# Reproducibility of Structural Brain Alterations Associated with Transdiagnostic Risk for Mental Illness: Evidence from a Population-Representative Birth Cohort

Adrienne L. Romer, MA[1,2], Annchen R. Knodt, MS[1,2], Maria L. Sison, BS[1], David Ireland, PhD[3], Sandhya Ramrakha, PhD[3], Richie Poulton, PhD[3], Ross Keenan, MB,Ch.B[4], Tracy R. Melzer, PhD[5,6], Terrie E. Moffitt, PhD[2,7,8,9], Avshalom Caspi, PhD[2,7,8,9], & Ahmad R. Hariri, PhD[1,2]

[1]*Laboratory of NeuroGenetics, Duke University, Durham, NC, USA*

[2]*Department of Psychology & Neuroscience, Duke University, Durham, NC, USA*

[3]*Dunedin Multidisciplinary Health and Development Research Unit, Department of Psychology, University of Otago, Dunedin, New Zealand*

[4]*Christchurch Radiology Group, Christchurch, New Zealand*

[5]*Department of Medicine, University of Otago, Christchurch, NZ*

[6]*New Zealand Brain Research Institute, Christchurch, New Zealand*

[7]*Department of Psychiatry & Behavioral Sciences, Duke University School of Medicine, Durham, NC, USA*

[8]*Center for Genomic and Computational Biology, Duke University, Durham, NC, USA*

[9]*Social, Genetic, and Developmental Psychiatry Research Center, Institute of Psychiatry, Psychology & Neuroscience, King's College London, England*

Address Correspondence to:

Adrienne L. Romer, Laboratory of NeuroGenetics, Department of Psychology and Neuroscience, Duke University, 2020 West Main St., Ste. 30, Durham, NC 27705. (Tel.: 484-557-1802; Fax: 919-660-5726; email: adrienne.romer@duke.edu).

**Abstract**

Transdiagnostic research has identified a general psychopathology factor – often called the '*p*' factor – that accounts for shared liability to internalizing, externalizing, and thought disorders in diverse samples. It has been argued that the *p* factor may reflect dysfunctional thinking present in serious mental illness. In support of this, we previously used a theory-free, data-driven multimodal neuroimaging approach to find that higher *p* factor scores are associated with structural deficits within a cerebello-thalamo-cortical circuit (CTCC) and visual association cortex, both of which are important for monitoring and coordinating information processing in the service of executive control. Here we attempt to replicate these associations by conducting region-of-interest analyses of the CTCC and visual association cortex using data from 831 members of the Dunedin Multidisciplinary Health and Development Study, a five-decade longitudinal study of a population-representative birth cohort now 45 years old. We further sought to replicate a more recent report that *p* factor scores can be predicted by patterns of distributed cerebellar morphology as estimated through independent component analysis. We successfully replicated associations between higher *p* factor scores and both reduced grey matter volume of the visual association cortex and fractional anisotropy of pontine white matter pathways within the CTCC. In contrast, we failed to replicate prior associations between cerebellar structure and *p* factor scores. Collectively, our findings encourage further focus on the CTCC and visual association cortex as core neural substrates and potential biomarkers of transdiagnostic risk for mental illness.

**Introduction**

A rapidly emerging body of research has identified a general factor that captures broad risk for experiencing all common forms of psychopathology across diverse samples (1). This general psychopathology or '*p*' factor (2) accounts for the high rates of comorbidity among internalizing, externalizing, and thought disorders. One compelling argument regarding the nature of the *p* factor is that it captures the extent of disordered or dysfunctional thinking present not only in thought disorders, but also in extreme presentations of internalizing and externalizing disorders (3). Consistent with this argument, we recently used a theory-free, data-driven approach to find that higher *p* factor scores are associated with structural deficits in a cerebello-thalamo-cortico circuit (CTCC) critical for monitoring and coordinating information processing in the service of executive control, which is impaired to a greater or lesser extent across all common forms of psychopathology (4).

Specifically, we found that higher *p* factor scores are associated with reduced grey matter volume in neocerebellar lobule VIIb. We also found evidence for decreased microstructural integrity of pontine white matter pathways, as indexed by lower fractional anisotropy (FA). These pontine pathways are crucial for communication between the cerebrum, particularly the prefrontal cortex, and neocerebellum (5), supporting higher-order cognitive and emotional processing (6–9). In particular, the CTCC has been theorized to compare the intention with the execution of thoughts, emotions, and actions by continuously updating internal models (10,11). Moreover, CTCC dysfunction has been consistently reported in disorders principally characterized by poor executive control and disorganized thought such as schizophrenia (e.g., 12,13), and individuals with cerebellar cognitive affective syndrome following damage to the neocerebellum experience executive control dysfunction symptoms referred to as "dysmetria of thought" (14–16).

In addition to these structural deficits within the CTCC, we found novel evidence for decreased grey matter volume in the visual association cortex of individuals with higher *p* factor scores (4). Subsequently, we found that higher *p* factor scores are associated with patterns of inefficient

intrinsic functional connectivity between visual association cortex and networks supporting executive control and self-referential processes, which are often impaired across categorical disorders (17). Collectively, these patterns are consistent with speculation that higher *p* factor scores ultimately represent the likelihood of experiencing disordered thought through a diminished capacity for basic monitoring and processing of information supported by the CTCC and connectome-wide intrinsic functional connectivity.

It is important to seek to replicate these associations, especially because our original associations were discovered in a convenience sample of high-functioning 18 to 22 year-old university students through the Duke Neurogenetics Study (4). Here we attempt to replicate our original associations between CTCC and visual association cortex structural deficits and higher *p* factor scores using data from 831 members of the Dunedin Multidisciplinary Health and Development Study, a five-decade longitudinal study of a population-representative birth cohort now 45 years old. We further sought to replicate a more recent report that *p* factor scores can be predicted by patterns of distributed cerebellar morphology as estimated through independent component analysis of data from 8 to 23 year-old community volunteers through the Philadelphia Neurodevelopmental Cohort (18).

## Materials and Methods

### *Participants*

Data were available from 831 (415 females) 45 year-old study members who completed neuroimaging assessments as part of the ongoing Dunedin Multidisciplinary Health and Development Study. The Dunedin Study is a longitudinal investigation of health and behavior in a complete birth cohort of 1,037 individuals (91% of eligible births; 52% male) born between April 1972 and March 1973 in Dunedin, New Zealand (NZ), and eligible based on residence in the province and participation in the first assessment at age three. The cohort represents the full range of socioeconomic status on NZ's South Island and matches the NZ National Health and Nutrition Survey on key health indicators (19). The cohort is primarily white; fewer than 7% self-identify as having non-Caucasian ancestry, matching the South Island (19). Assessments

were carried out at birth and ages 3, 5, 7, 9, 11, 13, 15, 18, 21, 26, 32, 38, and most recently 45 years. Of 997 still-living cohort members, 831 (83%) have thus far completed MRI scanning in 2017-2019 during their age-45 assessment. Scanned Study members resembled still-living cohort members on their childhood IQ (M=100 [standard error=.5] vs. M=101 [standard error=.5) and their family-of-origin's socio-economic status (20), coded on a 6-point scale ranging from unskilled laborer to professional (M=3.76 [standard error=.04] vs. M=3.75 [standard error=.04]).

*Measuring the General Factor of Psychopathology, the p factor*

The Dunedin Study longitudinally ascertains mental disorders every 2 to 6 years, interviewing members about past-year symptoms. Psychopathology symptoms were assessed through private structured interviews using the Diagnostic Interview Schedule (21) at 18, 21, 26, 32, and 38 years of age. Interviewers were health-care professionals. We studied Diagnostic and Statistical Manual of Mental Disorders (DSM)–defined symptoms of the following disorders that were repeatedly assessed in our longitudinal study: alcohol dependence, cannabis dependence, dependence on hard drugs, tobacco dependence (assessed with the Fagerström Test for Nicotine Dependence; 22), conduct disorder, major depression, generalized anxiety disorder, fears and/or phobias, obsessive compulsive disorder, mania, and positive and negative schizophrenia symptoms. Ordinal measures represented the number of possible DSM-defined symptoms associated with each disorder. Fears and/or phobias were assessed as the count of diagnoses for simple phobia, social phobia, agoraphobia, and panic disorder that a study member reported at each assessment. Symptoms were assessed without regard for hierarchical exclusionary rules to facilitate the examination of comorbidity. Each of the 11 disorders were assessed at least 3 times. The past-year prevalence rates of psychiatric disorders in the Dunedin cohort are similar to prevalence rates in nationwide surveys of the United States and New Zealand (23,24).

The method used to compute a general factor of psychopathology in the Dunedin cohort have been described previously (2). Briefly, we used confirmatory factor analysis to compute a bi-

factor model specifying a general psychopathology factor (labeled *p*). This model included our 11 observed variables: alcohol dependence, cannabis dependence, dependence on hard drugs, tobacco dependence, conduct disorder, major depression, generalized anxiety disorder, fears/phobias, obsessive-compulsive disorder, mania, and positive and negative schizophrenia symptoms. Each model also included method/state factors designed to pull out age-and assessment-related variance (e.g., interviewer effects, mood effects, and age-specific vulnerabilities) that was uncorrelated with trait propensity toward psychopathology. All analyses were performed in MPlus version 7.12 using the weighted least squares means and variance adjusted (WLSMV) algorithm. The bi-factor model indicates that the ordinal symptom measures reflect both general psychopathology (the *p* factor) and three narrower styles of psychopathology (internalizing, externalizing, and thought disorders).

General psychopathology is represented by a *p* factor that directly influences all of the diagnostic symptom factors. In addition, styles of psychopathology are represented by three factors, each of which influences a smaller subset of the symptom items. For example, depressive symptoms load jointly on the *p* factor and on the internalizing style factor. The specific factors represent the constructs of internalizing, externalizing, and thought disorder over and above *p*. After respecification for a Heywood case, the bi-factor model fit the data well: $\chi2(1012, N = 1000) = 1652.586$, CFI = .966, TLI = .963, RMSEA = .025, 90% confidence internal (CI) = [.023, .027]. Loadings on the *p* factor were high (all p's<.001) and averaged .650. For expository purposes, we scaled Study members' *p* factor scores to M=100, SD=15. The *p* factor allows us to test for structural brain alterations in relation to general psychopathology.

*MRI Data Acquisition*

Each study member was scanned using a Siemens Skyra 3T scanner equipped with a 64-channel head/neck coil at the Pacific Radiology imaging center in Dunedin, New Zealand. Diffusion-weighted images providing full brain coverage were acquired with 2.5 mm isotropic resolution and 64 diffusion weighted directions (4700 ms repetition time, 110.0 ms echo time, b value 3,000 s/mm2, 240 mm field of view, 96×96 acquisition matrix, slice thickness = 2.5 mm). Non-

weighted (b = 0) images were acquired in both the encoding (AP) and reverse encoding (PA) directions to allow for EPI distortion correction. High resolution structural images were obtained using a T1-weighted MP-RAGE sequence with the following parameters: TR = 2400 ms; TE = 1.98 ms; 208 sagittal slices; flip angle, 9°; FOV, 224 mm; matrix = 256×256; slice thickness = 0.9 mm with no gap (voxel size 0.9×0.875×0.875 mm); and total scan time = 6 min and 52 s. All neuroimaging data were visually inspected for quality. Data were excluded for study members who were unable to be scanned with the 64-channel head coil, had an incidental finding, or whose scans were of poor quality due to motion (as revealed by visual inspection for T1-weighted images or >3 mm frame-to-frame movements for diffusion images), resulting in a total of 811 study members eligible for diffusion analyses and 815 study members eligible for grey matter volume analyses.

*Fractional Anisotropy*

Following the methods of Romer et al. (2018), diffusion tensor imaging analyses were completed using SPM8 implemented in Matlab R2016a. All diffusion weighted scans were motion corrected and co-registered to the mean image to correct for head movement. The tensor model was used to calculate FA values for each voxel and non-brain tissue was removed. Each image was normalized to MNI space and smoothed using a 4 mm FWHM Gaussian kernel. We note that the tensor model for derivation of FA values is not optimized for our current diffusion-weighted image data (25), which was acquired with b = 3000 s/mm2 to facilitate future probabilistic tractography. We are unaware of any suitable alternatives for the derivation of FA values at higher b values. Moreover, these differences in acquisition parameters are of less concern because visual inspection of the preprocessed images revealed adequate registration and we did successfully replicate the association between higher *p* factor scores and lower pontine FA (see below).

*Grey Matter Volume*

Again, following the methods of Romer et al. (2018) regional grey matter volumes were determined using the unified segmentation (26) and DARTEL normalization (27) modules in

SPM12 (http://www.fil.ion.ucl.ac.uk/spm). Using this approach, individual T1-weighted images were segmented into grey, white, and CSF images, and then non-linearly registered to the existing IXI template of 550 healthy subjects averaged in standard Montreal Neurological Institute (MNI) space, available with VBM8 (http://dbm.neuro.uni-jena.de/vbm/). Subsequently, grey matter images were modulated for nonlinear effects of the high-dimensional normalization to preserve the total amount of signal from each region and smoothed with an 8 mm FWHM Gaussian kernel. The voxel size of processed images was 1.5×1.5×1.5 mm. A grey matter mask for subsequent analyses was created by thresholding the final stage (6th) IXI template at 0.1.

*Cerebellar Grey Matter Volume*

In addition to the above whole-brain voxel based grey matter volume analyses, the Spatially Unbiased Infratentorial Toolbox (SUIT) was used for high-resolution cerebellar-specific voxel-based morphometry analyses per the methods of Romer et al. (2018). For each Study member, the Isolate function of the toolbox was used to create a mask of the cerebellum and generate grey and white matter segmentation maps. The masked segmentation maps were then normalized to the SUIT template with non-linear DARTEL normalization. The resulting cerebellar grey matter image was resliced into the SUIT atlas space and smoothed with a 4 mm FWHM isotropic Gaussian kernel, a small kernel to preserve precision in the definition of cerebellar structures, in line with previous publications (28).

*Independent Component Analysis of Cerebellar Morphology*

Lastly, we conducted an independent component analysis (ICA) of SUIT-based cerebellar morphology using the method of Moberget et al. (2019). Briefly, we masked the SUIT-derived cerebellar grey matter maps using the SUIT toolbox's grey matter probability map thresholded at 0.1 and subjected them to ICA using FSL MELODIC (29). In our sample, a model order of 9 corresponded to the highest number of clearly bilateral components, and this model was used for further analyses.

*Statistical Analyses*

Exact masks were created from the three primary associations with *p* factor scores originally reported in Romer et al. (2018): a 272 voxel cluster in the pons, a 2353 voxel cluster in the visual association cortex, and a 706 voxel cluster in the cerebellum. A fourth mask was created for the 156 voxel cluster in neocerebellar lobule VIIb identified through the SUIT analysis. Moving to the Dunedin Study data, mean values for each of these four masks were extracted for each Study member from the FA (pons), grey matter volume (visual association cortex and neocerebellum), and SUIT maps, respectively. These mean extracted values were then used as the dependent variable in linear models with *p* factor scores, sex, and total intracranial volume or average total FA, respectively, as predictors to explicitly test for replication of the original findings of Romer et al. (2018).

Per the strategy of Moberget et al. (2019), we also tested whether weights on our nine ICA-derived cerebellar components could predict *p* factor scores using shrinkage linear regression with 10,000 iterations of 10-fold cross-validation on randomly partitioned data. In contrast to Moberget et al., we did not include sex and total intracranial volume in the regression models, since there was no effect of sex on any of the components, and only one component was marginally associated with total volume ($p = 0.0184$ for component 5, not corrected for the 9 tests run). Performance was evaluated by comparing the distribution of Pearson correlations between predicted and observed *p* factor scores to a null distribution of correlations obtained by randomly permuting the *p* factor scores.

**Results**

*White Matter Microstructural Integrity*

A significant negative correlation (standardized $\beta = -0.0873$; $p = 0.00995$) indicated an association between lower pontine FA and higher *p* factor scores (Figure 1A), replicating the finding of Romer et al. (2018).

*Grey Matter Volume*

A significant negative correlation (standardized $\beta$ = -0.0774; p = 0.00717) indicated an association between lower visual association cortex grey matter volume and higher *p* factor scores (Figure 1B), replicating the finding of Romer et al. (2018). An observed negative correlation between cerebellar grey matter volume and *p* factor scores was not statistically significant (standardized $\beta$ = -0.0445; p = 0.125; Figure 1C). This was also true for the SUIT-based neocerebellar lobule VIIb cluster (standardized $\beta$ = -0.0234; p = 0.4998; Figure 1D).

*ICA-Derived Cerebellar Morphology*

The nine independent components of cerebellar morphology collectively accounted for 41.63% of the total variance in the modulated grey matter maps used as input in our *p* factor analysis; each component explained between 4.03% and 5.03% of the total variance (and between 9.68% and 12.09% of the explained variance). The nine ICA-derived components predicted *p* factor scores beyond chance on average, but the difference from the empirical null distribution was not significant (mean correlations between predicted and observed values: r = 0.13, p = 0.33; mean r > 87.38% of the empirical null distribution; Figure 2).

**Discussion**

We successfully replicated two prior associations between variation in brain structure and *p* factor scores using data from a population-representative birth cohort now in midlife. Namely, we replicated associations between *p* factor scores and both pontine FA and visual association cortex GMV as originally reported by Romer et al. (2018). In contrast, we failed to replicate three prior associations between cerebellar structure and *p* factor scores. First, although nominally consistent with the original report of Romer et al. (2018), neither of two tested associations between GMV in a broad cerebellar cluster nor a smaller cluster in neocerebellar lobule VIIb were statistically significant. Second, an ICA-based measure of global cerebellar morphology did not significantly predict *p* factor scores above chance as was reported originally by Moberget et al. (2019).

The replication of a negative association between pontine FA and *p* factor scores further implicates deficits in the CTCC in broad risk for psychopathology. Thus, dysfunction in fundamental aspects of monitoring and coordinating executive functions (i.e., "forward control") through dynamic information processing between the cerebellum and prefrontal cortex through the CTCC appears to be a core feature of transdiagnostic risk. The second replication of a negative association between *p* factor scores and GMV in visual association cortex is consistent with the importance of executive dysfunction in broad risk for psychopathology. In particular, structural deficits in visual association cortex may manifest as more effortful or less efficient integration of bottom-up sensory information with attentional demands and executive control processes in those at higher risk for mental illness (17).

The non-significant associations between *p* factor scores and multiple indices of cerebellar grey matter volume and morphology do not necessarily undermine the importance of the CTCC in transdiagnostic risk for psychopathology. Rather these failures may indicate that higher risk as captured by the *p* factor is more a reflection of how information is communicated within the CTCC, particularly through pontine white matter pathways connecting the cerebellum and prefrontal cortex, and less a reflection of how information may be locally computed within the neocerebellum. This would be consistent with the emerging understanding that brain function may be best characterized by distributed patterns of network communication rather than discrete regional activity (30,31). However, there also are pragmatic factors that may have limited our ability to replicate prior associations between *p* factor scores and cerebellar structure.

First, the failure to replicate cerebellar associations with *p* factor scores may reflect different contributions of brain structure to risk across development. The discovery samples were comprised of young adults (4) or children, adolescents, and young adults (18). In contrast, our sample is comprised of individuals in midlife. Thus, the contribution of cerebellar GMV and morphology to the *p* factor may be greater earlier than later in life. This difference may reflect the still-active structural development of the cerebellum, which parallels that of the prefrontal

cortex, in both discovery samples (32). Developmental differences are hinted at by the observation of only nine independent components of cerebellar morphology in our sample but ten such components in the sample studied by Moberget et al. (2019). Longitudinal assessment of brain structure and *p* factor scores within the same individuals is necessary to evaluate a hypothesis of developmental differences (33). Second, the nature of the sampling strategy across the three samples also may influence replication. Unlike the population-representative birth cohort in our current study, both discovery samples represented narrow groups of select individuals (e.g., high-functioning university students or community volunteers). Additional replication efforts across diverse samples are necessary to probe the implications of such possible differences for the study of the brain basis of transdiagnostic risk. Lastly, we may simply have been underpowered to identify significant associations of small effect. Our current sample is smaller than either the discovery sample of Romer et al. (N=1227) or Moberget et al. (N=1401). Generally, successful replication is more likely if the test samples are larger and thus better powered to detect often smaller effects than reported in a discovery sample (34,35).

These limitations notwithstanding, the two reproducible associations of the theory-free, data-driven findings of Romer et al. (2018) reported herein point to specific features of brain structure that may be a core feature of transdiagnostic risk for common mental illness across populations. Deficits in the microstructural integrity of pontine white matter pathways may reflect dysfunction of executive control processes supported through dynamic communication within the CTCC. Likewise, deficits in GMV of visual association cortex may reflect impairments in the integration of bottom-up sensory information with top-down executive control and attentional processes. Notably, both of these neuroanatomical features are consistent with a model of the *p* factor as indexing increasingly disordered thought, which characterizes the most debilitating forms of mental illness.

**Acknowledgments**

The Dunedin Multidisciplinary Health and Development Study is supported by the NZ HRC and NZ MBIE. This research was supported by National Institute on Aging grant R01AG032282, R01AG049789, and UK Medical Research Council grant MR/P005918/1. Additional support was provided by the Jacobs Foundation. ALR received support from the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106401. Thank you to members of the Advisory Board for the Dunedin Neuroimaging Study. We thank the Dunedin Study members, unit research staff, Pacific Radiology staff, and Study founder Phil Silva. The study protocol was approved by the institutional ethical review boards of the participating universities. Study members gave informed consent before participating.

**Conflict of Interest**

The authors declare no conflict of interest.

**References**

1. Lahey BB, Krueger RF, Rathouz PJ, Waldman ID, Zald DH. A hierarchical causal taxonomy of psychopathology across the life span. Psychol Bull. 2017 Feb;143(2):142–86.

2. Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, et al. The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? Clin Psychol Sci. 2014 Mar;2(2):119–37.

3. Caspi A, Moffitt TE. All for One and One for All: Mental Disorders in One Dimension. Am J Psychiatry. 2018 Apr 6;appiajp201817121383.

4. Romer AL, Knodt AR, Houts R, Brigidi BD, Moffitt TE, Caspi A, et al. Structural alterations within cerebellar circuitry are associated with general liability for common mental disorders. Mol Psychiatry. 2018 Apr;23(4):1084–90.

5. Middleton FA, Strick PL. Cerebellar projections to the prefrontal cortex of the primate. J Neurosci Off J Soc Neurosci. 2001 Jan 15;21(2):700–12.

6. Buckner RL. The Cerebellum and Cognitive Function: 25 Years of Insight from Anatomy and Neuroimaging. Neuron. 2013 Oct 30;80(3):807–15.

7. D'Angelo E, Casali S. Seeking a unified framework for cerebellar function and dysfunction: from circuit operations to cognition. Front Neural Circuits [Internet]. 2013 [cited 2019 Mar 2];6. Available from: http://journal.frontiersin.org/article/10.3389/fncir.2012.00116/abstract

8. Keren-Happuch E, Chen S-HA, Ho M-HR, Desmond JE. A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. Hum Brain Mapp. 2014 Feb;35(2):593–615.

9. Marek S, Siegel JS, Gordon EM, Raut RV, Gratton C, Newbold DJ, et al. Spatial and Temporal Organization of the Individual Human Cerebellum. Neuron. 2018 Nov 21;100(4):977-993.e7.

10. Ito M. Movement and thought: identical control mechanisms by the cerebellum. Trends Neurosci. 1993 Nov;16(11):448–50; discussion 453-454.

11. Ito M. Control of mental activities by internal models in the cerebellum. Nat Rev Neurosci. 2008 Apr;9(4):304–13.

12. Bernard JA, Mittal VA. Dysfunctional Activation of the Cerebellum in Schizophrenia: A Functional Neuroimaging Meta-Analysis. Clin Psychol Sci. 2015 Jul 1;3(4):545–66.

13. Moberget T, Doan NT, Alnæs D, Kaufmann T, Córdova-Palomera A, Lagerberg TV, et al. Cerebellar volume and cerebellocerebral structural covariance in schizophrenia: a

multisite mega-analysis of 983 patients and 1349 healthy controls. Mol Psychiatry. 2018 Jun;23(6):1512.

14. Andreasen NC, Paradiso S, O'Leary DS. "Cognitive Dysmetria" as an Integrative Theory of Schizophrenia: A Dysfunction in Cortical-Subcortical-Cerebellar Circuitry? Schizophr Bull. 1998 Jan 1;24(2):203–18.

15. Schmahmann JD. Disorders of the cerebellum: ataxia, dysmetria of thought, and the cerebellar cognitive affective syndrome. J Neuropsychiatry Clin Neurosci. 2004;16(3):367–78.

16. Schmahmann JD, Weilburg JB, Sherman JC. The neuropsychiatry of the cerebellum - insights from the clinic. Cerebellum Lond Engl. 2007;6(3):254–67.

17. Elliott ML, Romer A, Knodt AR, Hariri AR. A Connectome Wide Functional Signature of Transdiagnostic Risk for Mental Illness. Biol Psychiatry [Internet]. 2018 Apr 10; Available from: http://www.sciencedirect.com/science/article/pii/S0006322318314161

18. Moberget T, Alnæs D, Kaufmann T, Doan NT, Córdova-Palomera A, Norbom LB, et al. Cerebellar grey matter volume is associated with cognitive function and psychopathology in adolescence. Biol Psychiatry [Internet]. 2019 Jan 30 [cited 2019 Mar 2];0(0). Available from: https://www.biologicalpsychiatryjournal.com/article/S0006-3223(19)30065-4/abstract

19. Poulton R, Moffitt TE, Silva PA. The Dunedin Multidisciplinary Health and Development Study: overview of the first 40 years, with an eye to the future. Soc Psychiatry Psychiatr Epidemiol. 2015 May 1;50(5):679–93.

20. Poulton R, Caspi A, Milne BJ, Thomson WM, Taylor A, Sears MR, et al. Association between children's experience of socioeconomic disadvantage and adult health: a life-course study. The Lancet. 2002 Nov 23;360(9346):1640–5.

21. Robins L, Cottler L, Bucholz K, Compton W. Diagnostic interview schedule for dsm-iv (dis-iv). 1995.

22. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom K-O. The Fagerström Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. Br J Addict. 1991 Sep 1;86(9):1119–27.

23. Schaefer JD, Caspi A, Belsky DW, Harrington H, Houts R, Horwood LJ, et al. Enduring Mental Health: Prevalence and Prediction. J Abnorm Psychol. 2017 Feb;126(2):212–24.

24. Moffitt TE, Caspi A, Taylor A, Kokaua J, Milne BJ, Polanczyk G, et al. How common are common mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. Psychol Med. 2010 Jun;40(6):899–909.

25. Steven AJ, Zhuo J, Melhem ER. Diffusion Kurtosis Imaging: An Emerging Technique for Evaluating the Microstructural Environment of the Brain. Am J Roentgenol. 2013 Dec 26;202(1):W26–33.

26. Ashburner J, Friston KJ. Unified segmentation. NeuroImage. 2005 Jul 1;26(3):839–51.

27. Ashburner J. A fast diffeomorphic image registration algorithm. NeuroImage. 2007 Oct 15;38(1):95–113.

28. D'Agata F, Caroppo P, Boghi A, Coriasco M, Caglio M, Baudino B, et al. Linking coordinative and executive dysfunctions to atrophy in spinocerebellar ataxia 2 patients. Brain Struct Funct. 2011 Sep 1;216(3):275–88.

29. Beckmann CF, Smith SM. Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging. IEEE Trans Med Imaging. 2004 Feb;23(2):137–52.

30. Mišić B, Sporns O. From regions to connections and networks: new bridges between brain and behavior. Curr Opin Neurobiol. 2016 Oct 1;40:1–7.

31. Bassett DS, Sporns O. Network neuroscience. Nat Neurosci. 2017 Feb 23;20(3):353–64.

32. Diamond A. Close Interrelation of Motor Development and Cognitive Development and of the Cerebellum and Prefrontal Cortex. Child Dev. 2000 Jan;71(1):44–56.

33. Schaie KW. Age Changes and Age Differences. The Gerontologist. 1967;7:128–32.

34. Anderson SF, Maxwell SE. Addressing the "Replication Crisis": Using Original Studies to Design Replication Studies with Appropriate Statistical Power. Multivar Behav Res. 2017 May 4;52(3):305–24.

35. Turner BO, Paul EJ, Miller MB, Barbey AK. Small sample sizes reduce the replicability of task-based fMRI studies. Commun Biol. 2018 Jun 7;1(1):62.

36. Diedrichsen J, Zotow E. Surface-Based Display of Volume-Averaged Cerebellar Imaging Data. PLOS ONE. 2015 Jul 31;10(7):e0133402.

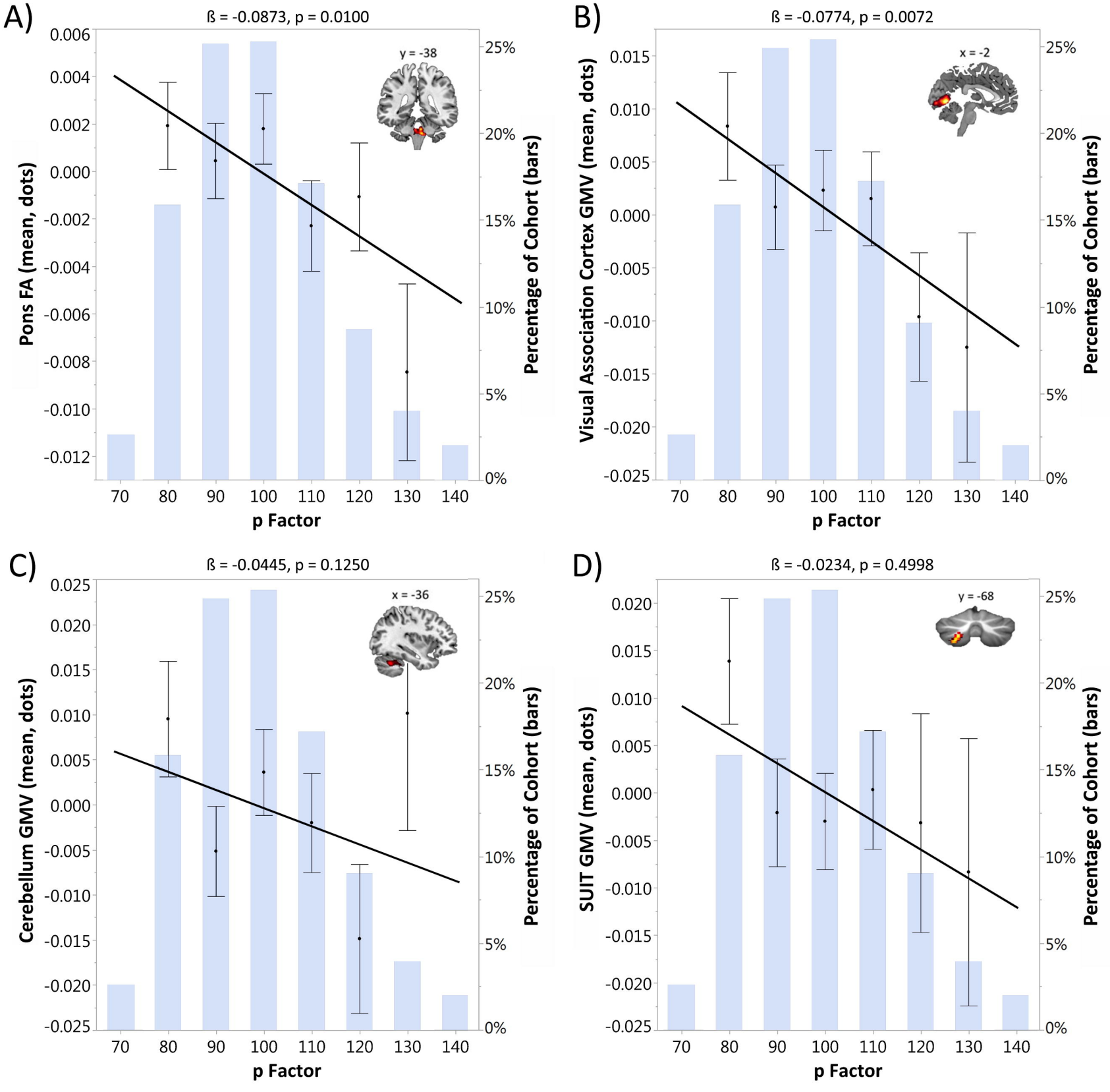
**Figure Captions**

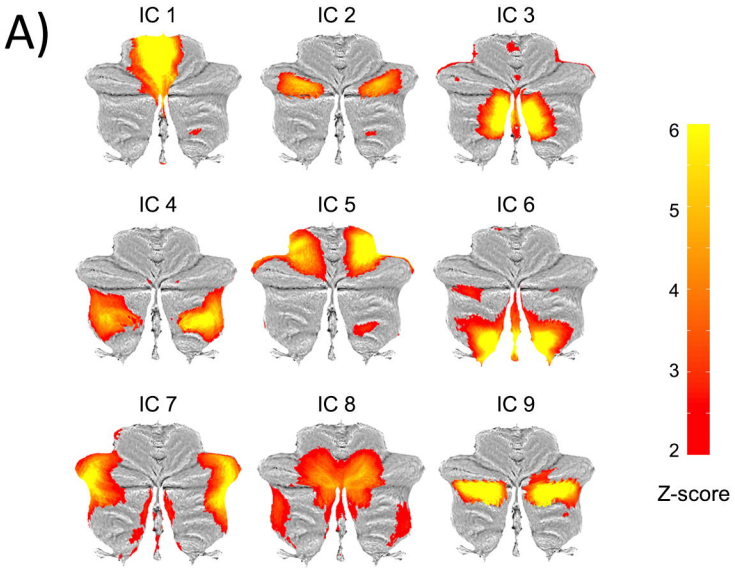
**Figure 1.** Replication analyses in the Dunedin Study of the original structural brain associations with *p* factor scores from Romer et al. (2018).

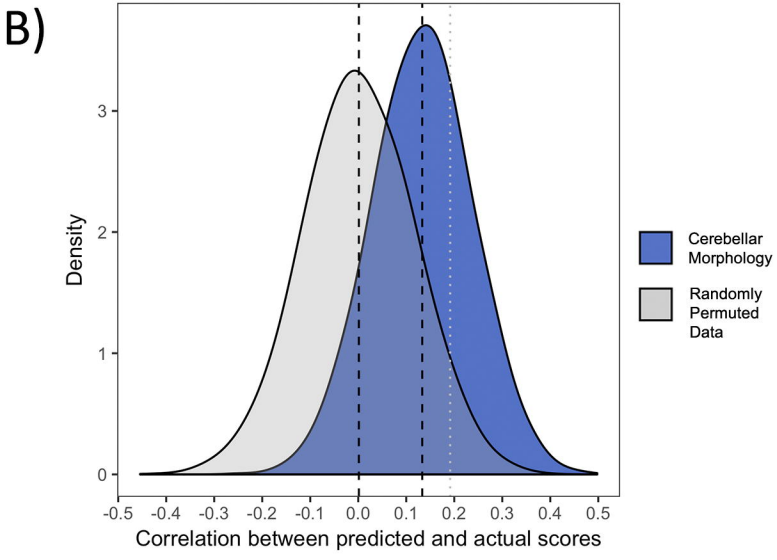
A) Replication of the negative association between pontine fractional anisotropy (FA) and *p* factor scores. B) Replication of the negative association between visual association cortex grey matter volume (GMV) and *p* factor scores. C) Non-significant replication of the negative association between cerebellar GMV and *p* factor scores. D) Non-significant replication of the negative association between SUIT-based neocerebellar lobule VIIb GMV and *p* factor scores. Per convention, *p* factor scores are normalized to a mean of 100 (SD = 15).

**Figure 2.** Replication analyses in the Dunedin Study of the original ICA-derived cerebellar morphology associations with *p* factor scores from Moberget et al. (2019).

A) The nine independent components resulting from data-driven decomposition of cerebellar grey matter maps projected onto flat-maps of the cerebellar cortex (36). B) Distributions of correlations between predicted and actual *p* factor scores across 10,000 iterations of the 10-fold cross-validated model using the average of the 9 independent components from A compared to the empirical null-distribution. The black dotted lines represent the mean for each distribution and the grey dotted line represents the one-tailed .05 threshold. The nine ICA-derived components predicted *p* factor scores beyond chance on average, but the difference from the empirical null distribution was p = 0.33, suggesting non-significant replication of Moberget et al. (2019).


A)
ß = -0.0873, p = 0.0100
y = -38
Pons FA (mean, dots)
Percentage of Cohort (bars)
p Factor
B)
ß = -0.0774, p = 0.0072
x = -2
Visual Association Cortex GMV (mean, dots)
Percentage of Cohort (bars)
p Factor
C)
ß = -0.0445, p = 0.1250
x = -36
Cerebellum GMV (mean, dots)
Percentage of Cohort (bars)
p Factor
D)
ß = -0.0234, p = 0.4998
y = -68
SUIT GMV (mean, dots)
Percentage of Cohort (bars)
p Factor


A)
IC 1
IC 2
IC 3
IC 4
IC 5
IC 6
IC 7
IC 8
IC 9
6
5
4
3
2
Z-score


B)
Density
0
1
2
3
-0.5
-0.4
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
0.5
Correlation between predicted and actual scores
Cerebellar Morphology
Randomly Permuted Data