

Identification of neurobehavioural symptom groups based on shared brain mechanisms

Alex Ing¹, Philipp G. Sämann², Congying Chu¹, Nicole Tay¹, Francesca Biondo¹, Gabriel Robert^{1,3}, Tianye Jia¹, Thomas Wolfers⁴, Sylvane Desrivieres¹, Tobias Banaschewski⁵, Arun L. W. Bokde⁶, Uli Bromberg⁷, Christian Büchel^{7,8}, Patricia Conrod^{9,10}, Tahmine Fadaei⁷, Herta Flor^{11,12}, Vincent Frouin¹³, Hugh Garavan¹⁴, Philip A. Spechler¹⁴, Penny Gowland¹⁵, Yvonne Grimmer⁵, Andreas Heinz¹⁶, Bernd Ittermann¹⁷, Viola Kappel¹⁸, Jean-Luc Martinot¹⁹, Andreas Meyer-Lindenberg²⁰, Sabina Millenet⁵, Frauke Nees^{5,11}, Betteke van Noort¹⁸, Dimitri Papadopoulos Orfanos¹³, Marie-Laure Paillère Martinot^{21,22}, Jani Penttilä²³, Luise Poustka²⁴, Erin Burke Quinlan¹, Michael N. Smolka²⁵, Argyris Stringaris^{26,27}, Maren Struve²⁵, Ilya M. Veer¹⁶, Henrik Walter¹⁶, Robert Whelan²⁸, Ole A. Andreassen^{29,30}, Ingrid Agartz^{30,31,32}, Hervé Lemaitre³³, Edward D. Barker^{1,34}, John Ashburner³⁵, Elisabeth Binder², Jan Buitelaar⁴, Andre Marquand⁴, Trevor W. Robbins³⁶, Gunter Schumann^{1,37,38,39,40*} and IMAGEN Consortium⁴¹

Most psychopathological disorders develop in adolescence. The biological basis for this development is poorly understood. To enhance diagnostic characterization and develop improved targeted interventions, it is critical to identify behavioural symptom groups that share neural substrates. We ran analyses to find relationships between behavioural symptoms and neuroimaging measures of brain structure and function in adolescence. We found two symptom groups, consisting of anxiety/depression and executive dysfunction symptoms, respectively, that correlated with distinct sets of brain regions and inter-regional connections, measured by structural and functional neuroimaging modalities. We found that the neural correlates of these symptom groups were present before behavioural symptoms had developed. These neural correlates showed case-control differences in corresponding psychiatric disorders, depression and attention deficit hyperactivity disorder in independent clinical samples. By characterizing behavioural symptom groups based on shared neural mechanisms, our results provide a framework for developing a classification system for psychiatric illness that is based on quantitative neurobehavioural measures.

Adolescence and its transition towards young adulthood form a critical period for the development of psychiatric illness, with half of the lifetime psychopathological burden emerging by the mid-teens and three-quarters by the mid-twenties¹. It coincides with major structural changes in grey and white matter² that are particularly pronounced in the limbic system and the prefrontal cortex³. Cognitive and (other) behavioural maturation reflect this brain-wide developmental process⁴. Because psychopathological symptoms during adolescent brain reorganization are often unspecific, and in many cases reversible, it has been difficult to identify unambiguously the early markers of sustained mental illness. Therefore most patients present during adulthood, often at a point when severe psychopathology has developed that gravely impairs their daily functioning. Presentation at this advanced stage increases individual suffering and renders therapeutic interventions more difficult.

At present, both adolescent and adult psychiatric diagnoses are made on the basis of combinations of behavioural symptoms that, while reflecting the psychopathological experience of generations of clinicians and patients, are not necessarily related to homogeneous pathophysiological or etiological processes. This results in biological

heterogeneity within diagnostic entities⁵, high rates of comorbidity between diagnoses^{6,7} and ill-defined targets for drug development. This is particularly relevant in adolescence, where there is evidence to suggest that psychiatric illness is more dimensional and less categorical than in adult psychopathology. Neuroimaging methods offer the opportunity to identify the biological mechanisms underpinning mental illness, without recourse to these categorizations^{8,9}.

One of the challenges in breaking up diagnostic borders, in favour of more homogeneous clusters of symptoms sharing common neural mechanisms, is that biological and behavioural data need to be combined in a meaningful way. A suitable method for this is canonical correlation analysis (CCA), which is formulated to maximize the correlation between variables in two views of a dataset. This technique has previously been used to link complex behavioural datasets with functional brain networks¹⁰. However, CCA has a number of limitations: it cannot be applied to data with more features than samples, results are difficult to interpret owing to a lack of localizability and it is only possible to find relationships between two sets of variables. The first two of these issues can be addressed using sparse CCA (sCCA)^{11,12}, which has been used to find modes of shared variation between resting-state functional connectivity

A full list of affiliations appears at the end of the paper.

magnetic resonance imaging (rs-fcMRI) and behavioural measures in adolescents and young adults¹². However, this approach is still limited in that it is only possible to identify relationships between psychiatric symptoms and one kind of biological measure at a time. We further enhanced sCCA by formulating a constrained form of multiple CCA, which maximizes the correlation between psychiatric symptoms, and several different neuroimaging modalities simultaneously¹³, before combining them in a linear regression model; we term this approach sparse multiple CCA regression (msCCA regression).

We investigated whether symptoms contributing to Diagnostic and Statistical Manual of Mental Disorders (DSM-5)/International Classification of Diseases (ICD-10) diagnoses can be reconfigured to identify neurobehavioural symptom groups that best represent specific underlying dysfunctional brain networks in adolescence. Here, we used a data-driven approach applied to a large general population neuroimaging sample to investigate direct relationships between neuroimaging measures of brain structure and function, yet without immediate recourse to diagnostic psychiatric categories. Following this, we sought to determine whether the regions we found to be related to psychiatric symptoms in adolescence were associated with full-blown clinical psychopathology in several independent clinical samples. Overall, this multistep approach enabled us to identify brain correlates of psychopathology in adolescence, probe their predictive value in the critical period between 14 and 19 years of age and characterize these brain correlates against the development of full-blown psychopathology.

Results

We used msCCA regression (see the Methods section ‘msCCA regression’) to link participant responses to the Development and Well Being Assessment (DAWBA; a structured interview for psychiatric DSM-5/ICD-10 diagnoses¹⁴; Supplementary Table 1) with voxel-based morphometry (VBM)¹⁵ measures of grey matter volume, fractional anisotropy (FA) along major white matter tracts using tract-based spatial statistics (TBSS)¹⁶ and functional connectivity between different brain regions, mapped from rs-fcMRI scans¹⁷. T₁-weighted and diffusion tensor imaging (DTI) data were preprocessed using voxel-wise VBM¹⁸ and TBSS¹⁶ methods respectively, as these procedures have been extensively studied and applied to real data. We mapped inter-regional rs-fcMRI connections across the brain using nodal maps defined by Miller et al.¹⁷; reasons for our pre-processing and analysis choices are detailed in the Methods section ‘Different neuroimaging pre-processing strategies’. We investigated 90 DAWBA items (symptoms) related to a broad range of psychiatric disorders, including affective and anxiety symptoms, attention deficit/hyperactivity and conduct symptoms, as well as substance use, eating disorders and symptoms of psychosis (Supplementary Table 1)¹⁴. This analysis was carried out on the general population IMAGEN sample, on participants of age 19. Following an in-depth quality control procedure (see Methods section ‘IMAGEN analysis’), data for $n = 666$ participants of age 19 was available.

To avoid overestimating the variance shared between psychiatric symptoms, and the neuroimaging modalities analysed (overfitting), we used a train/test analysis design, which allows us to estimate effect sizes in an unbiased way. Using a test set also allowed us to carry out further characterization of the data, without running into circularity problems. We carried out model selection in a training dataset of 70% of the data ($n = 467$), and model validation in the testing dataset of the remaining 30% ($n = 199$). To enhance stability, we resampled the data and retained only variables that contributed to the model in 90% of resamples (see Methods section ‘Stability selection’ and Supplementary Fig. 1)¹⁹. Demographic information on the full sample, training and testing sets is given in Supplementary Fig. 2. The msCCA-regression procedure we used in this investigation is designed to maximize associations between variable sets. For this

reason, all msCCA-regression significance values reported in the text are one-sided.

Using msCCA regression, we found a significant relationship between a subset of six DAWBA symptoms (see Fig. 1), and VBM, TBSS and rs-fcMRI measures ($r = 0.59$, $n = 465$, $P < 0.001$). The behavioural correlates derived from DAWBA covered symptoms linked to feelings of depression, anxiety and somatic problems, as well as temper and attentional problems (Fig. 1). The model was also significant when applied to the test dataset ($r = 0.23$, $n = 197$, $P < 0.001$, 95% confidence intervals (CIs) = 0.13, ∞) (Fig. 1), explaining 5.30% of the variance between psychiatric symptoms and the brain. Brain correlates derived from VBM, TBSS and rs-fcMRI measures were associated with this anxiety/depression symptom group with correlation values of: $r = 0.16$, $n = 197$, $P = 0.017$, 95% CIs = 0.040, ∞ ; $r = 0.14$ (197), $P = 0.040$, 95% CIs = 0.037, ∞ and $r = 0.15$, $n = 197$, $P = 0.029$, 95% CIs = 0.041, ∞ , respectively; all P -values are family-wise error (FWE)-corrected for multiple comparisons; see Methods section ‘Analysis design’ and Supplementary Fig. 3).

VBM, TBSS and rs-fcMRI modalities all showed an individually significant relationship to psychopathology. We carried out further localization analyses in each modality to identify brain regions that showed an individually significant relationship to psychopathology (see Methods section ‘Additional analyses to localize effects’). In this localization analysis, we identified one grey matter cluster in the right inferior temporal gyrus ($r = 0.16$, $n = 197$, $P = 0.032$ FWE corrected, 95% CIs = 0.041, ∞), and a single cluster of decreased FA in the genu of the corpus callosum ($r = 0.16$, $n = 197$, $P = 0.031$ FWE corrected, 95% CIs = 0.041, ∞). Both of these brain regions have been among those exhibiting the largest differences between healthy controls and patients with depression, in recent large, well powered meta-analyses^{20,21}. Further, we found an increase in functional connectivity between the default mode network, and the cerebellum ($r = 0.15$, $n = 197$, $P = 0.041$ FWE corrected, 95% CIs = 0.037, ∞); the default mode network has been implicated in several different psychiatric disorders, but particularly depression, with recent research showing that connectivity between the cerebellum and the default mode network is altered in patients with depression²². Information on the full set of regions found to be associated with psychiatric symptoms can be found in Supplementary Tables 2 and 3 and Supplementary Figs. 4 and 5.

We then removed the effects of the first canonical relationship and investigated the presence of a second dimension of shared variance between symptoms and the brain (see Methods section ‘Finding multiple modes of variation’). Here, we identified another behavioural correlate consisting of five items from the DAWBA, including problems with attention, fidgeting, rapidly changing moods and (lack of) conscientiousness, which was significantly associated with the neuroimaging modalities ($r = 0.46$, $n = 465$, $P = 0.004$). The test sample correlation is significant ($r = 0.19$, $n = 197$, $P = 0.002$, 95% CIs = 0.087, ∞), explaining 3.61% of the variance between psychiatric symptoms and the brain. Brain correlates derived from VBM, TBSS and rs-fcMRI measures were associated with the executive dysfunction symptom group with correlation values of $r = 0.19$, $n = 197$, $P = 0.012$, 95% CIs = 0.079, ∞ ; $r = 0.070$, $n = 197$, $P = 0.21$, 95% CIs = -0.029, ∞ and $r = 0.020$, $n = 197$, $P = 0.58$, 95% CIs = -0.090, ∞ , respectively. These results are displayed in Fig. 2.

Because the VBM modality was the only modality in this second canonical relationship to show an individually significant relation to psychopathology, we carried out a localization analysis for VBM data only in this modality; we found that executive dysfunction symptoms correlated with a single grey matter cluster in the right middle temporal gyrus ($r = 0.16$, $n = 197$, $P = 0.024$ FWE corrected, 95% CIs = 0.049), an area that has previously been shown to be associated with attention deficit hyperactivity disorder (ADHD) symptomology²³. Information on the full set of regions

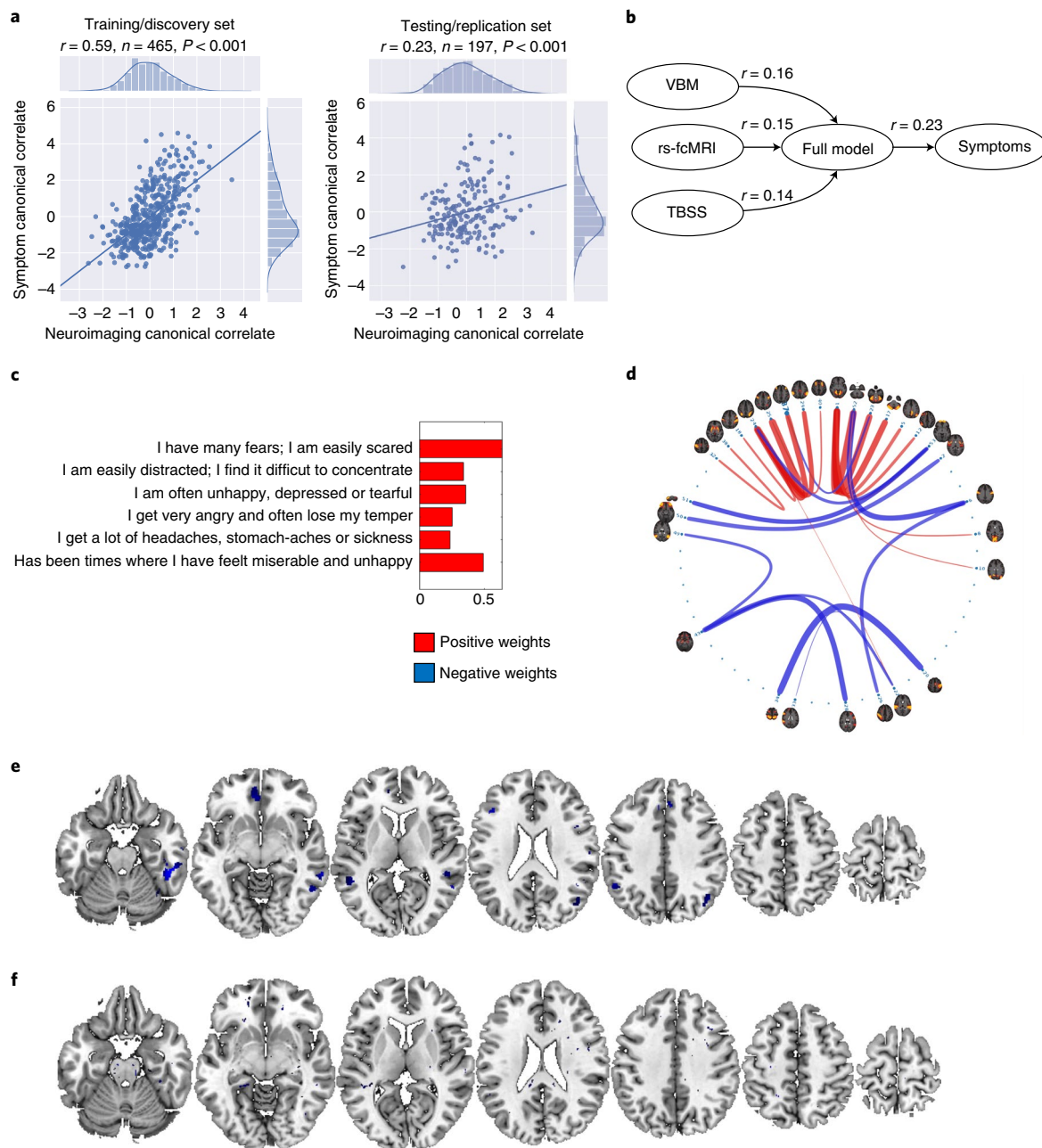


Fig. 1 | Results of the first msCCA-regression analysis showing relationships between anxiety/depression psychiatric symptoms and neuroimaging measures in the IMAGEN sample. a, The full msCCA-regression model linking psychiatric symptoms to VBM, TBSS and rs-fcMRI neuroimaging measures at age 19. We found associations between psychiatric symptoms and neuroimaging measures of $r = 0.59, n = 465, P < 0.001$ in the training set, and associations between symptoms and the brain of $r = 0.23, n = 197, P < 0.001, 95\% \text{ CIs} = 0.13, \infty$ in the test set. **b**, The msCCA-regression model linking psychiatric symptoms with the different neuroimaging measures. **c**, Psychiatric symptoms contributing to this relationship are shown on the left, with their canonical weights shown in red. **d**, rs-fcMRI measures of functional connectivity. Canonical weights are shown in red and blue. **e**, VBM measures of grey matter volume associated with symptoms. Canonical weights are shown in blue. **f**, TBSS measures of FA. Canonical weights are shown in blue.

found to be associated with psychiatric symptoms can be found in Supplementary Tables 4 and 5 and Supplementary Figs. 4 and 5. Associations between canonical anxiety/depression and executive dysfunction canonical correlates are given in Supplementary Table 6. Our results were robust to different rs-fcMRI atlas choices, as shown by repeated analyses using a different nodal definition²⁴, which generated similar results (Supplementary Fig. 6).

Hypothesis-driven analysis. To determine whether the canonical symptom groups identified in our data-driven analysis show a stronger

relationship to neuroimaging measures than existing means of organizing psychiatric symptoms, we carried out a hypothesis-driven analysis using internalizing and externalizing symptoms, which are often used in adolescent psychiatric diagnostics. We tested whether the canonical symptom groups identified with msCCA-regression were able to explain more variance than this widely used model of illness (see Methods section ‘Hypothesis-driven analysis’)²⁵. We term these pre-defined symptom groups as DAWBA-internalizing and DAWBA-externalizing. We found that the correlation of the DAWBA-internalizing dimension of psychopathology with

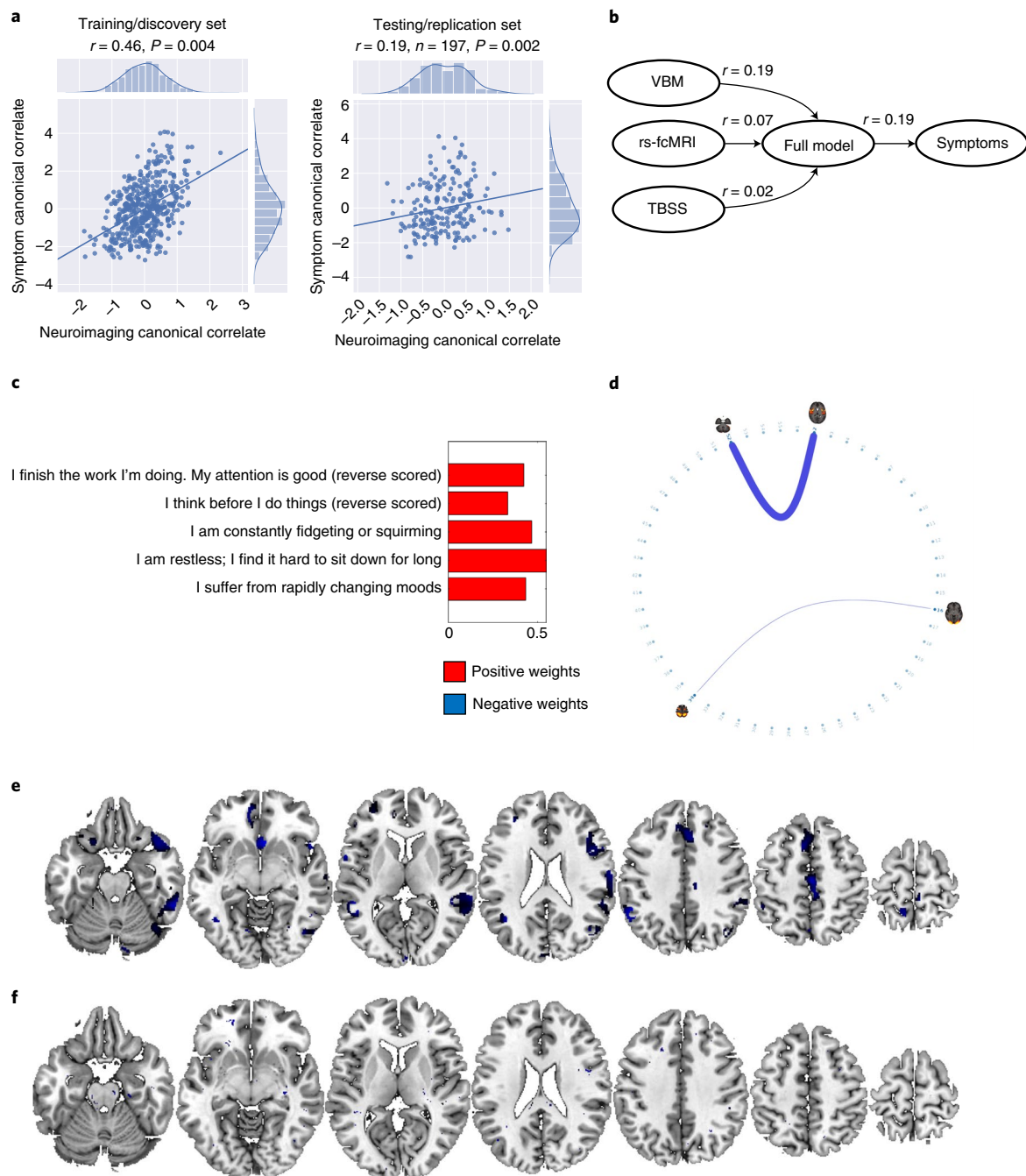


Fig. 2 | Results of the second msCCA-regression analyses showing relationships between executive dysfunction symptoms and neuroimaging measures in the IMAGEN sample, following the removal of the first canonical relation. a, The full msCCA-regression model linking psychiatric symptoms to VBM, TBSS and rs-fcMRI neuroimaging measures at age 19. We found associations between executive dysfunction symptoms and neuroimaging measures of $r = 0.46$ ($P = 0.004$) in the training set, and associations between symptoms and the brain of $r = 0.19, n = 197, P = 0.002, 95\% \text{ CIs} = 0.087, \infty$ in the test set. **b**, The msCCA-regression model linking psychiatric symptoms with the different neuroimaging measures. **c**, Symptoms contributing to this relationship are shown on the left with their canonical weights shown in red. **d**, rs-fcMRI measures of functional connectivity. Canonical weights are shown in red and blue. **e**, VBM measures of grey matter volume associated with symptoms. Canonical weights are shown in blue. **f**, TBSS measures of FA. Canonical weights are shown in blue.

neuroimaging measures shows only trend-level significance in the test set ($r = 0.12, n = 197, P = 0.060, 95\% \text{ CIs} = -0.02, \infty$) and explains 1.9% of variance. Similarly, DAWBA-externalizing dimensions of psychiatric illness correlated with neuroimaging measures at ($r = 0.040, n = 197, P = 0.28, 95\% \text{ CIs} = -0.095, \infty$) in the test set, explaining 0.16% of the variance (Supplementary Fig. 7). We then used a modified version of Dunn and Clarke's Z test^{26,27} to test

directly whether the association of the canonical symptom groups with the brain was significantly stronger than their pre-defined analogues. While the symptom-brain correlation of the executive-dysfunction symptom group was indeed significantly stronger than that of the DAWBA-externalizing symptom group ($Z = 1.95, n = 196, P = 0.029$), we did not find evidence that the strength of the association between the anxiety/depression symptom group and the

brain was significantly larger than that of the DAWBA-internalizing group ($Z=0.92$, $n=196$, $P=0.18$).

Longitudinal analysis. We carried out the initial cross-sectional analysis relating psychiatric symptoms to the brain at age 19, because most psychopathological symptoms will have manifested by this age. To investigate how adolescent brain development relates to the development of psychopathological symptoms, we analysed data from the same participants at age 14 years. After a conservative quality control procedure (see Methods section “Longitudinal analysis”), $n=412$ participants were available at this age in the training set and $n=182$ in the test set. First, we repeated the cross-sectional msCCA-regression analysis using VBM and TBSS (rs-fcMRI data was not available at age 14). We found a non-significant, trend-level association between symptoms and neuroimaging measures of $r=0.42$, $n=410$, $P=0.11$ in the training set. We found similarly non-significant results in the testing set ($r=0.10$, $n=180$, $P=0.090$, 95% CIs = $-0.017, \infty$). The results of these analyses are displayed in Supplementary Fig. 8.

There is previous evidence to suggest that neuroimaging measures precede the development of psychiatric symptoms in adolescence²⁸. We tested whether that was the case with the canonical symptom groups established in the present study by extracting the TBSS and VBM regions discovered at age 19 and using them as regions of interest at age 14. To obtain unbiased estimates of effect, we looked for associations in the test sample. Our data did not show any evidence of an association between anxiety/depression brain correlates and anxiety/depression symptoms at 14 years: $r=0.020$, $n=180$, $P=0.40$, 95% CIs = $-0.10, \infty$. However, the brain correlates taken from data at age 14 were predictive of symptoms at age 19: $r=0.14$, $n=180$, $P=0.023$, 95% CIs = $0.022, \infty$. These results are shown in Fig. 3. The difference in correlation between brain correlates at age 14 years with anxiety/depression symptoms at 14 years and 19 years was also significant, testing for a difference in association using a modified version of Dunn and Clarke’s Z test ($Z=1.74$, $n=179$, $P=0.041$)²⁷. We did not find evidence of an association between brain correlates and symptoms of executive dysfunction at age 14 years ($r=0.030$, $n=180$, $P=0.41$, 95% CIs = $-0.093, \infty$). Prediction of symptoms at 19 years showed a trend towards significance ($r=0.11$, $n=180$, $P=0.065$, 95% CIs = $-0.010, \infty$).

Clinical characterization. We investigated whether the canonical correlates of psychopathology we identified in a general population adolescent sample are correlated with fully developed psychiatric illnesses. In these analyses, we looked for case-control differences in the anxiety/depression and executive dysfunction canonical correlates, across four common psychiatric illnesses in several independent clinical samples. We carried out these analyses using VBM data alone, as this was the only data modality that showed an individually significant association with both symptom groups. Clinical and demographic information associated with the different clinical samples is displayed in Supplementary Fig. 9 and Supplementary Tables 7–9. Extensive information on quality control and data exclusion criteria for these clinical samples is given in the Methods section “Clinical analyses”. In assessing this data, we were looking for a directional effect, so we therefore report significance levels resulting from one-tailed tests in this section.

When analysing the data for case-control differences in grey matter correlates of anxiety/depression symptoms, we found significant reductions in regional grey matter volume in independent samples of patients with depression (t -statistic = 4.61, $n=612$, $P<0.001$, Cohen’s $D=0.39$, 95% CIs = $0.25, \infty$), schizophrenia (t -statistic = 2.54, $n=445$, $P=0.002$, Cohen’s $D=0.25$, 95% CIs = $0.087, \infty$) and in ADHD (t -statistic = 1.84, $n=203$, $P=0.034$, Cohen’s $D=0.26$, 95% CIs = $0.030, \infty$). In the executive dysfunction grey matter correlates, we found significant differences between patients

and healthy controls in ADHD (t -statistic = 2.19, $n=203$, $P=0.014$, Cohen’s $D=0.32$, 95% CIs = $0.070, \infty$), schizophrenia (t -statistic = 2.84, $n=445$, $P=0.0026$, Cohen’s $D=0.28$, 95% CIs = $0.11, \infty$) and depression (t -statistic = 1.65, $n=612$, $P=0.050$, Cohen’s $D=0.14$, 95% CIs = $0.001, \infty$). We did not find significant effects of bipolar disorder along either of these dimensions (t -statistic = -0.23 , $n=473$, $P=0.59$, Cohen’s $D=-0.02$, 95% CIs = $-0.17, \infty$) and (t -statistic = -1.33 , $n=473$, $P=0.90$, Cohen’s $D=-0.12$, 95% CIs = $-0.27, \infty$) respectively (Fig. 4). In these case-control analyses, the data distribution was assumed to be normal but this was not formally tested. To test whether the observed reduction in grey matter was specific to the brain correlates identified, as opposed to being a proxy for a generalized, brain-wide reduction in grey matter, we repeated the clinical comparisons using total grey matter as a covariate of no interest in addition to total intracranial volume (Supplementary Fig. 10). Results for ADHD and depression were unaffected by this change in pre-processing. In contrast, the results for schizophrenia were no longer significant.

Discussion

We ran analyses to establish direct relationships between psychiatric symptoms and neuroimaging measures of brain structure and function, without immediate reference to pre-defined psychiatric categories. This kind of dimensional, data-driven approach is particularly relevant in adolescence where there is a good deal of evidence suggesting that psychopathology is less differentiated than in adulthood and therefore does not fit into the traditional categorical conception of psychiatric disorder^{29,30}. We find two largely non-overlapping sets of brain regions that correlate with different sets of psychiatric symptoms. The first symptom dimension predominantly encompassed anxiety/depression symptoms, whereas the second dimension consisted mainly of executive dysfunction symptoms.

The anxiety/depression canonical symptom correlate was significantly associated with T_1 , rs-fcMRI and DTI data modalities. Participants scoring highly on this psychiatric scale showed decreased grey matter volume in the middle temporal gyrus, reduced FA in the genu of the corpus callosum, and increased functional connectivity between the default mode network and the cerebellum. A recent meta-analysis has demonstrated an association of depression with the right inferior temporal gyrus²⁰, a region exhibiting close connections with the limbic system, consistent with the theory that depression results from dysfunctional cortico-limbic circuits³¹. The genu of the corpus callosum is a commissural white matter pathway that links left and right prefrontal brain regions³². Changes in the structure of the corpus callosum are known to result in altered inter-hemispheric connectivity and impaired emotional control³³. The genu of the corpus callosum has been shown to be the white matter region with the largest difference in FA between controls and patients with major depression²¹. The default-mode network is a set of brain regions that reliably exhibit a decrease in activity when the brain is engaged in non-self-directed tasks; this network is thought to be primarily responsible for self-inspection and internal monitoring^{34,35}, which are processes overactive in depression³⁶. Increased connectivity between the default-mode network and the cerebellum has been previously reported in drug-naïve depressive patients²², consistent with its recently discovered involvement in complex cognitive and emotional processes³⁷.

We found that the executive dysfunction psychiatric symptom group was significantly correlated with neuroimaging measures derived from T_1 data. Here, decreased grey matter was localized to the right middle temporal gyrus, previously linked to ADHD²³. These results are more difficult to interpret as the function of this brain area is not well studied. As with the rest of the temporal lobe, this brain area is thought to be responsible for generating meaning from sensory inputs¹⁶. Furthermore, the temporal lobe

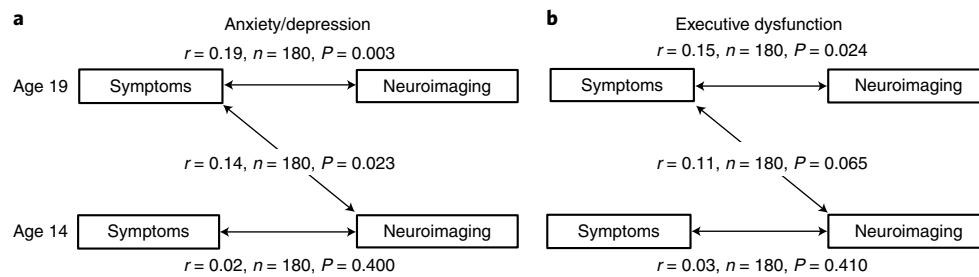


Fig. 3 | Longitudinal analysis of canonical correlates. a, Anxiety/depression symptom correlates. VBM and TBSS brain correlates established at age 19 are associated with anxiety/depression behavioural symptoms at age 19 ($r = 0.19, n = 180, P = 0.003$, 95% CIs = $0.069, \infty$), but not at age 14 ($r = 0.020, n = 180, P = 0.40$, 95% CIs = $-0.10, \infty$). Brain correlates at 14 years predict the manifestation of behavioural symptoms at 19 years ($r = 0.14, n = 180, P = 0.023$, 95% CIs = $0.022, \infty$). **b**, Executive dysfunction symptom correlates. VBM and TBSS correlates established at age 19 are associated with behavioural symptoms at age 19 ($r = 0.15, n = 180, P = 0.024$, 95% CIs = $0.028, \infty$), but not at age 14 ($r = 0.030, n = 180, P = 0.41$, 95% CIs = $-0.093, \infty$). Brain correlates at 14 years do not predict the manifestation of behavioural symptoms at 19 years ($r = 0.11, n = 180, P = 0.065$, 95% CIs = $-0.010, \infty$).

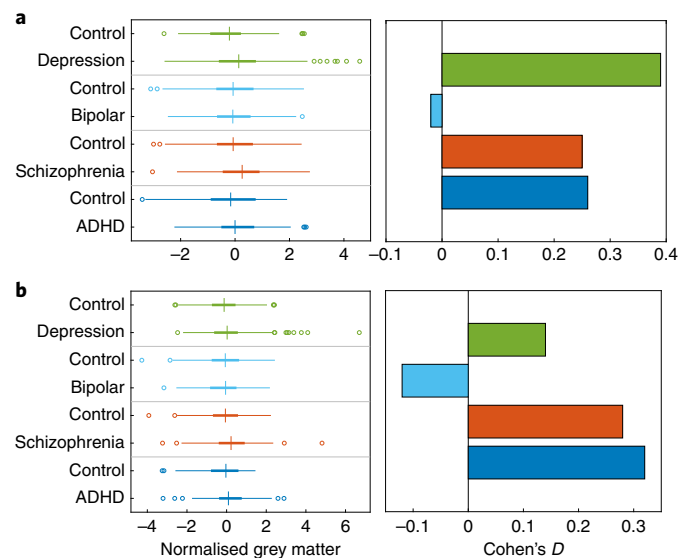


Fig. 4 | Differences in the grey matter correlates of anxiety/depression and executive dysfunction psychiatric symptoms between cases and controls for a range of psychiatric illnesses. For the box-and-whisker plots, the central mark in each box represents the median, with the top and bottom edges of the box indicating the 25th and 75th percentiles of the sample, respectively; whiskers represent 1.5x the interquartile range, and the hollow circles represent sample outliers. For display purposes, the total grey matter in each case-control comparison is shown in the right panel. The effect sizes (calculated using Cohen's *D*) relating to these differences are shown in the right panel. **a**, Differences in grey matter volume between patients and controls in the anxiety/depression set of grey matter correlates are shown in the left panel. Clinical psychiatric disorders exhibited the following case-control differences. Depression: t -statistic = 4.61, $n = 612, P < 0.001$, Cohen's $D = 0.39$, 95% CIs = $0.25, \infty$; schizophrenia: t -statistic = 2.54, $n = 445, P = 0.002$, Cohen's $D = 0.25$, 95% CIs = $0.087, \infty$; ADHD: t -statistic = 1.84, $n = 203, P = 0.034$, Cohen's $D = 0.26$, 95% CIs = $0.030, \infty$; bipolar: t -statistic = -0.23 , $n = 473, P = 0.59$, Cohen's $D = -0.02$, 95% CIs = $-0.17, \infty$. **b**, Differences in grey matter volume between patients and controls in the executive dysfunction set of grey matter correlates. Clinical psychiatric disorders exhibited the following case-control differences. Depression: t -statistic = 1.65, $n = 612, P = 0.050$, Cohen's $D = 0.14$, 95% CIs = $0.001, \infty$; schizophrenia: t -statistic = 2.84, $n = 445, P = 0.0026$, Cohen's $D = 0.28$, 95% CIs = $0.11, \infty$; ADHD: t -statistic = 2.19, $n = 203, P = 0.014$, Cohen's $D = 0.32$, 95% CIs = $0.070, \infty$; bipolar: t -statistic = $-1.33, n = 473, P = 0.90$, Cohen's $D = -0.12$, 95% CIs = $-0.27, \infty$.

functions closely with the hippocampus in the formation of memories¹⁶. Therefore, atrophied grey matter in this brain area may help to explain the learning deficits often observed with ADHD-like symptoms.

The identification of brain systems from a population-based cohort that is not suffering from any other psychiatric illness has major advantages. By identifying sub-clinical correlates of psychiatric illness, prior to the full manifestation of disorder, it is possible to avoid the potential impact of effects indirectly related to illness, such as substance use and medication effects. For example, 17% of the schizophrenic patients, and 21% of the bipolar patients, but none of the healthy controls studied here, have a history of alcohol

abuse, which has been linked to widespread decreases in grey matter³⁸. In addition, various psychiatric medicines, including lithium, which is often prescribed to bipolar patients, have also been linked to alterations in grey matter volume³⁹; it is possible that lithium-induced increases in grey matter volume may have contributed to the observed absence of significant findings in bipolar patients in this study.

We compared the efficacy of the data-driven msCCA-regression method with pre-defined psychiatric scales of internalizing and externalizing symptoms. We found that the data-driven approach identified relationships between symptoms and the brain that were significantly stronger than a similar approach using standard

internalizing and externalizing psychiatric symptom scales, defined without reference to any underlying biology. The fact that the canonical symptom groups show a stronger correlation with neuroimaging measures than pre-defined scales is important as it shows that data-driven methods may offer the potential to refine existing psychiatric categorizations⁶.

It is notable that grey matter correlates of psychopathology are already present at age 14 years, preceding the development of symptoms that only become manifest 5 years later, at 19 years. We also found that the brain correlates identified in the adolescent general population replicate in independent clinical samples of patients with corresponding psychiatric disorders, namely depression and ADHD. In addition to validating our primary results gained from population cohorts, these results raise the prospect of using neuroimaging measures, discovered in preclinical samples, as predictors of future psychopathology, thus enabling the development of targeted interventions in a young age group, where such measures are most effective in reducing the burden of mental illness⁴⁰.

It is important to note that the results of the msCCA-regression analysis applied here depend on the distribution of prevalence of psychopathological symptoms in each sample investigated. Thus, whereas a general population sample may yield an index of the normative variance in psychiatric symptoms from a broader range of different psychiatric disorders and their neural correlates, a patient sample might yield a narrower biological stratification within distinct clinical psychiatric categories, such as different biotypes of depression⁵, or symptoms of psychosis.

By basing symptom groups upon brain correlates, and by demonstrating specific associations of these correlates with clinical psychopathology, we have characterized stratification markers based on shared neural substrates. By discovering that these brain correlates identified in young adults are already established during adolescence, we have characterized biological risk markers prior to the manifestation of symptoms. Our work thus shows how major obstacles can be overcome in developing a taxonomy for psychiatric illness based on quantifiable neurobehavioural phenotypes.

Methods

Ethics. For IMAGEN, each site sought and received approval from the relevant local research ethics committee. Written consent was obtained from each participant and a parent or guardian.

For the depression studies in Munich, the studies were approved by the respective local ethics committees: the ethical committee of the Ludwig-Maximilians-Universität, Munich, Germany, and the ethical committee of the Bayerische Landesärztekammer, Munich, Germany. All participants provided written informed consent.

For the Thematically Organized Psychosis (TOP) study, all participants were recruited between 2003 and 2009 as part of an ongoing study of psychotic disorders (TOP). After receiving a complete description of the study, all participants gave informed consent to participate. The study was approved by the Regional Committee for Medical Research Ethics and the Norwegian Data Inspectorate.

The ADHD study was approved by the regional ethics committee (Centrale Commissie Mensgebonden Onderzoek, CMO Regio Arnhem, Nijmegen; 2008/163; ABR: NL23894.091.08) and the medical ethical committee of the VU University Medical Center. Informed written consent was obtained from each participant. For children under 18, both parents and children gave consent.

Study protocol. We developed a method, termed msCCA regression, of finding multivariate relationships between psychiatric symptoms, and multiple neuroimaging modalities simultaneously (VBM¹⁸ measures of grey matter volume, FA derived from DTI data and normalized using tract-based spatial statistics (TBSS)¹⁶, and resting-state functional connectivity neuroimaging measures⁴¹). msCCA-regression analysis was carried out in the general population IMAGEN sample, when participants were aged 19. Additional analyses were then applied in order to localize associations between psychiatric symptoms, and neuroimaging measures of brain structure and function. We then analysed neuroimaging and symptom data at age 14 in order to determine whether this multivariate relationship already existed at this earlier time point. We then assessed the clinical significance of our findings by conducting case–control comparisons of the structural markers found in the IMAGEN analysis, in several clinical samples. The following text gives a more detailed description of the methods described here.

IMAGEN analysis. IMAGEN is a large-scale neuroimaging–genetics cohort study conducted with the aim of understanding the biological basis of individual variability in psychological and behavioural traits, and their relationship to common psychiatric disorders⁴². The study involves a thorough neuropsychological, behavioural, clinical and environmental assessment of each participant. Participants also undergo biological characterization, with the collection of T₁-weighted structural magnetic resonance imaging (sMRI), DTI, task-based functional magnetic resonance imaging (fMRI), rs-fcMRI and genetic data. We used T₁-weighted MRI, DTI and rs-fcMRI data in the current investigation.

Participants. The analysis was carried out on participants drawn from the IMAGEN sample (see Schumann et al.⁴²). For IMAGEN, a general population sample of Caucasian adolescents were recruited from eight sites across France, Ireland, England and Germany. Data was collected at 14, 16 and 19 years of age. After a conservative quality control of MRI acquisitions and DAWBA questionnaires, participants with complete datasets were used in the subsequent data analysis. No statistical analyses were used to pre-define sample size. However, the sample size used was similar to that reported in previous studies^{10,12}.

DAWBA. Psychiatric symptoms of the IMAGEN participants were assessed using screening questions from the DAWBA, a wide-ranging psychiatric screening questionnaire⁴³. Participants were asked screening questions, assessing symptoms of specific fears, social fears, stress after a very frightening event, obsessions and compulsions, worrying, depression, rapidly changing mood, attention and activity, troublesome behaviour, drug and alcohol use, concern about appearance and strange/frightening experiences; if enough of these questions were answered in the affirmative, a more in-depth assessment of symptoms associated with that disorder was made. DAWBA screening questions have previously been used to define subthreshold clinical symptoms in neuroimaging studies of subclinical psychopathology⁴⁴. The SDQ was also used in the present investigation, as this questionnaire contributes to the assignment of diagnostic status in the DAWBA⁴⁵. Questions in the SDQ are categorized into broad internalizing and externalizing domains. The data from four of the questions asked had a standard deviation of zero among the participants asked, and were therefore not used in subsequent analyses. The full set of psychiatric questions asked in the present investigation can be found in Supplementary Table 1, where the questionnaire items that were omitted from the analysis are marked. At the time of the analysis conducted here, DAWBA/SDQ data had been collected for 1,510 participants. Of these, data were incomplete for 239 participants, and were not used.

T₁-weighted MRI acquisition. Scanning took place at eight different sites across Europe, using scanners built by four different manufacturers (Siemens, Philips, General Electric, Bruker). High-resolution, T₁-weighted images were obtained using a Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) sequence, based on the ADNI protocol (<http://adni.loni.usc.edu/methods/documents/mri-protocols/>). Scan parameters were standardized across sites to the highest degree possible (sagittal slice plane; repetition time 2.3 s; echo time 2.8 ms; flip angle 8°; 256 × 256 × 160 matrix; isotropic voxel size 1.1 mm).

VBM pre-processing. At the time this investigation was conducted, T₁ data had been acquired for 1,400 participants. All scans were visually inspected and manually reoriented. 285 scans were discarded from the analysis because they had movement artefacts, strong field inhomogeneities, abnormal field of view, abnormally reduced cerebellum or brace artefacts. The resulting 1,115 scans were used to build the study-specific template. Baseline and two follow-up scans were preprocessed using both the 2008 version of the Voxel Based Morphometry toolbox (VBM8)⁴⁵ running in SPM8 (version 5236)⁴⁶. Given that IMAGEN recruited young adults, we first used VBM8 in order to avoid using adult tissue probability maps to initiate the segmentation process. The VBM8 toolbox segmentation relies on an adaptive 'Maximum a Posterior' technique and tissue probability maps used in VBM8 are for registration purposes only. Diffeomorphic registration (Dartel, an option in SPM) was then used to register the 1,115 images, and to generate the study-specific population average template⁴⁷. We then re-sliced the data to 1.5 × 1.5 × 1.5 mm voxel size. Smoothing was carried out using an isotropic 8 mm full-width at half-maximum (FWHM) Gaussian smoothing kernel. We created a mask for the sample by taking the mean across all VBM maps included in the sample. We thresholded the mask at >0.4. We used a stringent mask to avoid overfitting the data⁴⁸. We then extracted all voxel values within this mask, resulting in 241,544 grey matter voxels.

DTI acquisition. The DTI acquisition sequence was based on the study by Jones et al.⁴⁹. Diffusion tensor images were acquired using an echo planar imaging sequence ($b = 0$ and 32 directions with b -value 1,300 s mm⁻²; axial slice plane; echo time = 104 ms; 128 × 128 × 60 matrix; voxel size 2.4 × 2.4 × 2.4 mm), adapted to tensor measurements (such as FA and mean diffusivity) and tractography analysis.

TBSS pre-processing. At the time this study was conducted, DTI data had been acquired for 1,412 participants. Of these, 71 were not usable owing to signal

dropouts or too much rotation. Diffusion imaging data was pre-processed using software from the FSL toolbox (www.fmrib.ox.ac.uk/fsl)⁵⁰. We preprocessed the remaining 1,341 scans using tract-based spatial statistics (TBSS)¹⁶. Pre-processing was carried out in the following manner. An affine registration was applied to the first B_0 image for head motion and eddy current correction. The B_0 image is an anatomical image measuring tissue signals and contrasts, but in the absence of a diffusion gradient. Brain extraction was carried out using the brain extraction tool (BET) in FSL. Diffusion tensor fitting was then used to obtain FA maps for each participant. All participants' FA data was aligned into a common space using the non-linear registration tool FNIRT (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FNIRT>), using a b -spline representation of the registration warp field. The mean was then taken across all FA maps to create an FA-averaged image. This map was then 'thinned' to create a mean FA skeleton, which was then thresholded at $FA > 0.2$, keeping only the major white matter tracts. Each participant's aligned FA data was then projected onto the mean skeleton. We then used these skeletonized maps in all subsequent analyses. The final mask used contained 106,812 voxels. A further ten scans were not used owing to masking or normalization issues in TBSS.

fMRI acquisition. fMRI scanning of the IMAGEN participants was carried out at multiple sites. The following parameters were standardized: number of volumes (164), repetition time (TR) = 2.2 s, echo time (TE) = 30 ms, flip angle = 75, number of slices/disabled discarded acquisitions (ddas) = 40/3, slice thickness = 2.4 mm, slice gap = 3.4 mm, voxel size = $3.4 \times 3.4 \times 2.4$ mm³, matrix size = 64^2 , field of view (FOV) = 218 mm.

fMRI preprocessing. At the time of this investigation, we had collected fMRI scans for 1,067 participants. Of these scans, 157 were not used, either because over 5% of scans from that participant exhibited artefacts of some kind, or because over 5% of the volume showed a fractional displacement of over 0.5 mm. Preprocessing of resting-state data was performed with routines from FMRIB's Software Library (FSL version 5.0.9)⁵⁰ and Advanced Normalization Tools (ANTs version 1.9.2)⁵¹.

1. Motion correction was carried out, applying a rigid body registration of each volume to the middle volume (using FSL routine MCFLIRT).
2. Non-brain tissue was removed (using FSL routine BET).
3. Spatial smoothing was applied using a 5 mm FWHM Gaussian kernel.
4. Independent component analysis (using FSL routine MELODIC) was run for each dataset. Artefact components were identified using an automatic classification algorithm, and subsequently regressed from the data (ICA-AROMA version 0.3)^{52,53}. ICA-AROMA⁵² has been shown to be as effective as motion parameter regression, with additional spike regression and 'scrubbing' in the removal of motion-related effects on functional connectivity measures derived from fMRI data. However, this procedure has the additional benefit that it preserves more signal of interest than these methods⁵³.
5. The resulting cleaned dataset was de-trended (up to a third-degree polynomial).
6. Co-registration to a high-resolution T_1 image (using FSL routine FLIRT and the BBR algorithm), and normalization to 2 mm isotropic Montreal Neurological Institute (MNI) standard space (ANTs) was carried out.
7. We used the CompCorr procedure to further clean the data of physiological noise⁵⁴. To do this we created white matter and cerebrospinal fluid masks by taking the mean of the white matter and cerebrospinal fluid segmentations from the VBM analysis, and thresholding them at 0.95; we then re-sliced these maps into the same space as the fMRI data. We then extracted timecourses from voxels within these regions, and took the first three principal components of this signal for both white matter and cerebrospinal fluid maps. These six principal component signals should represent the non-neuronal signal. We then regressed this non-neuronal signal from voxel timecourses across the rest of the brain.
8. Lastly, preprocessed and normalized resting-state datasets were re-sliced to 3 mm isotropic voxels.

Mapping rs-fMRI data.

1. We first generated 55 regional nodal timecourses using dual regression on nodal regions established in the UK Biobank sample¹⁷.
2. We mapped the correlation between nodal regions using Pearson's pairwise correlation coefficient, for each participant, thus producing a connectivity matrix for each participant. This connectivity matrix consists of 1,485 connections between nodes.
3. We then transformed these connectivity values using Fisher's Z -score transform.

Different neuroimaging processing strategies. A wide range of different preprocessing strategies can be applied in the analysis of neuroimaging data. Approaches to analysing DTI and T_1 can be categorized into two broad types: voxelwise and atlas-based approaches^{18,55}. We chose to analyse this data at the voxelwise level, as this allows for the highest level of spatial specificity. Although it is also technically possible to analyse rs-fMRI data across the whole brain at the voxelwise level, this approach results in an enormous number of features. When

mapping connectivity at the voxelwise level, in a dataset made up of N voxels, we are left with $(N \times (N-1))/2$ connections between those voxels. In the current investigation, $N = 57,053$, leading to $N \times (N-1)/2 = 1.63$ billion inter-regional connections. This would lead to a huge amount of redundancy in the data and computational, statistical and interpretational issues. For this reason, we mapped the connectivity between a pre-defined set of nodes. We used nodal definitions resulting from previous work applying independent component analysis to the UK Biobank sample¹⁷. We used this nodal definition because it derives from the largest extant sample of neuroimaging data. To test whether the results we obtained were robust to different nodal definitions, we also mapped inter-regional connectivity using the widely used Power atlas⁵⁶ and achieved similar results (Supplementary Fig. 6).

CCA and sCCA. CCA is a very general statistical method used to identify linear relationships between two or more sets of variables⁵⁷. It can be thought of as a generalization of multiple linear regression. The objective of CCA is to identify a relationship between two (or more) sets of variables, where there is no distinction between which variables are considered dependent, and which are considered independent. This method identifies weights for each variable, such that the weighted sum of variables in each set is maximally correlated with the weighted sum of variables from the opposite set, assuming a linear relationship.

Consider two matrices X_1 and X_2 , where each row denotes one of n observations, and each column denotes one of p_1 or p_2 features. CCA seeks to find the weight vectors w_1 and w_2 that maximize the correlation:

$$\rho = \text{corr}(X_1 w_1, X_2 w_2)$$

This optimization problem can be written as:

$$\rho = \max_{w_1, w_2} w_1^T X_1^T X_2 w_2$$

Subject to the constraints:

$$w_1^T X_1^T X_1 w_1 = 1 \text{ and } w_2^T X_2^T X_2 w_2 = 1$$

We assume that the columns of X_1 and X_2 have been standardized to have a mean of zero and a standard deviation of one. The vectors $X_1 w_1$ and $X_2 w_2$ are referred to as canonical variates.

Classical CCA is extremely powerful, but cannot be applied in situations where there are more features than samples (that is, $p_1 > n$ or $p_2 > n$, which is typically the case in neuroimaging studies). Interpreting and describing results from CCA can be difficult because the estimated weights are not sparse. This means that some variables may make negligible but non-zero contributions to the variance explained between sets. sCCA was developed to address these issues^{11,58,59}.

sCCA uses an L_1 penalty on canonical weights, which forces some of them to take a value of exactly zero. Furthermore, sCCA can also be applied in scenarios where there are more features than samples ($p > n$). The optimization criteria for sCCA can be written in the following manner:

$$\rho = \max_{w_1, w_2} w_1^T X_1^T X_2 w_2$$

Subject to the constraints:

$$\|w_1\|^2 = 1, \|w_2\|^2 = 1, \|w_1\|_1 \leq c_1 \text{ and } \|w_2\|_1 \leq c_2$$

Here, c_1 and c_2 are assumed to fall within the bounds $1 \leq c_1 \leq \sqrt{p_1}$ and $1 \leq c_2 \leq \sqrt{p_2}$, where p_1 and p_2 are the number of features in views X_1 and X_2 respectively.

msCCA regression. The formulation of sCCA described in the text above is designed to find relationships between two views of a dataset. However, we have collected data from several different neuroimaging modalities, and would like to use information from each of them. A naive approach to finding relationships between psychiatric symptoms and multiple neuroimaging measures would be to include all available neuroimaging modalities in one view of the canonical relation, with psychiatric symptoms in the other view. However, this approach is likely to be problematic because different modalities are associated with very different numbers of features. For example, the functional connectivity data used in the present investigation has only 0.6% of the number of features that the VBM data has. Therefore, if these modalities were entered into the same model, the VBM data would overwhelm the functional connectivity data.

We developed an approach designed to maximize the cross-correlation between psychiatric symptoms, and multiple neuroimaging modalities simultaneously, and then combined these modalities in a linear regression model. Formulations of CCA that are able to find relationships between more than two sets of data are termed multiple or generalized canonical correlation procedures. A widely used optimization criteria for multiple CCA is to maximize the sum of correlations between each of the different views of a dataset⁶⁰. Witten et al. have formulated a sparse version of multiple CCA³⁶; this formulation is designed to maximize the sum of correlations between all views of the data. However, in the present investigation, we are only interested in finding correlations between neuroimaging

measures, and psychiatric questionnaire responses; we do not wish to optimize the correlation between different neuroimaging measures.

As such, we seek to maximize the following relationship:

$$\max_{\mathbf{w}_1, \dots, \mathbf{w}_n} \mathbf{w}_1^T \mathbf{X}_1^T \sum_{i=2}^n \mathbf{X}_i \mathbf{w}_i$$

Subject to the constraints:

$$\|\mathbf{w}_1\|^2 = 1, \|\mathbf{w}_i\|^2 = 1, \|\mathbf{w}_1\|_1 \leq c_1 \text{ and } \|\mathbf{w}_i\|_1 \leq c_i$$

This method simultaneously optimizes the correlation between a weighted sum of variables in the target set, \mathbf{X}_1 , with a weighted sum of variables in the other sets. In the present investigation, \mathbf{X}_1 is a matrix of psychiatric symptoms and \mathbf{X}_2 to \mathbf{X}_n are neuroimaging measures of brain structure and function. Using this method, we are able to maximize the correlation between psychiatric symptoms, and between several different neuroimaging modalities within the same integrated model. A natural choice for the statistic of interest, in any inference carried out using this procedure, would be the sum of correlations between the symptom data, and the neuroimaging measures of brain structure and function. However, a sum of correlations is of less practical benefit than understanding how much total variance is shared between neuroimaging measures of brain structure and function, and psychiatric symptoms. Therefore, in the final step of this process, we combine canonical neuroimaging variables in an ordinary linear regression model. Canonical variables are defined as:

$$C_i = \mathbf{X}_i \mathbf{w}_i$$

The canonical variables are then combined in the prediction of psychiatric symptoms using ordinary linear regression:

$$C_1 = \beta_0 + C_2 \beta_2 \dots + C_n \beta_n + \epsilon$$

We used this approach to establish relationships between psychiatric symptoms (C_1), and TBSS (C_2), VBM (C_3), and connectivity measures (C_4) derived from rs-fcMRI data and β_n are the associated weights estimated using ordinary linear regression (β_0 is the constant estimated in regression).

msCCA regression was carried out using in-house codes written in MATLAB. This algorithm requires an initialization value. In the present study, initial weights were randomly generated. Weight values associated with psychiatric symptoms were always constrained to be positive to ensure interpretability.

This study is designed to be exploratory in nature. Nevertheless, given the very large quantity of data we sought to integrate, it is likely that some simple priors will help to improve the stability of our results, as long as those priors are well supported. There is a great deal of evidence suggesting that psychopathology is associated with decreases in both grey matter, and FA, across psychiatric disorders^{61,62}. For this reason, we constrained the canonical weights on VBM volume and FA to be negative. This will help to reduce variance in the model and will help increase interpretability of our results. In contrast, there is no clear evidence that psychiatric illness is associated with increases or decreases in connectivity measures derived from blood-oxygen-level-dependent (BOLD)-fMRI. Therefore, we did not add constraints to the functional connectivity data.

Stability selection. Although msCCA-regression (and sCCA) have advantages over classical CCA in terms of interpretability, they can suffer from instabilities due to their use of an L_1 penalty to introduce sparsity¹⁹. This is particularly true when $p \gg n$, and when there is a high degree of collinearity in the data. Stability selection is a widely applicable feature selection procedure that can address this problem¹⁹. This procedure has the added benefit that it makes the results less sensitive to the choice of L_1 penalty.

The conceptual underpinning of stability selection is very simple: if a model is repeatedly resampled, features exhibiting a 'real' effect will be selected more often than noise. Using stability selection, data is repeatedly split into random subsamples of size $n_i/2$ (where n_i is the total number of participants in the training dataset). In this work, resampling was carried out a hundred times. msCCA was applied to each resample, and those features that appear more often are deemed to be more stable. Deciding which variables are stable requires a threshold: π , is defined as the fraction of samples in which a particular variable must be observed to be considered stable. We set π to 0.9, which means that a particular variable must be present in 90% of resamples to be considered stable. The outcome of this stability selection procedure is a set of stable features. A benefit of stability selection is that it is insensitive to tuning parameters. Here, we simply set the L_1 penalty at $\sqrt{p}/2$, which is halfway along the regularization path running from 1 to \sqrt{p} . It is worth noting that the stability selection procedure is easily parallelizable here as it simply involves re-applying the msCCA-regression algorithm to multiple different resamples of the same data.

Analysis design. The L_1 penalty used in sCCA means that the parametric tests used for significance testing in classical CCA (for example Wilk's Lambda)⁶³ cannot be used here, necessitating the use of permutation testing to determine whether

results are significant. We assessed the in-sample significance of the results we obtained here, then replicated these findings using an out-of-sample, hold-out set design. This kind of experimental design has a number of advantages in the present context: using a training/testing design, it is possible to obtain an unbiased estimate of effect size. We used a hold-out set design in preference to a cross-validation procedure. This is because cross-validation involves the training and testing of multiple statistical models, one for each cross-validation fold, which precludes the use of a single model for further validation/characterization. A related advantage is that it is possible to carry out further characterization of the test set results, owing to the fact that we are able to estimate effect size in an unbiased way.

In detail, the analysis design was carried out as follows:

1. Psychiatric symptom data, and data from the VBM, TBSS and rs-fcMRI neuroimaging modalities were extracted and transformed into $n_i \times p_i$ matrices, where n_i is the number of participants included in the training dataset, and p_i is the number of features included in each of the views of the data.
2. The full dataset was randomly split into training and testing sets. The training set was made up of 70% of the data whilst the testing set was made up of the remaining 30%.
3. The training data was then randomly split into a hundred further resamples. Each resample was made up of $n_i/2$ participant scans, where n_i is the total number of participants in the training dataset.
4. The first stage of the msCCA-regression algorithm (see above) was then applied to each resample, with a sparsity constraint of $\sqrt{p_i}/2$ in each view of the data.
5. We then recorded which variables, in each view of the data, are present in over 90% of resamples. These variables are considered to be stable, and are retained.
6. We then re-applied the msCCA algorithm to the data, without sparsity constraints, on the variables that survived more than 90% of the resampling.
7. We then combined the neuroimaging canonical variates we found in the previous step in a prediction model on the symptom canonical variate, using ordinary least-squares regression. We then recorded the correlation between the neuroimaging prediction model, and the symptom canonical correlate.
8. We then permuted the training data, and repeated steps 3)–7). This was done for 10,000 different permutations of the training data labelling. In each case, we recorded the correlation between the neuroimaging model, and the canonical correlate of psychiatric symptoms. In this way, we built up a permutation distribution to assess the significance of the relationship between symptom and neuroimaging data in the experimental labelling, within the training dataset.
9. We then applied the trained model to the test set to produce canonical correlates of symptom and neuroimaging measures. We recorded associations for both the full model, and between the psychiatric symptom score and each of the individual neuroimaging canonical correlates.
10. We then randomly permuted the data rows in the test set and re-calculated correlation values between symptom and brain canonical correlates. We recorded associations between psychiatric symptoms and the full neuroimaging model, for each of 10,000 permutations of the experimental labelling.
11. It is also interesting to find the significance of the individual neuroimaging modalities. However, as we are testing the significance of multiple neuroimaging modalities, it is necessary to correct for multiple comparisons across these different modalities. This is easily done using the distribution of the maximal statistic: for each permutation of the experimental labelling, we calculate the association between the symptom score and each of the neuroimaging canonical correlates; the largest of these associations is then recorded. This is done for each of the 10,000 permutations of the test labelling, producing a distribution of the maximal statistic. Correlations between symptom and neuroimaging measures in the experimental labelling are then significant at the FWE-corrected level α if they are above the $100 \times (1 - \alpha)$ percentile of this distribution.

This process is illustrated in Supplementary Fig. 1.

Confounds. It is important to account for the effects of confounds, which might otherwise lead to spurious relationships between the different data views⁶⁴. Here, we regressed age, gender, site and intracranial volume from all data views prior to the sCCA analysis^{65–67}. For the connectivity measures derived from rs-fcMRI data, we also regressed the mean between-volume fractional displacement, and the percentage of slices corrupted by artefacts, from the scans.

Additional analyses to localize effects. We used msCCA regression to find multivariate relationships between psychiatric symptomatology and neuroimaging measures of brain structure and function. In using msCCA-regression, it is possible to make inferences about relationships between sets of psychiatric symptoms and neuroimaging measures across the brain, but it is not possible to make inferences on individual brain regions/connections or individual questionnaire items. For this reason, we conducted additional analyses to further deconstruct the relationship between psychiatric symptomatology and the brain. This procedure is similar to a redundancy analysis^{68,69}. In particular, we were interested in localizing which

brain regions exhibited an individually significant association with psychiatric symptomatology.

Conducting further tests on the whole dataset would introduce circularity into the analysis. Therefore, additional inferences must be carried out on the testing dataset alone. Nevertheless, the training dataset is still likely to contain useful information, which can be used to guide analyses carried out on the testing dataset, thus decreasing the multiple comparison problem, and increasing the likelihood of finding significant effects in the testing dataset. In the present investigation, we looked for significant localizable effects in the training dataset, and then used these results to inform analyses carried out on the testing dataset. In this sense, the training dataset was used as a 'discovery dataset'.

For the TBSS and VBM data, we sought to localize associations between symptoms and the brain to the cluster-wise level. In the case of the rs-fcMRI data, we sought to localize changes to individual inter-regional connections. VBM and TBSS clusters were defined using an 18-connectivity scheme. This means that voxels must be connected by a face or an edge to be considered a part of the same cluster.

This analysis was carried out in the manner described below:

1. We calculated the grey matter volume and FA in spatially distinct clusters identified in the sCCA analysis applied to VBM and TBSS, respectively. We extracted connectivity values with non-zero canonical weights. This was done in both the training and testing datasets.
2. We calculated Pearson's correlation coefficient between the mean of each spatially distinct cluster/connection, and the sum of symptom score values. This was done separately in the training and testing datasets.
3. Rows associated with neuroimaging data in the training set were permuted and correlations between clusters/connections, and symptom clusters were recalculated. The maximal value was recorded. Training data were permuted 10,000 times; the maximum correlation value across all clusters/connections was recorded for each permutation. Clusters/connections exhibiting a significant effect in the training dataset were then determined by comparing correlation values to the distribution of the maximal statistic^{50,51}. Because model selection was carried out in the training dataset, conducting inference on the training dataset would constitute 'double dipping'.
4. Clusters/connections exhibiting a significant effect in the training dataset were taken forward for an analysis carried out in the testing set.
5. We calculated correlation values between clusters/connections in the testing dataset, and the symptom score.
6. Testing data was permuted 10,000 times; the maximum correlation value across all clusters/connections was recorded for each permutation. Clusters/connections exhibiting a significant effect in the testing dataset were determined by comparing correlation values to the distribution of the maximal statistic. Cluster/connection correlations in the testing dataset were then compared to correlations in the distribution of the maximal statistic. Cluster/connection correlations in the experimental labelling that were in the top 5% of the distribution of the maximal statistic were considered significant at the FWE corrected level.

This process is illustrated in Supplementary Fig. 3.

Finding multiple modes of variation. Using CCA, it is possible to uncover multiple modes of variation between datasets. After determining the significance of the first canonical correlate, we remove the effect of the first set of canonical vectors, and repeat the analysis. Witten et al. used Hotelling's deflation to remove the effect of the first vector; this approach has been criticized by Monteiro et al., who propose the projection deflation procedure as an alternative^{1,70}; this is the procedure we use in the present investigation. Correlations between the different canonical relationships are given in Supplementary Table 6.

It is possible to ascertain the significance of all canonical relationships after the first by comparing the correlations of subsequent associations to the permutation distribution of the first relationship. The first canonical relationship between sets is by definition the strongest; any subsequent associations between sets will be weaker than the canonical relationship that preceded it. A common means of correcting for multiple comparisons is to compare test statistics in the experimental labelling to the maximal statistic across all tests in the permutation distribution; this distribution is usually termed the distribution of the maximal statistic^{71,72}. In the present investigation, we can find this distribution by recording the strength of the first canonical relationship, for each permutation. Significance values that are corrected for multiple comparisons can then be found by comparing associations of subsequent modes of variation, with this distribution⁷³.

Hypothesis-driven analysis. A major advantage of the approach described here is that it allows the grouping of psychiatric illnesses to be driven by their biological underpinnings. Nevertheless, it is an open question whether the symptom groups discovered in the data-driven analysis we ran here show a stronger relationship to neuroimaging measures of brain structure and function than pre-defined symptom groups. For this reason, we tested whether the widely used internalizing/externalizing organization of psychiatric illness is able to explain as much variance in psychiatric symptomatology as can this purely data-driven method.

To do this, we used an approach that is as similar as possible to the primary data analysis followed in the main part of the investigation, yet still makes use of the internalizing/externalizing illness structure: we replaced the symptom matrices used in the main part of the investigation with symptom vectors based on previously defined internalizing and externalizing symptom sub-scales from the DAWBA; no sparsity was applied to psychiatric symptom sub-scales. Used in this manner, the msCCA algorithm reduces to something like a sparse partial least-squares regression⁷⁴, where the neuroimaging features are predictors and the pre-defined internalizing/externalizing vectors are the targets. This method was applied twice, once to predict the internalizing symptom dimension, and once to predict the externalizing. We term the internalizing and externalizing symptom scales as DAWBA-internalizing and DAWBA-externalizing, respectively. We defined symptoms as belonging to broad internalizing or externalizing categories in the same way as Aebi et al.⁷⁵: the DAWBA-internalizing scale was created by summing 'specific fears', 'social fears', 'panic attacks', 'stress after a frightening event', 'worrying' and 'depression'. The DAWBA-externalizing scale was created by summing 'attention and activity', 'behaviours and attitudes that can get people into trouble', and the 'cigarettes, alcohol and drugs' sections of the DAWBA. The SDQ is already split into broad internalizing and externalizing domains⁴³. Therefore, internalizing and externalizing SDQ scores were simply added to these scores to create DAWBA-internalizing and DAWBA-externalizing scores respectively. The sections on 'rapidly changing mood', 'dieting and bingeing', and 'strange experiences that are surprisingly common' were not used to create scores because these symptoms do not fit neatly into an internalizing/externalizing dichotomy. All of these questions can be found in Supplementary Table 1.

Longitudinal analysis. The msCCA-regression analysis described above was used to find relationships between psychiatric symptoms and neuroimaging measures of brain structure at age 19, when participants were young adults. However, the developmental time period immediately preceding this time point is also of potential interest, with the brain going through important maturational processes and participants being at increased risk for the development of psychopathology⁷⁶. Thus, we applied the msCCA-regression algorithm between psychiatric symptoms and neuroimaging measures at age 14. The results of this analysis are shown in Supplementary Fig. 8. We did not find a significant relationship between psychiatric symptoms and the brain at this age. As fMRI data is only available for a small sub-sample of the full dataset at age 14, we used only VBM and TBSS data in this analysis.

It is possible that neuroimaging markers of psychiatric illness precede the development of full-blown psychiatric symptomatology. To determine whether this was the case in the present investigation, we took the TBSS and VBM regions identified as being associated with psychopathology at age 19, then extracted the appropriate neuroimaging data from these brain regions at age 14, and correlated the output with symptoms at age 19. In this way, we showed that neuroimaging measures at age 14 have predictive value for psychopathology at age 19.

For these analyses, we used the same subjects as were included in our analysis at age 19. We also used the same training-test split within this subject group. We subjected the age 14 data to the same quality control procedures as the data taken at age 19. Of the $n = 666$ subjects used in the msCCA-regression analysis carried out at age 19, 72 subjects had data that did not pass quality control at age 14. This left $n = 594$ subjects for age 14 analyses, with $n = 412$ subjects in the training group and $n = 182$ in the testing/replication group.

Clinical analyses. Using msCCA-regression, we found a set of neuroimaging features that correlate with a set of questions assessing psychiatric health. At the group level, participants who score more highly on the vector derived from neuroimaging data will suffer a larger number of psychiatric symptoms (as measured by the DAWBA). It might therefore be expected that participants with a clinical diagnosis of a psychiatric disorder would score more highly on this neuroimaging vector than would healthy control participants. To discover whether this was the case, we subjected clinical data to exactly the same pre-processing as the IMAGEN data; we then looked for changes in grey matter volume in the regions identified in the initial analysis. A (one-sided) two-sample t -test was used to determine whether patients and controls differed significantly on this one-dimensional measure. We only used grey matter data here as this data type showed the strongest relationship to psychopathology in the IMAGEN sample. Furthermore, this data type is widely available and the number of degrees of freedom in the MRI scan acquisition parameters is low. The case-control tests we used here make the assumption of data normality, although this was not formally tested here.

We used the same confounds in this analysis as we did on the IMAGEN data, including the use of total grey matter as a covariate of no interest. However, it could still be argued that regional changes are only acting as a proxy for total grey matter. To determine whether this is the case, we repeated all pertinent analyses, using total grey matter as a regressor in addition to total intracranial volume. The results of these analyses are shown in Supplementary Fig. 10.

Depression sample. The Munich sample consisted of patients with first episode and recurrent unipolar depression treated as in-patients at the Max Planck Institute

of Psychiatry in Munich, and healthy control participants. The data for 13 of the participants assessed was not used as it was deemed to be of insufficient quality, which left $N=614$; 400 patients, age 48 (standard deviation, s.d. 13.8) years, 53% women; 214 control participants age 49 (s.d. 13.3) years, 58% women, for the most part overlapping with imaging genetic and major depressive disorder (MDD) association studies reported in collaboration with the ENIGMA consortium^{30,77}. Other than in the two flagship studies, no bipolar patients were included for reasons of clinical homogeneity. MDD diagnoses were based on clinical consensus in addition to Munich-Composite International Diagnostic Interview (M-CIDI) or Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interviews, depending on the original study protocols. The Munich sample comprised images acquired in subsamples of the Munich Antidepressant Response Signature Study and the Recurrent Unipolar Depression Case-Control study, both performed at the Max Planck Institute of Psychiatry. We did not use any statistical analyses to decide on the sample size used here. However, the sample used was among the largest of any single study investigating alterations in brain structure in depressed participants⁷⁷.

Schizophrenia/bipolar sample. Participants with schizophrenia and bipolar disorder were recruited from the TOP study. This is a collaborative study based at the University of Oslo in Norway. The data for two participants were not used as they were considered to be of insufficient quality, which left 286 control participants (aged 34 (s.d. 9.5) years, 46% women), 161 schizophrenics (aged 32 (s.d. 8.8) years, 35% women) and 189 participants with bipolar disorder (aged 34 (s.d. 11.5) years, 58% women). Patients were recruited from the psychiatric unit of Oslo University Hospital and were assessed for psychiatric illness using the Structural Clinical Interview for DSM-IV Axis I disorders (SCID-I). This assessment was either administered by a doctor or a clinically trained psychologist, and was used to assess the presence of AXIS I disorders. Before participation, control participants were screened to exclude serious somatic and psychiatric illness, substance abuse, or MRI-incompatibility. All participants gave written informed consent before participation. Further information about this sample and the scan protocols used can be found in Rimol et al.⁷⁸. We did not use any statistical methods to pre-define the sample size used in this investigation. Nevertheless, the sample used is among the largest of any investigating structural brain alterations in schizophrenia⁷⁹ and bipolar disorder³⁰.

ADHD sample. Data for the ADHD sample was taken from the NeuroIMAGE project, a clinical cohort study. The study is made up of individuals tested at two different sites in the Netherlands, The Donders Centre for Cognitive Neuroimaging in Nijmegen and the Vrije Universiteit in Amsterdam. The total sample consisted of 184 participants suffering from ADHD, 103 unaffected siblings, and 128 healthy controls. Further information on the participants and the protocols used can be found in von Rhein et al.⁸⁰. This sample includes a number of very young participants, which is likely to introduce a large degree of heterogeneity into the analysis. For this reason, we did not analyse the data from participants under the age of fifteen. This age divide point was considered to offer a reasonable trade-off between sample homogeneity and size. The data for 12 of the participants was not used as they were deemed to be of insufficient quality. Case-control analyses were made between 74 healthy controls (aged 18 (s.d. 2.0) years, 50% women) and 131 ADHD participants (aged 18 (s.d. 2.3) years, 27% women). No formal statistical methods were used to determine the size of this sample. However, this sample is large compared to similar samples investigating case-control differences in brain structure in patients with ADHD⁸¹.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The IMAGEN data used in this investigation will be made available on reasonable request to the corresponding author. All other data are available on reasonable request to the appropriate study leader.

Code availability

The core code used to run the analyses reported in this study are available as Supplementary Software. The supporting code can be found at: <https://github.com/alexjamesing/mscca-regression-code>.

Received: 13 May 2019; Accepted: 22 August 2019;
Published online: 07 October 2019

References

- Kessler, R. C. et al. Age of onset of mental disorders: a review of recent literature. *Curr. Opin. Psychiat.* **20**, 359–364 (2007).
- Giedd, J. N. et al. Brain development during childhood and adolescence: a longitudinal MRI study. *Nat. Neurosci.* **2**, 861–863 (1999).
- Steinberg, L. Risk taking in adolescence: new perspectives from brain and behavioral science. *Curr. Dir. Psychol. Sci.* **16**, 55–59 (2007).
- Gogtay, N. et al. Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl Acad. Sci. USA* **101**, 8174–8179 (2004).
- Drysdale, A. T. et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* **23**, 28–38 (2016).
- Insel, T. et al. Research Domain Criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).
- Lahey, B. B. et al. Is there a general factor of prevalent psychopathology during adulthood? *J. Abnorm. Psychol.* **121**, 971–977 (2012).
- Zhang, X. et al. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **113**, E6544 (2016).
- Rosenberg, M. D. et al. A neuromarker of sustained attention from whole-brain functional connectivity. *Nat. Neurosci.* **19**, 165–171 (2016).
- Smith, S. M. et al. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.* **18**, 1565–1567 (2015).
- Witten, D. M., Tibshirani, R. & Hastie, T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10**, 515–534 (2009).
- Xia, C. H. et al. Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat. Commun.* **9**, 3003 (2018).
- Kettenring, J. R. Canonical analysis of several sets of variables. *Biometrika* **58**, 433–451 (1971).
- Goodman, R., Ford, T., Richards, H., Gatward, R. & Meltzer, H. The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J. Child Psychol. Psychiat.* **41**, 645–655 (2000).
- Ashburner, J. A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95–113 (2007).
- Smith, S. M. et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* **31**, 1487–1505 (2006).
- Miller, K. L. et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
- Ashburner, J. & Friston, K. J. Voxel-based morphometry—the methods. *Neuroimage* **11**, 805–821 (2000).
- Meinshausen, N. & Bühlmann, P. Stability selection. *J. R. Stat. Soc. Ser. B* **72**, 417–473 (2010).
- Schmaal, L. et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA major depressive disorder working group. *Mol. Psychiat.* **22**, 900–909 (2016).
- Chen, G. et al. Disorganization of white matter architecture in major depressive disorder: a meta-analysis of diffusion tensor imaging with tract-based spatial statistics. *Sci. Rep.* **6**, 21825 (2016).
- Guo, W. et al. Increased cerebellar-default-mode-network connectivity in drug-naïve major depressive disorder at rest. *Medicine* **94**, e560 (2015).
- Carmona, S. et al. Global and regional gray matter reductions in ADHD: a voxel-based morphometric study. *Neurosci. Lett.* **389**, 88–93 (2005).
- Power, J. D., Fair, D. A., Schlaggar, B. L. & Petersen, S. E. The development of human functional brain networks. *Neuron* **67**, 735–748 (2010).
- Krueger, R. F., Caspi, A., Moffitt, T. E. & Silva, P. A. The structure and stability of common mental disorders (DSM-III-R): a longitudinal-epidemiological study. *J. Abnorm. Psychol.* **107**, 216 (1998).
- Diedenhofen, B. & Musch, J. Cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS One* **10**, e0121945 (2015).
- Dunn, O. J. & Clark, V. Correlation coefficients measured on the same individuals. *J. Am. Stat. Assoc.* **64**, 366–377 (1969).
- Whelan, R. et al. Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature* **512**, 185–189 (2014).
- Lahey, B. B., Van Hulle, C. A., Singh, A. L., Waldman, I. D. & Rathouz, P. J. Higher-order genetic and environmental structure of prevalent forms of child and adolescent psychopathology. *Arch. Gen. Psychiat.* **68**, 181–189 (2011).
- Kessler, R. C. et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's world mental health survey initiative. *World Psychiat.* **6**, 168–176 (2007).
- Mayberg, H. S. Modulating dysfunctional limbic-cortical circuits in depression: towards development of brain-based algorithms for diagnosis and optimised treatment. *Br. Med. Bull.* **65**, 193–207 (2003).
- Witelson, S. F. Hand and sex differences in the isthmus and genu of the human corpus callosum: a postmortem morphological study. *Brain* **112**, 799–835 (1989).
- Tham, M. W., San Woon, P., Sum, M. Y., Lee, T. & Sim, K. White matter abnormalities in major depression: evidence from post-mortem, neuroimaging and genetic studies. *J. Affect Disord.* **132**, 26–36 (2011).
- Raichle, M. E. et al. A default mode of brain function. *Proc. Natl Acad. Sci.* **98**, 676–682 (2001).
- Buckner, R. L., Andrews-Hanna, J. R. & Schacter, D. L. The brain's default network. *Ann. N. Y. Acad. Sci.* **1124**, 1–38 (2008).

36. Ray, R. D. et al. Individual differences in trait rumination and the neural systems supporting cognitive reappraisal. *Cogn. Affect. Behav. Neurosci.* **5**, 156–168 (2005).
37. Stoodley, C. J. The cerebellum and cognition: evidence from functional imaging studies. *Cerebellum* **11**, 352–365 (2012).
38. Guggenmos, M. et al. Quantitative neurobiological evidence for accelerated brain aging in alcohol dependence. *Transl. Psychiat.* **7**, 1279–1286 (2017).
39. Hibar, D. P. et al. Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA bipolar disorder working group. *Mol. Psychiat.* **23**, 932–942 (2017).
40. McGorry, P. D., Hickie, I. B., Yung, A. R., Pantelis, C. & Jackson, H. J. Clinical staging of psychiatric disorders: a heuristic framework for choosing earlier, safer and more effective interventions. *Aust. N. Z. J. Psychiatry* **40**, 616–622 (2006).
41. Biswal, B., Zerrin Yetkin, F., Haughton, V. M. & Hyde, J. S. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* **34**, 537–541 (1995).
42. Schumann, G. et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol. Psychiat.* **15**, 1128–1139 (2010).
43. Goodman, R. The strengths and difficulties questionnaire: a research note. *J. Child Psychol. Psychiat.* **38**, 581–586 (1997).
44. Vulser, H. et al. Subthreshold depression and regional brain volumes in young community adolescents. *J. Am. Acad. Child Adolesc. Psychiat.* **54**, 832–840 (2015).
45. Kurth, F. & Lüders, E. VBM8. <http://www.neuro.uni-jena.de/vbm/download/> (2010).
46. The FIL Methods Group. SPM8. <https://www.fil.ion.ucl.ac.uk/spm/software/spm8/> (2009).
47. Ashburner, J. & Friston, K. J. Unified segmentation. *Neuroimage* **26**, 839–851 (2005).
48. Grellmann, C. et al. Comparison of variants of canonical correlation analysis and partial least squares for combined analysis of MRI and genetic data. *Neuroimage* **107**, 289–310 (2015).
49. Jones, D. K. et al. Isotropic resolution diffusion tensor imaging with whole brain acquisition in a clinically acceptable time. *Hum. Brain Mapp.* **15**, 216–230 (2002).
50. Smith, S. M. et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, S219 (2004).
51. Avants, B. B. et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011).
52. Pruim, R. H. et al. ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* **112**, 267–277 (2015).
53. Pruim, R. H., Mennes, M., Buitelaar, J. K. & Beckmann, C. F. Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI. *Neuroimage* **112**, 278–287 (2015).
54. Behzadi, Y., Restom, K., Liu, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101 (2007).
55. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
56. Power, J. D. et al. Functional network organization of the human brain. *Neuron* **72**, 665–678 (2011).
57. Hotelling, H. Relations between two sets of variates. *Biometrika* **28**, 321–377 (1936).
58. Witten, D. M. & Tibshirani, R. J. Extensions of sparse canonical correlation analysis with applications to genomic data. *Stat. Appl. Genet. Mol. Biol.* **8**, 1–27 (2009).
59. Parkhomenko, E., Tritchler, D. & Beyene, J. Sparse canonical correlation analysis with application to genomic data integration. *Stat. Appl. Genet. Mol. Biol.* **8**, 1–34 (2009).
60. Gifi, A. *Nonlinear Multivariate Analysis* (Wiley, 1990).
61. Jenkins, L. M. et al. Shared white matter alterations across emotional disorders: a voxel-based meta-analysis of fractional anisotropy. *NeuroImage Clin.* **12**, 1022–1034 (2016).
62. Goodkind, M. et al. Identification of a common neurobiological substrate for mental illness. *JAMA Psychiat.* **72**, 305–315 (2015).
63. Everitt, B. S. & Dunn, G. *Applied Multivariate Data Analysis Vol 2* (Arnold, 2001).
64. Timm, N. H. & Carlson, J. E. Part and bipartial canonical correlation analysis. *Psychometrika* **41**, 159–176 (1976).
65. O'Brien, L. M. et al. Statistical adjustments for brain size in volumetric neuroimaging studies: some practical implications in methods. *Psychiatry Res. Neuroimag.* **193**, 113–122 (2011).
66. Pell, G. S. et al. Selection of the control group for VBM analysis: influence of covariates, matching and sample size. *Neuroimage* **41**, 1324–1335 (2008).
67. Voevodskaya, O. et al. The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. *Front. Aging Neurosci.* **6**, 264 (2014).
68. Van Den, Wollenberg & Arnold, L. Redundancy analysis: an alternative for canonical correlation analysis. *Psychometrika* **42**, 207–219 (1977).
69. Stewart, D. & Love, W. A general canonical correlation index. *Psychol. Bull.* **70**, 160–163 (1968).
70. Monteiro, J. M., Rao, A., Shawe-Taylor, J. & Mourão-Miranda, J. Alzheimer's Disease Initiative. A multiple hold-out framework for sparse partial least squares. *J. Neurosci. Methods* **271**, 182–194 (2016).
71. Holmes, A. P., Blair, R. C., Watson, G. & Ford, I. Nonparametric analysis of statistic images from functional mapping experiments. *J. Cereb. Blood Flow. Metab.* **16**, 7–22 (1996).
72. Westfall, P. H. & Troendle, J. F. Multiple testing with minimal assumptions. *Biometrical J.* **50**, 745–755 (2008).
73. Westfall, P. H. & Young, S. S. *Resampling-based Multiple Testing: Examples and Methods for P-value Adjustment* Vol. 279 (Wiley, 1993).
74. Friedman, J., Hastie, T. & Tibshirani, R. *The Elements of Statistical Learning*. Springer Series in Statistics, Vol. 1 (Springer, 2001).
75. Aebi, M. et al. The use of the development and well-being assessment (DAWBA) in clinical practice: a randomized trial. *Eur. Child Adolesc. Psychiat.* **21**, 559–567 (2012).
76. Steinberg, L. Cognitive and affective development in adolescence. *Trends Cogn. Sci. Regul. Ed.* **9**, 69–74 (2005).
77. Schmaal, L. et al. Subcortical brain alterations in major depressive disorder: findings from the ENIGMA major depressive disorder working group. *Mol. Psychiat.* **21**, 806 (2016).
78. Rimol, L. M. et al. Cortical volume, surface area, and thickness in schizophrenia and bipolar disorder. *Biol. Psychiat.* **71**, 552–560 (2012).
79. van Erp, T. G. et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol. Psychiat.* **21**, 547 (2016).
80. von Rhein, D. et al. The NeuroIMAGE study: a prospective phenotypic, cognitive, genetic and MRI study in children with attention-deficit/hyperactivity disorder: design and descriptives. *Eur. Child Adolesc. Psychiat.* **24**, 265–281 (2015).
81. Hoogman, M. et al. Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *Lancet Psychiat.* **4**, 310–319 (2017).

Acknowledgements

This work received support from the following sources: the European Union-funded FP6 Integrated Project IMAGEN (reinforcement-related behaviour in normal brain function and psychopathology; grant no. LSHM-CT-2007-037286), the Horizon 2020-funded ERC Advanced Grant STRATIFY (brain-network-based stratification of reinforcement-related disorders; grant no. 695313), ERANID (understanding the interplay between cultural, biological and subjective factors in drug use pathways; grant no. PR-ST-0416-10004), BRIDGET (JPND: brain imaging, cognition, dementia and next generation genomics; grant no. MR/N027558/1), the Human Brain Project (HBP SGA 2, grant no. 785907), the FP7 project MATRICS (grant no. 603016), the Medical Research Council Grant c-VEDA (Consortium on Vulnerability to Externalizing Disorders and Addictions; grant no. MR/N000390/1), the National Institute for Health Research Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, the Bundesministerium für Bildung und Forschung (BMBF grant nos. 01GS08152 and 01EV0711; Forschungsnetz AERIAL 01EE1406A, 01EE1406B, 01ZX1314G, 01GS08147), the Deutsche Forschungsgemeinschaft (DFG grant nos. SM 80/7-2, SFB 940/2), the Medical Research Foundation and Medical Research Council (grant nos. MR/R00465X/1 and MR/S020306/1), the National Institutes of Health (NIH)-funded ENIGMA (grant nos. 5U54EB020403-05 and 1R56AG058854-01). Further support was provided by grants from: ANR (project AF12-NEUR0008-01-WM2NA and ANR-12-SAMA-0004), the Fondation de France, the Fondation pour la Recherche Médicale, the Mission Interministérielle de Lutte-contre-les-Drogues-et-les-Conduites-Addictives (MILDECA), the Assistance-Publique-Hôpitaux-de-Paris and INSERM (interface grant), Paris Sud University IDEX 2012; the NIH, Science Foundation Ireland (grant no. 16/ERC/D/3797), USA (Axon, Testosterone and Mental Health during Adolescence; grant no. RO1 MH085772-01A1) and NIH Consortium grant no. U54 EB020403, supported by a cross-NIH alliance that funds Big Data to Knowledge Centres of Excellence. A.M. gratefully acknowledges funding from the Netherlands Organization for Scientific Research via the Vernieuwingsimpuls VIDI programme (grant no. 016.156.415). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.I., C.C., I.M.V., P.G.S., H.L., T.J. and G.R. preprocessed the data. A.I. and P.G.S. analysed the data. A.I., G.S., F.B. and P.G.S. wrote the manuscript. A.I., G.S., T.W.R., A.M., J.A. and E.B. conceptualized the study. N.T., E.B.Q., T.W., S.D., T.B., A.L.W.B., U.B., C.B., P.C., T.F., H.F., V.F., H.G., P.S., P.G., Y.G., A.H., B.L., V.K., J.-L.M., A.M.-L., F.N., B.v.N., D.P.O., M.-L.P.M., S.M., J.P., L.P., M.S., A.S., M.N.S., H.W., R.W., O.A.A., I.A.,

E.D.B. and J.B. collected data. A.I. and N.T. prepared the figures. All authors revised the manuscript.

Boehringer Ingelheim, Brainsway, Elsevier, Lundbeck Int. Neuroscience Foundation and Science Advances. The other authors declare no competing interests.

Competing interests

T.B. served in an advisory or consultancy role for Lundbeck, Medice, Neurim Pharmaceuticals, Oberberg GmbH and Shire. He received conference support or a speaker's fee from Lilly, Medice, Novartis and Shire. He has been involved in clinical trials conducted by Shire and Viforpharma. He received royalties from Hogrefe, Kohlhammer, CIP Medien and Oxford University Press. The present work is unrelated to the above grants and relationships. E.D.B. received honoraria from General Electric Healthcare for teaching on scanner programming courses and acts as a consultant for IXICO. O.A.A. received a speaker's honorarium from Lundbeck. G.R. received financial support from scientific meetings (Janssen & Janssen, Otsuka–Lundbeck). A.M.-L. received consultant fees from

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41562-019-0738-8>.

Correspondence and requests for materials should be addressed to G.S.

Peer review information Primary Handling Editor: Mary Elizabeth Sutherland.

Reprints and permissions information is available at www.nature.com/reprints.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

¹Centre for Population Neuroscience and Precision Medicine, Institute of Psychiatry, Psychology & Neuroscience, Social, Genetic and Developmental Psychiatry Centre, King's College London, London, UK. ²Neuroimaging, Max Planck Institute of Psychiatry, Munich, Germany. ³Behaviour and Basal Ganglia Research Unit, University of Rennes, Rennes, France. ⁴Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands. ⁵Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. ⁶Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland. ⁷Systems Neuroscience, University Medical Centre Hamburg–Eppendorf, Hamburg, Germany. ⁸Department of Psychology, Stanford University, Stanford, CA, USA. ⁹Department of Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience King's College London, London, UK. ¹⁰Department of Psychiatry, CHU Ste Justine Hospital, Université de Montréal, Montreal, Quebec, Canada. ¹¹Institute of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. ¹²Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany. ¹³NeuroSpin, CEA, Université Paris–Saclay, Gif-sur-Yvette, France. ¹⁴Departments of Psychiatry and Psychology, University of Vermont, Burlington, VT, USA. ¹⁵Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, Nottingham, UK. ¹⁶Department of Psychiatry and Psychotherapy CCM, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany. ¹⁷Biomedical Magnetic Resonance, Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin, Germany. ¹⁸Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Charité, Humboldt University, Berlin, Germany. ¹⁹Institut National de la Santé et de la Recherche Médicale, INSERM Unit 1000 Neuroimaging & Psychiatry, University Paris Saclay, University Paris Descartes; Dlgiteo-Labs, Gif-sur-Yvette; and Maison de Solenn, Paris, France. ²⁰Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. ²¹Institut National de la Santé et de la Recherche Médicale, INSERM Unit 1000 Neuroimaging & Psychiatry, University Paris Saclay, University Paris Descartes; Dlgiteo-Labs, Gif-sur-Yvette, France. ²²AP-HP Sorbonne Université, Department of Child and Adolescent Psychiatry, Pitié-Salpêtrière Hospital, Paris, France. ²³Department of Social and Health Care, Psychosocial Services Adolescent Outpatient Clinic, Lahti, Finland. ²⁴Department of Child and Adolescent Psychiatry and Psychotherapy, University Medical Centre Göttingen, Göttingen, Germany. ²⁵Department of Psychiatry and Neuroimaging Centre, Technische Universität Dresden, Dresden, Germany. ²⁶Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. ²⁷Mood Brain and Development Unit, National Institute of Mental Health/NIH, Bethesda, MD, USA. ²⁸School of Psychology and Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland. ²⁹Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. ³⁰NORMENT, Institute of Clinical Medicine, University of Oslo, Oslo, Norway. ³¹Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway. ³²Department of Clinical Neuroscience, Centre for Psychiatric Research, Karolinska Institutet, Stockholm, Sweden. ³³Institut National de la Santé et de la Recherche Médicale, UMR 992 INSERM, CEA, Faculté de médecine, Université Paris-Sud, Université Paris-Saclay, NeuroSpin, Gif-sur-Yvette, France. ³⁴Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. ³⁵Wellcome Centre for Human Neuroimaging, UCL Institute of Neurology, University College London, London, UK. ³⁶Department of Psychology and Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK. ³⁷PONS Research Group, Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Humboldt University, Berlin and Leibniz Institute for Neurobiology, Magdeburg, Germany. ³⁸PONS Research Group, Institute for Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University, Shanghai, China. ³⁹PONS Research Group, Dept. of Psychiatry and Psychotherapy, Campus Charité Mitte, Humboldt University, Berlin, Germany. ⁴⁰PONS Research Group, Leibniz Institute for Neurobiology, Magdeburg, Germany. ⁴¹A list of authors and affiliations for the IMAGEN Consortium appears in the Supplementary Note. *e-mail: gunter.schumann@kcl.ac.uk

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection DAWBA questionnaire data was collected using an online questionnaire, more details can be found at: <http://dawba.info>

Data analysis Data preprocessing was conducted using standardized software packages such as FSL, SPM and ANTs. Multiple sparse canonical correlation analysis regression was implemented using custom written code. This code will be made available on publication of this paper.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data used in this study is available at the discretion of the principal investigators responsible for data collection.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

We utilized data from several different neuroimaging studies. Details pertaining to these different studies are given here:

IMAGEN: IMAGEN is a large-scale neuroimaging-genetics cohort study conducted with the aim of understanding the biological basis of individual variability in psychological and behavioural traits, and their relation to common psychiatric disorders. The study involves a thorough neuropsychological, behavioural, clinical and environmental assessment of each subject. Participants also undergo biological characterisation, with a collection of T1 weighted MRI (sMRI), diffusion tensor imaging (DTI), task-based fMRI (t-fMRI), resting-state fMRI (rs-fcMRI) and genetic data. We used T1 weighted, DTI, and rs-fcMRI data in the current investigation.

Munich Depression Sample: This is a clinical neuroimaging study, conducted with the aim of uncovering case-control differences in grey matter, associated with depression. In the present study, we used T1 weighted MRI data from this study.

Oslo Schizophrenia/Bipolar Sample (TOP): This is a clinical neuroimaging study, conducted with the aim of uncovering case-control differences in grey matter, associated with Schizophrenia/Bipolar. In the present study, we used T1 weighted MRI data from this study.

Netherlands ADHD Sample (NeuroImage): This is a clinical neuroimaging study, conducted with the aim of uncovering case-control differences in grey matter, associated with ADHD. In the present study, we used T1 weighted MRI data from this study.

Research sample

IMAGEN: For IMAGEN, a general population sample of Caucasian adolescents were recruited from eight sites across France, Ireland, England and Germany. Data was collected at age 14, 16 and 19 years. After a conservative quality control of MRI acquisitions and questionnaires used (see below) 666 (age: 18.8 [SD 0.61], years, 52% women). These participants can be considered to be representative of adolescents of this age drawn from the general population in Europe.

Munich Depression Sample: The Munich sample consisted of patients with first episode and recurrent unipolar depression treated as in-patients at the Max Planck Institute of Psychiatry, Munich, and healthy control subjects (N=614; 400 patients, age 48 [SD 13.8] years, 53% women; 214 control subjects age 49 [SD 13.3] years, 58% women).

Oslo Schizophrenia/Bipolar Sample (TOP): Participants with Schizophrenia and Bipolar Disorder were recruited from the Thematically Organised Psychosis (TOP) study. This is a collaborative study based at the University of Oslo in Norway. The study involved 286 controls (aged 34 [SD 9.5] years, 46% women), 161 schizophrenics (aged 32 [SD 8.8] years, 35% women) and 189 subjects with bipolar disorder (aged 34 [SD 11.5] years, 58% women).

Netherlands ADHD Sample (NeuroImage): Data for the ADHD sample was taken from the NeuroIMAGE project, a clinical cohort study. The study is made up of individuals tested at two different sites in the Netherlands, The Donders Centre for Cognitive Neuroimaging in Nijmegen, and the Vrije Universiteit in Amsterdam. Analyses were made on between 74 healthy controls (aged 18 [SD 2.0] years, 50% women) and 131 ADHD subjects (aged 18 [SD 2.3] years, 27% women).

Sampling strategy

IMAGEN participants were community-recruited from schools. Researchers introduced the objectives of the study to students via brain-science related classroom presentations, advertisements within school newsletters/websites and via presentations during school parent events. Extra effort was made to recruit adolescents prone to truancy to capture a particularly problematic slice of the population and to ensure representation.

Data collection

For IMAGEN, clinically-relevant data were collected using the DAWBA questionnaire administered online (www.dawba.info). Behavioural self-report/substance use/etc data were collected using Psytools software via its internet-based platform (Delosis Ltd, London, UK). Clinically relevant and behavioural questionnaires were reported when participant was alone. Parents were present for MRI collection at age 14. Researchers were present for all MRI time points.

Timing

For IMAGEN, age 14 data were collected beginning in 2008 ending in 2009. Age 19 data were collected beginning in 2013 ending in 2015.

Data exclusions

IMAGEN: At the time of the analysis conducted here, DAWBA/SDQ data had been collected for 1510 subjects. Of these, data was incomplete for 239 subjects, and was not used. At the time this investigation was conducted, T1 data had been acquired for 1400 subjects. All scans were visually inspected and manually reoriented. 285 scans were discarded from the analysis for either movement artifacts, strong field inhomogeneities, abnormal field of view, abnormally reduced cerebellum and for brace artefacts. At the time this study was conducted, DTI data had been acquired for 1412 subjects. Of these, 71 were not usable due to: signal dropouts or too much rotation. At the time of this investigation, we had collected rsfMRI scans for 1067 subjects. Of these scans, 157 were not used, either because over 5% of scans in that subject exhibited artifacts of some kind, or if over 5% of volumes showed a fractional displacement of over 0.5mm. Overall, 666 subjects had complete data after these QC procedures.

Munich Depression Sample: The T1 data for 13 of the subjects assessed was not used as it was deemed to be of insufficient quality, this left: N=614 subjects, of which 400 were patients.

Oslo Schizophrenia/Bipolar sample (TOP): The T1 data for 2 subjects was not used as it was considered to be of insufficient quality, this left: 286 Controls, 161 Schizophrenics (aged 32 [SD 8.8] years, 35% women) and 189 subjects with Bipolar Disorder.

The T1 data for 12 of the subjects was not used as it was deemed to be of insufficient quality, leaving data for 74 healthy controls and 131 ADHD subjects.

Non-participation

N/A

Randomization

The IMAGEN data was split into a training set and a hold out set. The training set was made up of 70% of the data, whilst the hold-out set constituted the remaining 30%. This train/test data split was made at random in the data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- n/a Involved in the study
- ☐ ☐ Antibodies
- ☐ ☐ Eukaryotic cell lines
- ☐ ☐ Palaeontology
- ☐ ☐ Animals and other organisms
- ☐ ☐ Human research participants
- ☐ ☒ Clinical data

Methods

- n/a Involved in the study
- ☐ ☐ ChIP-seq
- ☐ ☐ Flow cytometry
- ☐ ☒ MRI-based neuroimaging

Antibodies

Antibodies used

Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.

Validation

Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

State the source of each cell line used.

Authentication

Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.

Mycoplasma contamination

Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines
(See [ICLAC](#) register)

Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Palaeontology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.

Wild animals

Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

See above

Recruitment

In IMAGEN, participants were recruited through schools. Therefore, there is a potential self-selection bias in who was finally recruited. However, extra effort was made to recruit adolescents prone to truancy to capture a particularly problematic slice of the population and to ensure representation.

Ethics oversight

For each of the samples, ethics approval was gained at each local site.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection

Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.

Outcomes

Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session
(e.g. [UCSC](#))

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

Confirm that:

- ☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ All plots are contour plots with outliers or pseudocolor plots.
- ☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.

Instrument

Identify the instrument used for data collection, specifying make and model number.

Software

Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.

Cell population abundance

Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

Gating strategy

Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

- ☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type

Resting state, T1 weighted structural, DTI

Design specifications

N/A

Behavioral performance measures

N/A

Acquisition

Imaging type(s)

Resting state, T1 weighted structural, DTI

Field strength

IMAGEN-3T, Oslo (TOP)-1.5T, Netherlands (NeuroImage)-1.5T, Munich Depression Sample-1.5T

Sequence & imaging parameters

IMAGEN structural: T1 weighted, field of view: 286.1mm², matrix size: 256x256x170, slice thickness: 1.1mm, orientation: sagittal, TR: 2.3s, TE: 2.8ms, flip angle: 8 degrees.

NeuroImage structural: T1 weighted, field of view: 256mm², matrix size: 176x256x256, slice thickness: 1mm, orientation: sagittal, TR: 2.7s, TE: 2.9ms, flip angle: 7 degrees

TOP structural: T1 weighted, field of view: 240mm², matrix size: 180x255x160, slice thickness: 1mm, orientation: sagittal, TR: 2.73s, TE: 3.93msec, flip angle: 7 degrees

Depression structural: T1 weighted, field of view: 230mm², matrix size: 256x256, slice thickness: 1.2mm, orientation: sagittal, TR: 1.03s, TE: 2.1ms, flip angle: 90 degrees

IMAGEN rs-fMRI: EPI sequence, field of view: 218mm², matrix size: 64 x 64, slice thickness: 2.4mm, orientation: sagittal, TR: 2.2s, TE: 30ms, flip angle: 75 degrees

IMAGEN diffusion MRI: EPI sequence (b=0 and 32 directions with b-value 1300 s/mm²; axial slice plane; echo time: 104ms; 128x128x60 matrix; voxel size 2.4x2.4x2.4 mm), adapted to tensor measurements (for example, FA, mean diffusivity (MD)) and tractography analysis.

Area of acquisition

In all cases, a whole brain acquisition was used.

Diffusion MRI

☒ Used☐ Not used

Parameters EPI sequence (b=0 and 32 directions with b-value 1300 s/mm²; axial slice plane; echo time = 104ms; 128x128x60 matrix; voxel size 2.4x2.4x2.4 mm), adapted to tensor measurements (for example, FA, mean diffusivity (MD)) and tractography analysis.

Preprocessing

Preprocessing software

IMAGEN structural data: Baseline and Follow up two scans were preprocessed using both the 2008 version of the Voxel Based Morphometry toolbox (VBM8) running in SPM8 (v.5236). Given the young adults recruited in IMAGEN, we first used VBM8 in order to avoid using adult tissue probability maps (TPM) to initiate the segmentation process. The VBM8 toolbox segmentation relies on an adaptive Maximum a Posterior technique and TPMs used in VBM8 are for registration purposes only. Diffeomorphic registration (Dartel) was then used to register the images, and to generate the study-specific population average template. We then resliced the data to 1.5x1.5x1.5mm voxel size. Smoothing was carried out using an isotropic 8 mm full width at half maximum Gaussian smoothing kernel.

IMAGEN DTI data: Diffusion imaging data was pre-processed using software from the FSL toolbox (www.fmrib.ox.ac.uk/fsl). Tract based spatial statistics (TBSS) was used in all voxelwise analyses carried out on the diffusion tensor data. Pre-processing was carried out in the following manner: An affine registration was applied to the first B0 image for head motion and eddy current correction. Brain extraction was carried out using BET. Diffusion tensor fitting was then used to obtain fractional anisotropy (FA) maps for each subject. All participants' FA data was aligned into a common space using the non-linear registration tool FNIRT, using a b-spline representation of the registration warp field. The mean was then taken across all FA maps to create an FA averaged image. This map was then 'thinned' to create a mean FA skeleton, which was then thresholded at FA > 0.2, keeping only the major white matter tracts. Each subject's aligned FA data was then projected onto the mean skeleton. We then resliced these maps to 1.5x1.5x1.5mm voxel resolution, matching the VBM maps. We then used these skeletonised maps in all subsequent analyses.

IMAGEN rs-fcMRI data:

- 1) Motion correction was carried out, applying a rigid body registration of each volume to the middle volume (FSL MCFLIRT).
- 2) Non-brain tissue was removed (FSL BET).
- 3) Spatial smoothing was applied using a 5mm FWHM Gaussian kernel.
- 4) Independent component analysis (FSL MELODIC) was run for each data set. Artifact components were identified using an automatic classification algorithm, and subsequently regressed from the data (ICA-AROMA v0.3). ICA-AROMA has been shown to be as effective as motion parameter regression, with additional spike regression and 'scrubbing', in the removal of motion related effects on functional connectivity measures derived from resting state fMRI data. However, this procedure has the additional benefit that it preserves more signal of interest than these methods⁵³.
- 5) The resulting cleaned data set was de-trended (up to a third degree polynomial).
- 6) Co-registration to a high-resolution T1 image (FSL FLIRT using the BBR algorithm), and normalization to 2mm isotropic MNI standard space (ANTs) was carried out.
- 7) We used the CompCorr procedure to further clean the data of physiological noise. To do this: we created white matter (WM) and cerebrospinal fluid (CSF) masks by taking the mean of the WM and CSF segmentations from the VBM analysis, and thresholding them at 0.95, we then resliced these maps into the same space as the rsfMRI data. We then extracted timecourses from voxels within these regions, and took the first three principal components of this signal for both WM and CSF maps. These six principal component signals should represent non-neuronal signal. We then regressed this non-neuronal signal from voxel timecourses across the rest of the brain.
- 8) Lastly, preprocessed and normalized resting-state data sets were resliced to 3mm isotropic voxels.

Clinical structural data: all T1-weighted images underwent unified segmentation, before iterative, diffeomorphic image registration based on grey matter (GM) and white matter (WM) segments using Dartel. Smoothing was carried out using an isotropic 8 mm full width at half maximum Gaussian smoothing kernel. This was carried out in SPM12.

Normalization

IMAGEN structural Data: T1 weighted structural MRI data was normalised to MNI space using the DARTEL procedure implemented in VBM8.

IMAGEN resting state data: Co-registration was applied to a high resolution T1 image (FSL FLIRT using the BBR algorithm), and normalization to 2mm isotropic MNI standard space (ANTs) was carried out.

IMAGEN DTI data: DTI scans were normalised to MNI space using the tract based spatial statistics method implemented in FSL.

Clinical Structural Data: T1 weighted structural MRI data was normalised to MNI space using the DARTEL procedure implemented in SPM12.

Normalization template

Grey matter maps were aligned to a study specific template using DARTEL, before being transformed to MNI space.

DTI data was normalised to the FMRIB58_FA template defining MNI space.

Resting state data was normalised to the 2mm isotropic MNI standard space template (ICBM 152).

Noise and artifact removal

T1 weighted data was visually inspected, any data with serious artefacts was not used in any subsequent analyses.

Noise and artifact removal

DTI data was visually inspected, any data with serious artefacts was not used in any subsequent analyses.

Artifacts in the rs-fMRI data were identified using ArtRepair. Subjects where more than 5% of the scan slices were compromised by artifacts were removed from subsequent analyses. Further artefact filtering was carried out using ICA-AROMA 0.3. Artefact components were identified using an automatic classification algorithm, and subsequently regressed from the data. The resulting data was de-trended (up to a third degree polynomial).

Volume censoring

IMAGEN: At the time this investigation was conducted, T1 data had been acquired for 1400 subjects. All scans were visually inspected and manually reoriented. 285 scans were discarded from the analysis for either movement artifacts, strong field inhomogeneities, abnormal field of view, abnormally reduced cerebellum and for brace artefacts. At the time this study was conducted, DTI data had been acquired for 1412 subjects. Of these, 71 were not usable due to: signal dropouts or too much rotation. At the time of this investigation, we had collected rsfMRI scans for 1067 subjects. Of these scans, 157 were not used, either because over 5% of scans in that subject exhibited artifacts of some kind, or if over 5% of volumes showed a fractional displacement of over 0.5mm. Overall, 666 subjects had complete data after these QC procedures.

Munich Depression Sample: The T1 data for 13 of the subjects assessed was not used as it was deemed to be of insufficient quality, this left: N=614 subjects, of which 400 were patients.

Oslo Schizophrenia/Bipolar sample (TOP): The T1 data for 2 subjects was not used as it was considered to be of insufficient quality, this left: 286 Controls, 161 Schizophrenics (aged 32 [SD 8.8] years, 35% women) and 189 subjects with Bipolar Disorder.

ADHD Sample: The T1 data for 12 of the subjects was not used as it was deemed to be of insufficient quality, leaving data for 74 healthy controls and 131 ADHD patients.

Statistical modeling & inference

Model type and settings

We developed a new method, termed msCCA-regression to find multivariate relationships between psychiatric symptoms, and multiple neuroimaging modalities simultaneously; In this case, voxel-based morphometry (VBM), tract based spatial statistics (TBSS) and resting state functional connectivity neuroimaging measures. msCCA-regression analysis was carried out in the general population IMAGEN sample, when subjects were aged 19. Additional analyses were then applied in order to localize changes in brain structure. We then analyzed neuroimaging and symptom data at age 14 in order to determine whether this multivariate relationship already existed at this earlier time-point. Following this, we assessed the clinical significance of our findings by conducting case-control comparisons of the structural markers found in the IMAGEN analysis, in a clinical sample.

Effect(s) tested

In all cases, we sought to find relations between psychiatric symptoms and neuroimaging measures of brain structure and function.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☒ Both

Anatomical location(s)

In all cases, we sought to find relations between psychiatric symptoms and neuroimaging measures of brain structure and function.

Statistic type for inference
(See [Eklund et al. 2016](#))

A wide range of statistical tests were used in the present study. All msCCA-regression analyses used permutation testing to assess the significance of Pearson's correlation coefficient between psychiatric symptoms and neuroimaging measures.

Case-control analyses between healthy subjects and subjects suffering from clinical disorders used standard t-tests (one-sided).

Correction

Permutation testing was used to control the FWE rate in all cases where multiple comparison correction was required.

Models & analysis

n/a | Involved in the study

- ☐ ☒ Functional and/or effective connectivity
☐ ☒ Graph analysis
☐ ☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Pearson's pairwise correlation coefficient was used to map connectivity between different brain regions.

Graph analysis

Pearson's pairwise correlation coefficient was the functional connectivity measure. The dependent variable was a set of psychiatric symptoms.

Multivariate modeling and predictive analysis

Initial multivariate modeling was carried out using msCCA-regression with the aim of finding relations

between sets of psychiatric symptoms, and neuroimaging measures of brain structure and function. We assessed the in-sample correlation between symptoms and neuroimaging measures using permutation testing. We also assessed the significance of this model in an out-of-sample test.