weighted cell-type enrichment (EWCE) R package (https://github.com/NathanSkene/EWCE) represents a method to statistically evaluate if a set of genes (e.g. our MEF2C gene-set) has higher expression within a particular cell type than can be reasonably expected by chance (58). The probability distribution for this is estimated by randomly generating gene-sets of equal length from a set of background genes. We used scRNA-seq data from 19 regions across the central and peripheral nervous system of post-natal day (P) 12-30 and 6-8 week old mice (29) and from nine regions of the adult (P60-70) mouse brain (30). Tissue analysed by Zeisel et al. (29) included the anterior, middle and posterior cortex, hippocampus, amygdala, striatum, thalamus, hypothalamus,

midbrain, cerebellum, dorsal root ganglion, myenteric plexus, submucosal plexus, muscle layer, sympathetic chain, spinal cord, pons, medulla and olfactory bulb. This resulted in 265 cell clusters represented by 160,796 single cell transcriptomes. These expression datasets were sourced from http://mousebrain.org/. Regions analysed by Saunders et al. (30) included the frontal and posterior cortex, hippocampus, thalamus, cerebellum, striatum, globus pallidus externus and nucleus basalis, entopeduncular nucleus and subthalamic nucleus, and substantia nigra and ventral tagmental area. This resulted in 565 transcriptionally distinct populations from 690,000 individual cells. These expression datasets were sourced from http://dropviz.org/. We used the EWCE package to test whether MEF2C DEGs are enriched in any particular cell type from the scRNA-seq data. The analysis is carried out in three steps. First, the relevant scRNA-seq data is loaded. Next, a target gene-set and suitable background set is chosen. Finally, a bootstrapping enrichment test is run, controlling for transcript length and GC-content. The package controls for these by selecting bootstrap lists containing genes with similar properties to genes in the target list. For our analysis we ran 100,000 permutations to generate the P values in Supplementary Tables 9-10. The package can convert gene IDs and provides a background set of mouse/human genes from Biomart. Genes in the target or background set without mouse/human orthologs were dropped. To counter the potential bias of using a MEF2C gene-set that represents a set of genes that are expressed in the somatosensory cortex, we generated and tested 1,000 randomized gene-sets derived from the Harrington dataset of 7,977 expressed genes (each containing n=1,055 genes expressed in the somatosensory cortex to match the size of our test MEF2C DEG gene-set). For each cell type, across the 1,000 analyses, we counted the number of times a random gene-set produced a SD from mean value greater than the test MEF2C DEG gene-set and used this to generate a simulated p value for each cell type (Supplementary Tables 9-10).

Functional annotation

ConsensusPathDB-human (http://cpdb.molgen.mpg.de/) was used to perform overrepresentation analysis of gene-sets and we report on enriched GO terms (59).

We have set up a GitHub page containing details on how the different analyses were run (https://github.com/laurafahey02/MEF2C-gene-set-analysis-DNM-analysis-cell-type-enrichment-analysis).

Acknowledgements

This work was supported by a NARSAD Independent Investigator Award from the Brain and Behavior Research Foundation to DWM (#25916) and a Government of Ireland Postgraduate Scholarship to LF (GOIPG/2018/1501).

Conflict of Interest Statement:

The authors declare no conflict of interest.

References

1 Li, H., Radford, J.C., Ragusa, M.J., Shea, K.L., McKercher, S.R., Zaremba, J.D., Soussou, W., Nie, Z., Kang, Y.J., Nakanishi, N. *et al.* (2008) Transcription factor MEF2C influences neural stem/progenitor cell differentiation and maturation in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 9397-9402.

2 Potthoff, M.J. and Olson, E.N. (2007) MEF2: a central regulator of diverse developmental programs. *Development (Cambridge, England)*, **134**, 4131-4140.

3 Consortium, U. (2015) UniProt: a hub for protein information. *Nucleic acids research*, **43**, D204-212.

4 Lyons, G.E., Micales, B.K., Schwarz, J., Martin, J.F. and Olson, E.N. (1995) Expression of mef2 genes in the mouse central nervous system suggests a role in neuronal maturation. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **15**, 5727-5738.

5 Haberland, M., Arnold, M.A., McAnally, J., Phan, D., Kim, Y. and Olson, E.N. (2007) Regulation of HDAC9 gene expression by MEF2 establishes a negative-feedback loop in the transcriptional circuitry of muscle differentiation. *Molecular and cellular biology*, **27**, 518-525.

6 Lin, Q., Schwarz, J., Bucana, C. and Olson, E.N. (1997) Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science (New York, N.Y.)*, **276**, 1404-1407.

7 Adachi, M., Lin, P.Y., Pranav, H. and Monteggia, L.M. (2016) Postnatal Loss of Mef2c Results in Dissociation of Effects on Synapse Number and Learning and Memory. *Biological psychiatry*, **80**, 140-148.

8 Chen, Y.C., Kuo, H.Y., Bornschein, U., Takahashi, H., Chen, S.Y., Lu, K.M., Yang, H.Y., Chen, G.M., Lin, J.R., Lee, Y.H. *et al.* (2016) Foxp2 controls synaptic wiring of corticostriatal circuits and vocal communication by opposing Mef2c. *Nature neuroscience*, **19**, 1513-1522.

9 Tu, S., Akhtar, M.W., Escorihuela, R.M., Amador-Arjona, A., Swarup, V., Parker, J., Zaremba, J.D., Holland, T., Bansal, N., Holohan, D.R. *et al.* (2017) NitroSynapsin therapy for a mouse MEF2C haploinsufficiency model of human autism. *Nature communications*, **8**, 1488.

10 Barbosa, A.C., Kim, M.S., Ertunc, M., Adachi, M., Nelson, E.D., McAnally, J., Richardson, J.A., Kavalali, E.T., Monteggia, L.M., Bassel-Duby, R. *et al.* (2008) MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 9391-9396.

11 Lipton, S.A., Li, H., Zaremba, J.D., McKercher, S.R., Cui, J., Kang, Y.J., Nie, Z., Soussou, W., Talantova, M., Okamoto, S. *et al.* (2009) Autistic phenotype from MEF2C knockout cells. *Science (New York, N.Y.)*, **323**, 208.

12 Cho, E.G., Zaremba, J.D., McKercher, S.R., Talantova, M., Tu, S., Masliah, E., Chan, S.F., Nakanishi, N., Terskikh, A. and Lipton, S.A. (2011) MEF2C enhances dopaminergic neuron differentiation of human embryonic stem cells in a parkinsonian rat model. *PloS one*, **6**, e24027.

13 Harrington, A.J., Raissi, A., Rajkovich, K., Berto, S., Kumar, J., Molinaro, G., Raduazzo, J., Guo, Y., Loerwald, K., Konopka, G. *et al.* (2016) MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. *Elife*, **5**.

14 Paciorkowski, A.R., Traylor, R.N., Rosenfeld, J.A., Hoover, J.M., Harris, C.J., Winter, S., Lacassie, Y., Bialer, M., Lamb, A.N., Schultz, R.A. *et al.* (2013) MEF2C Haploinsufficiency features consistent hyperkinesis, variable epilepsy, and has a role in dorsal and ventral neuronal developmental pathways. *Neurogenetics*, **14**, 99-111.

15 Rocha, H., Sampaio, M., Rocha, R., Fernandes, S. and Leao, M. (2016) MEF2C haploinsufficiency syndrome: Report of a new MEF2C mutation and review. *Eur J Med Genet*, **59**, 478-482.

16 Mikhail, F.M., Lose, E.J., Robin, N.H., Descartes, M.D., Rutledge, K.D., Rutledge, S.L., Korf, B.R. and Carroll, A.J. (2011) Clinically relevant single gene or intragenic deletions encompassing critical neurodevelopmental genes in patients with developmental delay, mental retardation, and/or autism spectrum disorders. *American journal of medical genetics. Part A*, **155A**, 2386-2396.

17 Le Meur, N., Holder-Espinasse, M., Jaillard, S., Goldenberg, A., Joriot, S., Amati-Bonneau, P., Guichet, A., Barth, M., Charollais, A., Journel, H. *et al.* (2010) MEF2C haploinsufficiency caused by

either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *Journal of medical genetics*, **47**, 22-29.

18 Mitchell, A.C., Javidfar, B., Pothula, V., Ibi, D., Shen, E.Y., Peter, C.J., Bicks, L.K., Fehr, T., Jiang, Y., Brennand, K.J. *et al.* (2017) MEF2C transcription factor is associated with the genetic and epigenetic risk architecture of schizophrenia and improves cognition in mice. *Mol Psychiatry*.

19 Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R. *et al.* (2019) Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics*, **51**, 431-444.

20 Pardinas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L. *et al.* (2018) Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature genetics*, **50**, 381-389.

21 Sniekers, S., Stringer, S., Watanabe, K., Jansen, P.R., Coleman, J.R.I., Krapohl, E., Taskesen, E., Hammerschlag, A.R., Okbay, A., Zabaneh, D. *et al.* (2017) Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature genetics*, **49**, 1107-1112.

22 Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B., Coleman, J.R.I. *et al.* (2018) Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature genetics*, **50**, 912-919.

23 Davies, G., Lam, M., Harris, S.E., Trampush, J.W., Luciano, M., Hill, W.D., Hagenaars, S.P., Ritchie, S.J., Marioni, R.E., Fawns-Ritchie, C. *et al.* (2018) Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature communications*, **9**, 2098.

24 Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R. *et al.* (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics*, **50**, 1112-1121.

25 Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., Turley, P., Chen, G.-B., Emilsson, V., Meddens, S.F.W. *et al.* (2016) Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, **533**, 539-542.

26 Zou, F., Chai, H.S., Younkin, C.S., Allen, M., Crook, J., Pankratz, V.S., Carrasquillo, M.M., Rowley, C.N., Nair, A.A., Middha, S. *et al.* (2012) Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet*, **8**, e1002707.

27 Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y. *et al.* (2018) Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*, **362**, eaat8127.

28 Darnell, Jennifer C., Van Driesche, Sarah J., Zhang, C., Hung, Ka Ying S., Mele, A., Fraser, Claire E., Stone, Elizabeth F., Chen, C., Fak, John J., Chi, Sung W. *et al.* (2011) FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. *Cell*, **146**, 247-261.

29 Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G. *et al.* (2018) Molecular Architecture of the Mouse Nervous System. *Cell*, **174**, 999-1014.e1022.

30 Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S. *et al.* (2018) Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell*, **174**, 1015-1030 e1016.

31 Skene, N.G., Bryois, J., Bakken, T.E., Breen, G., Crowley, J.J., Gaspar, H.A., Giusti-Rodriguez, P., Hodge, R.D., Miller, J.A., Muñoz-Manchado, A.B. *et al.* (2018) Genetic identification of brain cell types underlying schizophrenia. *Nature Genetics*, **50**, 825-833.

32 Arion, D., Huo, Z., Enwright, J.F., Corradi, J.P., Tseng, G. and Lewis, D.A. (2017) Transcriptome Alterations in Prefrontal Pyramidal Cells Distinguish Schizophrenia From Bipolar and Major Depressive Disorders. *Biological psychiatry*, **82**, 594-600.

33 Fazel Darbandi, S., Robinson Schwartz, S.E., Qi, Q., Catta-Preta, R., Pai, E.L., Mandell, J.D., Everitt, A., Rubin, A., Krasnoff, R.A., Katzman, S. *et al.* (2018) Neonatal Tbr1 Dosage Controls Cortical Layer 6 Connectivity. *Neuron*, **100**, 831-845 e837.

34 Tendolkar, I., Weis, S., Guddat, O., Fernandez, G., Brockhaus-Dumke, A., Specht, K., Klosterkotter, J., Reul, J. and Ruhrmann, S. (2004) Evidence for a dysfunctional retrosplenial cortex in patients with schizophrenia: a functional magnetic resonance imaging study with a semantic-perceptual contrast. *Neuroscience letters*, **369**, 4-8.
35 Mitelman, S.A., Shihabuddin, L., Brickman, A.M., Hazlett, E.A. and Buchsbaum, M.S. (2005) Volume of the cingulate and outcome in schizophrenia. *Schizophr Res*, **72**, 91-108.
36 Spreng, R.N., Mar, R.A. and Kim, A.S. (2009) The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of cognitive neuroscience*, **21**, 489-510.
37 Bluhm, R.L., Miller, J., Lanius, R.A., Osuch, E.A., Boksman, K., Neufeld, R.W., Theberge, J., Schaefer, B. and Williamson, P.C. (2009) Retrosplenial cortex connectivity in schizophrenia. *Psychiatry Res*, **174**, 17-23.
38 Hogeveen, J., Krug, M.K., Elliott, M.V. and Solomon, M. (2018) Insula-Retrosplenial Cortex Overconnectivity Increases Internalizing via Reduced Insight in Autism. *Biological psychiatry*, **84**, 287-294.
39 Pfeiffer, B.E., Zang, T., Wilkerson, J.R., Taniguchi, M., Maksimova, M.A., Smith, L.N., Cowan, C.W. and Huber, K.M. (2010) Fragile X Mental Retardation Protein Is Required for Synapse Elimination by the Activity-Dependent Transcription Factor MEF2. *Neuron*, **66**, 191-197.
40 Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M. *et al.* (2019) Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics*, **51**, 63-75.
41 Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H.A. *et al.* (2019) Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature genetics*, **51**, 793-803.
42 Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F. *et al.* (2018) Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetics*, **50**, 668-681.
43 International Obsessive Compulsive Disorder Foundation Genetics, C., Studies, O.C.D.C.G.A., Arnold, P.D., Askland, K.D., Barlassina, C., Bellodi, L., Bienvenu, O.J., Black, D., Bloch, M., Brentani, H. *et al.* (2017) Revealing the complex genetic architecture of obsessive–compulsive disorder using meta-analysis. *Molecular Psychiatry*, **23**, 1181.
44 Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B. *et al.* (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics*, **45**, 1452-1458.
45 Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R. *et al.* (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature genetics*, **42**, 1118-1125.
46 Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J.a., Kutalik, Z. *et al.* (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics*, **46**, 1173.
47 Traylor, M., Farrall, M., Holliday, E.G., Sudlow, C., Hopewell, J.C., Cheng, Y.C., Fornage, M., Ikram, M.A., Malik, R., Bevan, S. *et al.* (2012) Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *The Lancet. Neurology*, **11**, 951-962.
48 Mahajan, A., Go, M.J., Zhang, W., Below, J.E., Gaulton, K.J., Ferreira, T., Horikoshi, M., Johnson, A.D., Ng, M.C., Prokopenko, I. *et al.* (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature genetics*, **46**, 234-244.
49 Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T. *et al.* (2015) Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics*, **47**, 979-986.

50 Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T. *et al.* (2018) Comprehensive functional genomic resource and integrative model for the human brain. *Science*, **362**, eaat8464.
51 O'Brien, H.E., Hannon, E., Hill, M.J., Toste, C.C., Robertson, M.J., Morgan, J.E., McLaughlin, G., Lewis, C.M., Schalkwyk, L.C., Hall, L.S. *et al.* (2018) Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biology*, **19**, 194.
52 Ward, L.D. and Kellis, M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*, **40**, D930-934.
53 de Leeuw, C.A., Mooij, J.M., Heskes, T. and Posthuma, D. (2015) MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology*, **11**, e1004219.
54 Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V. *et al.* (2016) Schizophrenia risk from complex variation of complement component 4. *Nature*, **530**, 177-183.
55 de Leeuw, C.A., Neale, B.M., Heskes, T. and Posthuma, D. (2016) The statistical properties of gene-set analysis. *Nat Rev Genet*, **17**, 353-364.
56 Genovese, G., Fromer, M., Stahl, E.A., Ruderfer, D.M., Chambert, K., Landen, M., Moran, J.L., Purcell, S.M., Sklar, P., Sullivan, P.F. *et al.* (2016) Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nature neuroscience*, **19**, 1433-1441.
57 Ware, J.S., Samocha, K.E., Homsy, J. and Daly, M.J. (2015) Interpreting de novo Variation in Human Disease Using denovolyzeR. *Current protocols in human genetics*, **87**, 7.25.21-27.25.15.
58 Skene, N.G. and Grant, S.G. (2016) Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. *Front Neurosci*, **10**, 16.
59 Herwig, R., Hardt, C., Lienhard, M. and Kamburov, A. (2016) Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nat. Protocols*, **11**, 1889-1907.

Legends to Figures

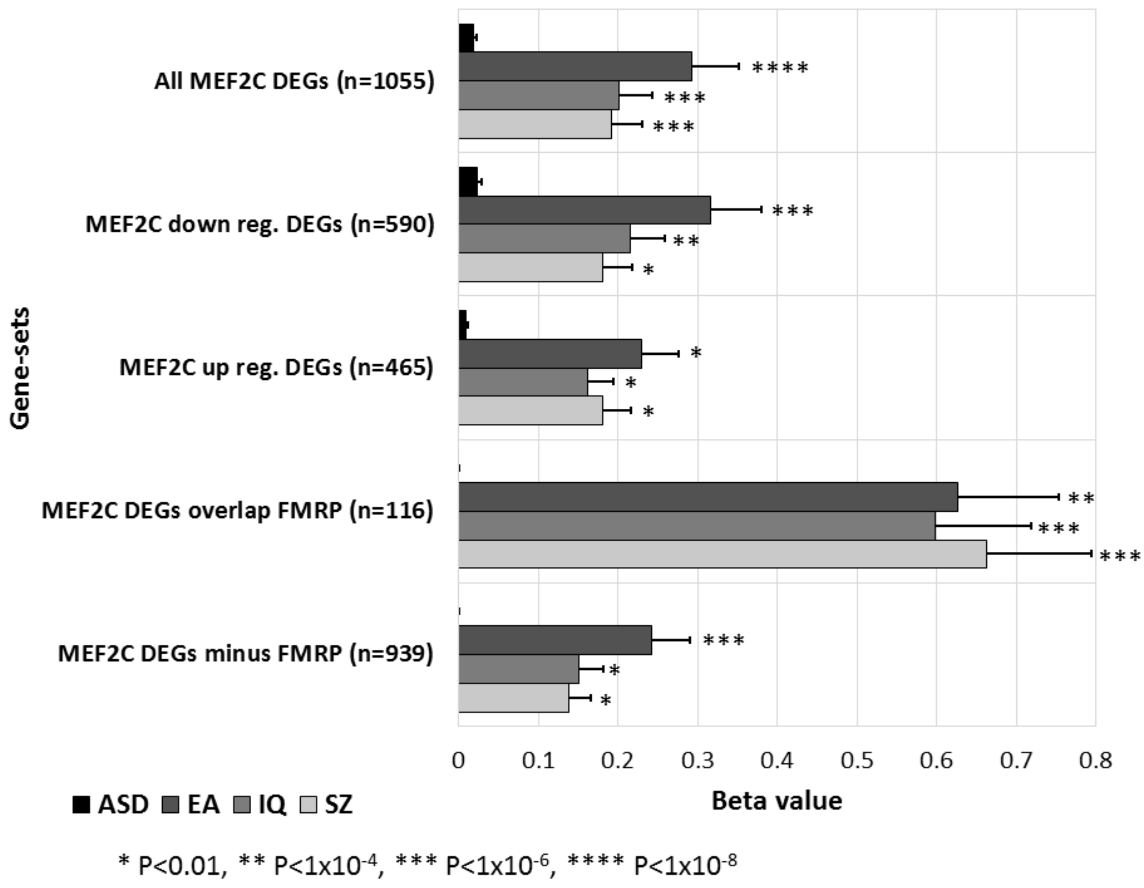

Figure 1: Gene-set analysis of the MEF2C gene-set in ASD, EA, IQ and SZ. Five gene-sets were tested as identified on the y axis along with the number of genes in each gene-set (n). The ASD gene-set was not included in the analyses using the FMRP target genes as it showed no evidence of enrichment in the first three analyses of MEF2C DEGs. Beta values (effect sizes) as calculated by MAGMA are plotted on the x-axis with P-value categories shown beside each bar. Horizontal bars indicate standard error. All data for this graph are in Supplementary Table 2.

Tables

Table 1: De novo mutation analysis of the MEF2C gene-set

Autism Spectrum Disorder (3,985 trios)				
Mutation class	**# Observed Mutations**	**# Expected Mutations**	**Enrichment value**	**P value[a]**
Silent	69	81.2	0.85	0.924
Missense	212	177.2	1.2	0.0060
LoF	45	24.0	1.88	**$7.96\text{x}10^{-05}$**
Intellectual Disability (192 trios)				
Mutation class	**# Observed Mutations**	**# Expected Mutations**	**Enrichment value**	**P value[a]**
Silent	4	3.9	1.02	0.549
Missense	25	8.5	2.93	**$3.57\text{x}10^{-06}$**
LoF	2	1.2	1.73	0.321
Schizophrenia (1,024 trios)				
Mutation class	**# Observed Mutations**	**# Expected Mutations**	**Enrichment value**	**P value[a]**
Silent	19	20.9	0.911	0.688
Missense	49	45.5	1.08	0.322
LoF	5	6.2	0.812	0.735
Unaffected siblings (1,995 trios)				
Mutation class	**# Observed Mutations**	**# Expected Mutations**	**Enrichment value**	**P value[a]**
Silent	39	40.7	0.959	0.623
Missense	93	88.7	1.05	0.338
LoF	18	12.0	1.5	0.0627
Controls (54 trios)				
Mutation class	**# Observed**	**# Expected**	**Enrichment value**	**P value[a]**
Missense	1	2.4	0.417	0.909

[a] P-values in bold survived Bonferroni multiple test correction for the 13 tests.

Table 2: Enriched GO terms for genes in the MEF2C gene-set contributing to different phenotypes

Gene Subset	# Genes	Category[a]	Selected Enriched Terms of Shared Ancestry (GO:IDs)
MEF2C-SZ	41	BP	neurogenesis (0022008), regulation of nervous system development (0051960), neuron differentiation (0030182), neuron development (0048666), trans-synaptic signalling (0099537), synaptic signalling (0099536)
MEF2C-IQ	37	CC	postsynapse (0098794), neuron to neuron synapse (0098984), presynaptic cytoskeleton (0099569),
		BP	neuron differentiation (0030182), neuron development (0048666), trans-synaptic signalling (0099537)
MEF2C-EA	109	CC	synaptic membrane (0097060), integral component of synaptic membrane (0099699), plasma membrane region (0098590), intrinsic component of synaptic membrane (0099240)
		BP	anterograde trans-synaptic signalling (0098916), trans-synaptic signalling (0099537), synaptic signalling (0099536), cell-cell signalling (0007267), nervous system development (0007399)
MEF2C-ASD	45	BP	cell projection morphogenesis (0048858), cell morphogenesis (0000902), cell part morphogenesis (0032990), cellular component morphogenesis (0032989), neuron differentiation (0030182), neuron development (0048666)
MEF2C-ID	25	BP	forebrain neuron fate commitment (0021877), neuron differentiation (0030182), cell morphogenesis (0000902), cell morphogenesis involved in differentiation (0000904), neuron development (0048666)

[a] BP = biological process; CC = cellular compartment

Table 3: Enriched cell types from analysis of scRNA-seq data using the MEF2C gene-set

Zeisel et al. 2018 study of 265 cell types from the mouse nervous system			
Cell Type	**Neurotransmitter(s)**	**# Cell Types Enriched**	**Region - Locations**
Excitatory neurons	Glutamate	15	Cortex - Cortical pyramidal (layers 4,5,6); Lateral cortex layer 6; Cingulate/Retrosplenial area (layers 2,5,6); Piriform pyramidal; Subiculum; Anterior olfactory nucleus.
Excitatory neurons	Glutamate, Nitric Oxide	2	Cortex - Piriform pyramidal; Subiculum.
Excitatory neurons	Glutamate, Nitric Oxide	2	Hippocampus - CA1
Excitatory neurons	Glutamate	1	Medulla - Spinal nucleus of the trigeminal.
Excitatory neurons	Glutamate	1	Midbrain - Superior colliculus, sensory related, deep layer.
Excitatory neurons	Glutamate	1	Thalamus - Posterior thalamus.
Granule neuroblasts	Glutamate	1	Dentate gyrus.
D2 medium spiny neurons	GABA	2	Striatum - Dorsal; Ventral.
Neurogliaform cells	GABA	1	Hippocampus, Cortex.
Inner horizontal cell	GABA	1	Olfactory bulb - Mitral cell layer.
Saunders et al. 2018 study of 565 cell types from the mouse brain			
Cell Type	**Neurotransmitter(s)**	**# Cell Types Enriched**	**Region - Locations**
Excitatory neurons	Glutamate	5	Frontal Cortex - Deep layer pyramidal cells; Superficial layer pyramidal cells.
Excitatory neurons	Glutamate	13	Posterior Cortex - Layers 2/3, 5; Subplate pyramidal cells; Subiculum, Entorhinal cortex; Retrosplenial cortex; Claustrum.
Excitatory neurons	Glutamate	10	Hippocampus - CA1 Principal cells; Anterior Subiculum, proximal to CA1; Medial entorhinal cortex.
Inhibitory neurons	GABA	1	Striatum - Indirect spiny projection neurons.

Abbreviations

AD	Alzheimer’s disease
ADHD	attention deficit hyperactivity disorder
ASD	autism spectrum disorder
BPD	bipolar disorder
cKO	conditional knockout
CA1	cornu ammonis 1
CD	Crohn’s disease
CNS	central nervous system
DEG	differentially expressed gene
DNM	*de novo* mutation
EA	educational attainment
eQTL	expression quantitative trait locus
EWCE	expression weighted cell-type enrichment
FMRP	Fragile X mental retardation protein
GO	Gene Ontology
GSA	gene-set analysis
GTEx	Genotype-Tissue Expression
GWAS	genome-wide association studies
HDAC	histone deacetylase
IBD	inflammatory bowel disease
ID	intellectual disability
IQ	human intelligence
LD	linkage disequilibrium
LoF	loss-of-function
MEF2C	myocyte enhancer factor 2 C
MDD	major depressive disorder
MSN	medium spiny neuron
OCD	obsessive-compulsive disorder
P	post-natal day
scRNA-seq	single cell RNA sequencing
SNP	single nucleotide polymorphism
SZ	schizophrenia
T2D	type 2 diabetes
TBR1	transcription factor T-brain-1
WT	wild-type