BIROn - Birkbeck Institutional Research Online

Mihalik, A. and Ferreira, F.S. and Rosa, M.J. and Moutoussis, M. and Ziegler, G. and Monteiro, J.M. and Portugal, L. and Adams, R.A. and Romero-Garcia, R. and Vértes, P.E. and Kitzbichler, M.G. and Váša, F. and Vaghi, Matilde M. and Bullmore, E.T. and Fonagy, P. and Goodyer, I.M. and Jones, P.B. and Hauser, T. and Neufeld, S. and St. Clair, M. and Whitaker, K. and Inkster, B. and Prabhu, G. and Ooi, C. and Toseeb, U. and Widmer, B. and Bhatti, J. and Villis, L. and Alrumaithi, A. and Birt, S. and Bowler, A. and Cleridou, K. and Dadabhoy, H. and Davies, E. and Firkins, A. and Granville, S. and Harding, E. and Hopkins, A. and Isaacs, D. and King, J. and Kokorikou, D. and Maurice, C. and McIntosh, C. and Memarzia, J. and Mills, H. and O'Donnell, C. and Pantaleone, S. and Scott, J. and Fearon, P. and Suckling, J. and van Harmelen, A.-L. and Kievit, R. and Dolan, R. and Mourão-Miranda, J. (2019) Brain-behaviour modes of covariation in healthy and clinically depressed young people. Scientific Reports 9 (1), ISSN 2045-2322.

SCIENTIFIC REPORTS
natureresearch

OPEN

Brain-behaviour modes of covariation in healthy and clinically depressed young people

Agoston Mihalik[1,2], Fabio S. Ferreira[1,2], Maria J. Rosa[1,2], Michael Moutoussis[2,3], Gabriel Ziegler[4,5], Joao M. Monteiro[1,2], Liana Portugal[1,6], Rick A. Adams[1,2,3], Rafael Romero-Garcia[7,9], Petra E. Vértes[7,9], Manfred G. Kitzbichler[7,9], František Váša[7,9], Matilde M. Vaghi[2,3], Edward T. Bullmore[7,8,9,10], Peter Fonagy[11], Ian M. Goodyer[7,8], Peter B. Jones[7,8], NSPN Consortium*, Raymond Dolan[2,3] & Janaina Mourão-Miranda[1,2]

Understanding how variations in dimensions of psychometrics, IQ and demographics relate to changes in brain connectivity during the critical developmental period of adolescence and early adulthood is a major challenge. This has particular relevance for mental health disorders where a failure to understand these links might hinder the development of better diagnostic approaches and therapeutics. Here, we investigated this question in 306 adolescents and young adults (14–24 y, 25 clinically depressed) using a multivariate statistical framework, based on canonical correlation analysis (CCA). By linking individual functional brain connectivity profiles to self-report questionnaires, IQ and demographic data we identified two distinct modes of covariation. The first mode mapped onto an externalization/internalization axis and showed a strong association with sex. The second mode mapped onto a well-being/distress axis independent of sex. Interestingly, both modes showed an association with age. Crucially, the changes in functional brain connectivity associated with changes in these phenotypes showed marked developmental effects. The findings point to a role for the default mode, frontoparietal and limbic networks in psychopathology and depression.

Adolescence and early adulthood are periods of high risk for onset of many psychiatric disorders[1,2], with up to a fifth of 18 to 25 year olds seeking professional help for psychological distress[3]. Despite this there are, as yet, no biological measures that inform early diagnosis and treatment. Neuroimaging techniques, especially functional Magnetic Resonance Imaging (fMRI)[4], enable researchers to relate biological measures, such as patterns of functional brain connectivity, to the continuum of healthy to pathological states[5]. Here, we applied these techniques to uniquely identify underlying dimensions of brain-behaviour variation during a key developmental period.

Multivariate statistical methods[6], such as canonical correlation analysis (CCA)[7], allow an investigation of links between multiple sets of measures, such as brain imaging and behavioural data, collected from the same individuals. Recently, an emerging number of studies report links between individual patterns of functional brain connectivity and item-level measures of behaviour and mental symptoms[8–11]. To the best of our knowledge no study has investigated this relationship in adolescents and young people, including those with clinically diagnosed mental health problems.

[1]Centre for Medical Image Computing, Department of Computer Science, University College London, London, United Kingdom. [2]Max Planck University College London Centre for Computational Psychiatry and Ageing Research, University College London, London, United Kingdom. [3]Wellcome Centre for Human Neuroimaging, University College London, London, United Kingdom. [4]Institute of Cognitive Neurology and Dementia Research, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany. [5]German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany. [6]Department of Physiology and Pharmacology, Federal Fluminense University, Niteroi, Brazil. [7]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom. [8]Cambridgeshire and Peterborough National Health Service Foundation Trust, Cambridge, United Kingdom. [9]Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, United Kingdom. [10]ImmunoPsychiatry, GlaxoSmithKline Research and Development, Stevenage, United Kingdom. [11]Research Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom. Agoston Mihalik and Fabio S. Ferreira contributed equally. *A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to A.M. (email: a.mihalik@ucl.ac.uk)

In the current study, we exploited resting-state fMRI (rsfMRI) and extensive item-level self-report questionnaire data, IQ and demographic information (that we refer as 'behavioural data' for simplicity) to investigate relationships between individual patterns of functional brain connectivity and individual sets of psychometrics/IQ/demographics. We used a sample of 281 healthy and 25 clinically depressed subjects, comprising adolescents and young adults (14–24 y, 165 females) from the NeuroScience in Psychiatry Network (NSPN) study[12].

All 306 subjects completed self-report questionnaires assessing well-being, affective symptoms, anxiety, impulsivity, compulsivity, self-esteem, self-harm, personality characteristics, psychotic spectrum symptoms, substance use, relations with peers and family and experience of trauma. These item-level measures were supplemented with measures of subjects' fluid and crystallized intelligence as well as additional demographic information (age, gender and socioeconomic deprivation index) amounting to a total of 364 behavioural (i.e. psychometrics/IQ/demographics) measures for each subject.

We acquired anatomical MRI scans and approximately 11 minutes of rsfMRI from the 306 subjects, which were then pre-processed as described in Methods. From the rsfMRI data, we extracted averaged time-series from 348 brain regions using subcortical regions from Freesurfer[13] and the multi-modal parcellation of the Human Connectome Project (HCP)[14]. We then estimated functional brain connectivity for each individual using full correlation (Pearson-correlation) between all pairs of regional time-series. This resulted in a single connectivity profile (348×347 regions, i.e. 60378 connections) for each subject.

We used canonical correlation analysis (CCA) to investigate the relationships between brain connectivity and behaviour profiles across individuals. CCA is a multivariate analysis technique, which seeks maximal correlations between linear combinations of two or more sources of data (e.g. brain connectivity and behavioural data). To reduce the high dimensionality of both brain and behavioural data (60378 and 364 variables, respectively) we applied principal component analysis (PCA). We then performed a CCA resulting in pairs of canonical variates, which define modes of covariation between linear combinations of brain connectivity and behaviour profiles (in short 'CCA modes'). On these canonical variates, subjects are represented by brain and behaviour scores, which describe the subject specific loadings in each CCA mode. We used a permutation approach[15] to estimate both the optimal number of PCA components and the statistical significance of the CCA modes. Finally, we applied CCA embedded in a multiple hold-out framework[16] to investigate the generalizability of the model, i.e. to assess whether the CCA mode represent associations that can be found on new data.

Given the age range of our sample, we expected a strong age (or developmental) effect on the brain-behaviour modes of covariation. We suspected that variation in these modes might be related to the presence of depression, given that our sample also included clinically depressed subjects. Finally, we hypothesized that psychopathological symptoms might be associated with a core set of abnormal functional brain networks incorporating default mode, frontoparietal and limbic networks as suggested by recent literature[17,18].

Results

We found two significant modes of covariation (Fig. 1) between patterns of functional brain connectivity and sets of behavioural measures, using a permutation approach (Supplementary Methods and Fig. S1). These were based on an optimal number of PCA components, $d = 25$, which explained 53% and 56% of the behaviour and brain connectivity variance, respectively. The first and second modes yielded canonical correlations of $q = 0.62$ ($p_{FWE} < 0.0001$) and $q = 0.58$ ($p_{FWE} < 0.0134$), respectively.

Figure 1 shows the two significant brain-behaviour modes of covariation, representing the correlation between brain and behaviour scores of individual subjects. The first mode is associated with sex and has an interaction with depression, with healthy males clustering towards higher scores and depressed females clustering towards lower scores (Fig. 1a). Additionally, younger adolescents can be seen to have lower scores whereas older ones are distributed more towards higher scores (Fig. 1c). The characteristics of the second mode were qualitatively different. Although depressed females seemed to cluster towards lower scores (Fig. 1b) again, both males and females were evenly distributed along this mode, and younger adolescents had higher scores whereas older ones were more distributed towards lower scores (Fig. 1d).

To inform the association captured by each mode, we correlated the original behavioural and connectivity variables with the subjects' brain and behaviour scores, respectively, which revealed the behavioural variables or brain connections most strongly associated with each CCA mode (Figs 2–4). Figure 2a shows that the first CCA mode is positively associated with being male, age, measures of impulsivity, sensation seeking, drinking habits, and negatively associated with being female, depression-related symptoms and suicidal thoughts (for details, see Supplementary Table S1). Thus, the first mode has characteristic of an externalization/internalization axis, where extreme positive and negative scores represent vulnerability for males and females, respectively. Importantly, sex was weakly associated with the other top identified behavioural items (i.e. items most positively or negatively correlated with the behavioural variables) suggesting that those items are present due to an association with brain connectivity and not because of their association with sex (Supplementary Fig. S2). Brain connections most positively correlated with the first CCA mode (denoted by red edges in Fig. 3) included nodes within the dorsal and ventral attention networks and a somatomotor network; brain connections most negatively correlated (denoted by blue edges in Fig. 3) included nodes of the default mode, limbic and frontoparietal networks. Similar overall patterns were observed using different thresholds of top connections (Supplementary Fig. S3). In addition, when looking at the 0.5% most negatively correlated connections (top 302 connections), the subcortical network (mostly thalamus and caudate nucleus) also appeared negatively correlated with the first mode (including subcortical-subcortical connections and cortical connections with the default mode network, Supplementary Fig. S3). The list of 20 brain connections most positively/negatively associated with the first mode and their assignment to anatomical regions are described in Supplementary Table S2 and displayed on Supplementary Fig. S4.
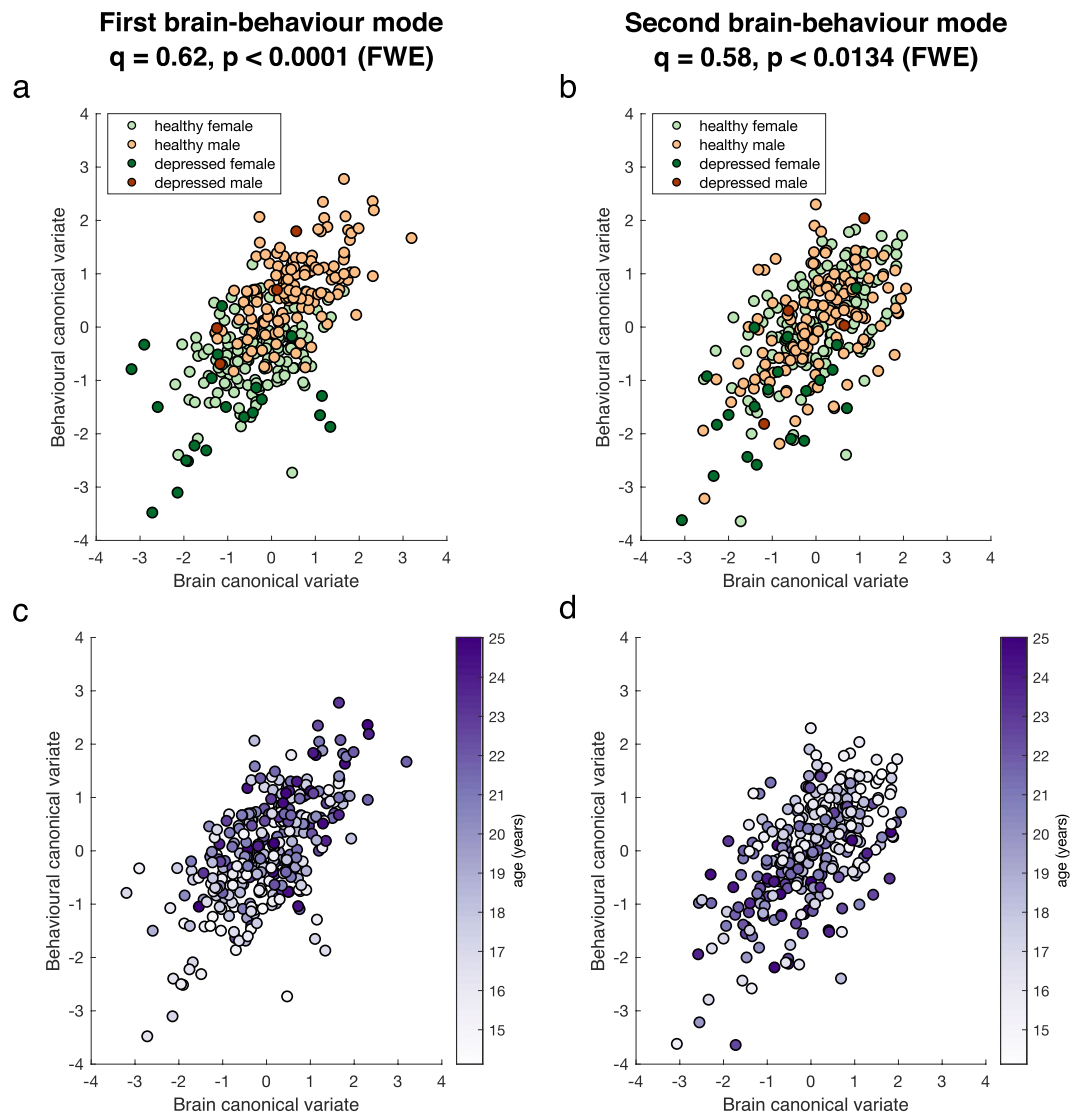
Figure 1. Significant brain-behaviour modes of covariation. Scatter plots showing the brain and behaviour scores for the first (**a,c**) and second (**b,d**) mode, where each dot represents an individual subject. Subjects are colour coded by: sex and clinical diagnosis (**a,b**); age (**c,d**). The canonical correlation, *q*, and corresponding *p*-value are shown on the top of each plot.

For the second mode, the most positively correlated behavioural variables (Fig. 2b) related to measures of mental well-being, self-esteem, confidence, while the most negatively associated related to age, depression-related symptoms, drinking habits, suicidal thoughts and sexual abuse. Thus, this second mode captures a well-being/distress axis, along which individuals vary from high mental well-being through to distress (for details, see Supplementary Table S3). The brain connections most positively correlated (depicted in red edges in Fig. 4) with this CCA mode included nodes involving mainly the default mode and subcortical networks (thalamus); brain connections most negatively correlated (depicted in blue edges in Fig. 4) included nodes within the dorsal and ventral attention networks and the visual and somatomotor networks. A largely similar overall pattern of networks was observed using different thresholds of top connections (Supplementary Fig. S3). In addition, when looking at the 0.5% most positively correlated connections (top 302 connections), the limbic and frontoparietal networks also appeared positively correlated with the second mode (including cortico-cortical connections and subcortical connections mostly with the thalamus, putamen and accumbens nucleus). The list of 20 brain connections most positively/negatively associated with the second mode and their assignment to anatomical regions are described in Supplementary Table S4 and displayed on Supplementary Fig. S5.

As validation of our model, we also applied a multiple hold-out framework (Supplementary Methods and Fig. S6) and found one brain-behaviour mode of covariation (Supplementary Fig. S7), based on ten PCA components ($d = 10$), which explained 40% and 47% of the behaviour and brain connectivity variance, respectively. Importantly, the distribution of subjects along the CCA main axis showed the same trend in the training and test sets (Supplementary Fig. S8). When ranking the original behavioural and connectivity variables according to their correlation with the subjects' brain and behaviour scores, we obtained a very similar overall ranking for
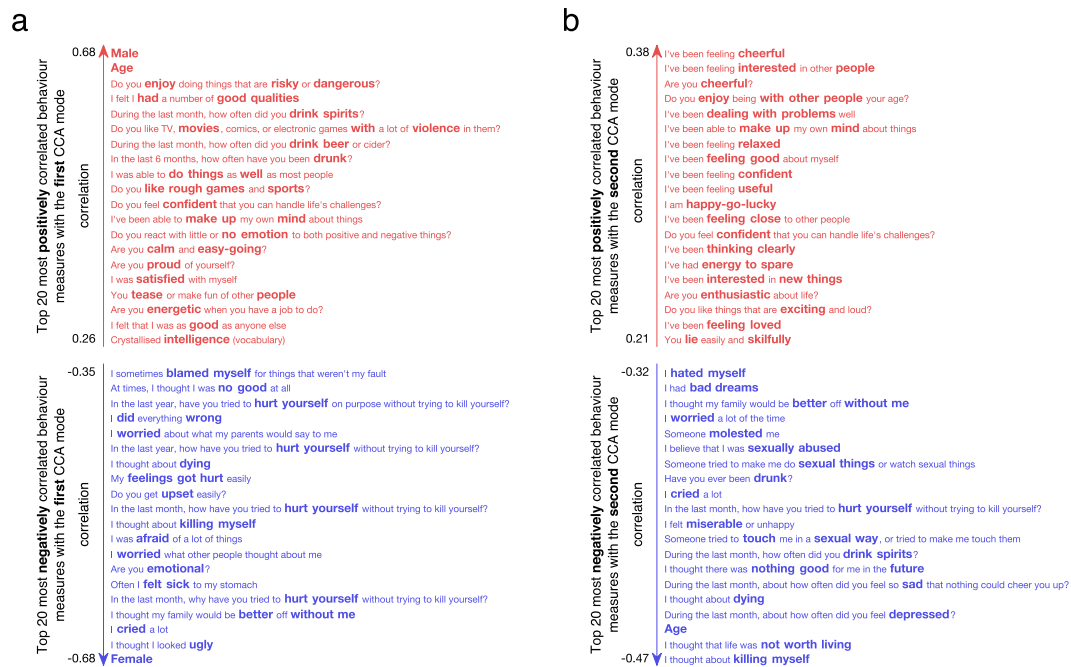
Figure 2. Correlations between the behavioural variables and behavioural canonical variate (behaviour scores of all subjects) of the first (**a**) and second (**b**) CCA modes. Top 20 most positively and top 20 most negatively correlated variables are shown only. The list of correlation values and questionnaire items can be found in Supplementary Tables S1 and S3.
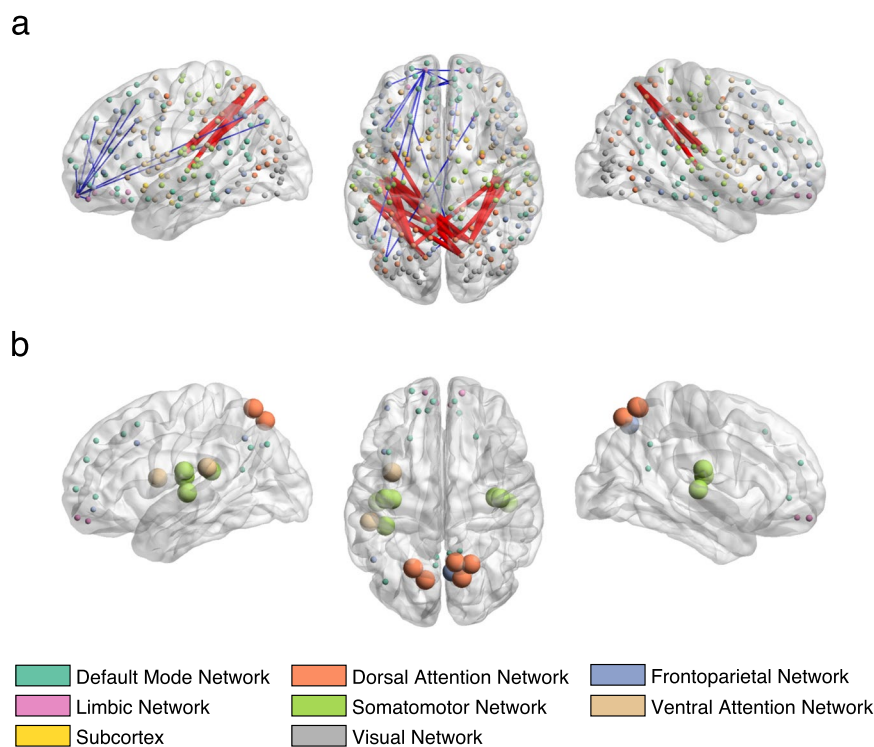

Figure 3. Correlations between the brain connectivity variables and brain canonical variate (brain scores of all subjects) of the first CCA mode in sagittal (left and right) and axial view (middle). (**a**) Top 20 most positively and top 20 most negatively correlated brain connections. The thickness of the edges is proportional to the absolute correlation (red for positive correlations and blue for negative correlations). (**b**) Top 20 most positively and top 20 most negatively correlated brain connections summarized by nodes. Node size is proportional to the mean absolute correlation. Nodes are colour coded by resting state networks assigning each node to one of the 7 cortical networks (based on the maximal surface based overlap) described in Yeo *et al.*[44] or the subcortex. The list of correlation values and respective labels can be found in Supplementary Table S2.

a

b

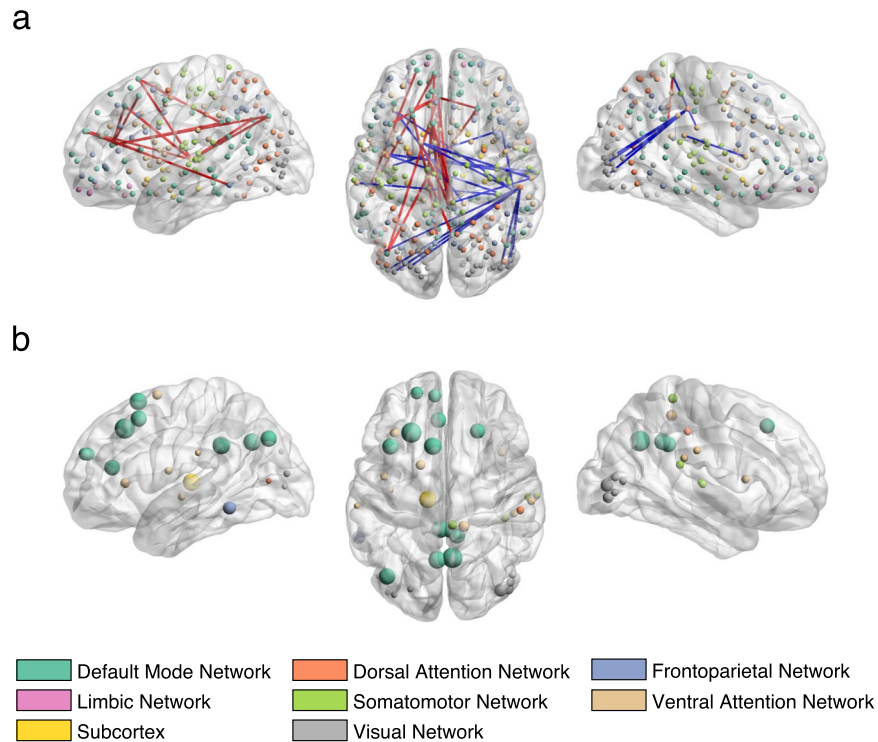| ■ Default Mode Network | ■ Dorsal Attention Network | ■ Frontoparietal Network |
| ■ Limbic Network | ■ Somatomotor Network | ■ Ventral Attention Network |
| ■ Subcortex | ■ Visual Network | |

Figure 4. Correlations between the brain connectivity variables and brain canonical variate (brain scores of all subjects) of the second CCA mode in sagittal (left and right) and axial view (middle). (**a**) Top 20 most positively and top 20 most negatively correlated brain connections. The thickness of the edges is proportional to the absolute correlation (red for positive correlations and blue for negative correlations). (**b**) Top 20 most positively and top 20 most negatively correlated brain connections summarized by nodes. Node size is proportional to the mean absolute correlation. Nodes are colour coded by resting state networks assigning each node to one of the 7 cortical networks (based on the maximal surface based overlap) described in Yeo *et al.*[44] or the subcortex. The list of correlation values and respective labels can be found in Supplementary Table S4.

both the permutation and hold-out frameworks (Supplementary Fig. S9). However, the overlap was not very large when we considered alone the top 20 most positively and top 20 most negatively correlated behavioural and brain variables (Supplementary Tables S1 and S2 and Tables S5 and S6). Indeed, the scatter plots in Supplementary Fig. S9 show that only 11 brain connectivity and 11 behavioural variables overlapped when those top variables were chosen. This might be explained by the fact that we are just displaying the very top variables ranked according to their correlation value, and some correlation values simply differ from each other on the fourth decimal place. However, we can see that the overlap is more extended when the top 5% most positively and top 5% most negatively correlated variables are selected (Supplementary Fig. S9).

Discussion

In summary, leveraging both resting-state fMRI and behavioural data within a multivariate analysis framework, we identified two brain-behaviour modes of covariation in a sample of 306 adolescents and young adults. The first CCA mode relates to an externalization/internalization axis which is associated with sex. Specifically, it suggests that males are more susceptible to disruptive behaviour and alcohol use, whilst females are more susceptible to depression and self-harm. The second CCA mode relates to a well-being/distress axis which covers positive symptoms of well-being on one side and negative symptoms related to depression, suicidal thoughts, history of sexual abuse and alcohol use on the other side. Both modes are also associated with age, which could be expected considering that the sample age range covers an important developmental period. Importantly, the brain networks related to both CCA modes align well with models of brain development highlighting the sequential maturation of subcortical and cortical regions in adolescence[19,20] and models of psychopathology[17,18].

Both CCA modes are conceptually associated with broadly described depressive psychopathology and can hence be seen as helping refine this clinical concept. It is therefore important to understand whether they capture distinctions in brain connectivity profiles alone or capture also distinctions in descriptive psychopathology. At first glance, the behavioural items common to both modes of depression, such as e.g. "…life was not worth living", "I thought about dying", "I cried a lot" might support the former hypothesis. Nevertheless, there are three clear differences:

Firstly, the first mode is associated with a more anxious, agitated and behaviourally-activated expression of depression (four self-harm items, "I felt sick…", "I worried…", "I was afraid…", "Are you emotional?"). Conversely, the second mode is associated with a more anhedonic and amotivational state (positively correlated

with "…life was not worth living", "…nothing good for me in the future", "…feel so sad…" and negatively correlated with "…feeling interested in other people"). Interestingly, similar 'anxious' and 'anhedonic' axes have been found in other large data-driven depression studies[9,21].

Secondly, the first CCA mode is strongly correlated with sex, but the second mode is not. Thus the latter is a more sex independent dimension of psychopathology. Furthermore, depression-related variables of the first mode are associated with younger age, whilst depression-related variables of the second mode are associated with older age (Fig. 1c,d and Fig. 2). Accordingly, depression in the first CCA mode is related to behavioural items, such as e.g. "…I looked ugly", "…my family would be better off without me", "I worried about what my parents would say…", which are more likely to be hallmarks of depression at a younger age. On the contrary, distress in the second CCA mode is related to items thought to characterise depression at an older age (e.g. "I thought about killing myself", being drunk and drinking spirits).

Thirdly, depression in the second mode is associated with sexual abuse and is negatively associated with feeling loved, confident and close to other people, perhaps indicating that sexual abuse affects these traits (however, causal attributions are not possible in this dataset).

The strong relationship between sex and the first CCA mode is striking in light of recent findings that there is <10% overlap in gene expression changes in the brains of male and female humans with depression – at least in the prefrontal cortex and insula (other cortical areas were not sampled)[22]. Moreover, the authors demonstrated that a similar lack of overlap between the sexes also exists in a chronic variable stress mouse model[22]. It is interesting that both insula and the prefrontal cortex dominate the connections of the first CCA mode, being either positively (insula) or negatively (prefrontal cortex) correlated with depression. This suggests that sex interacts with depression risk in these (and likely other) areas in a way that might be fundamental to the disorder.

Adolescence and early adulthood is the peak age of onset for many psychiatric disorders[1,2], rendering understanding vulnerability of individuals at this age of particular relevance. Importantly, most items correlated with the CCA modes are related to psychopathology, and so the identified CCA modes might represent a two-dimensional space not only related to current depressive symptoms (or their absence), but to a latent vulnerability to psychopathology. Deeper understanding of this vulnerability may powerfully inform biologically informed interventions in young people[23].

Substance use is highly correlated with psychiatric disorders[24,25], and it is especially detrimental in adolescence. Personality traits have an etiological role in the development of alcohol and substance use, and a vast body of research implicates two broad personality domains with opposing action tendencies, namely inhibition and disinhibition[26,27]. Our results concur with such a model. Alcohol usage is associated with both of our CCA modes in opposing directions. Behavioural items resembling to a disinhibited personality (first CCA mode) are positive correlations with e.g. "…enjoy doing things that are risky and dangerous?", "…like TV, movies, comics, or electronic games with a lot of violence in them?" or "Do you like rough games and sports?"; whereas items suggestive of an inhibited personality (second CCA mode) are negative correlations with e.g. being interested in or enjoying the company of other people, or being interested in new things.

As discussed above, age was associated with both CCA modes. The first CCA mode correlated positively with age (depicted in red in Fig. 2), attentional and frontoparietal networks (depicted in red in Supplementary Fig. S3a–c) and negatively correlated with subcortical-subcortical connections as well as connections within the limbic system (depicted in blue in Supplementary Fig. S3a–c). These results are consistent with models of adolescent brain development, demonstrating that subcortical and limbic regions mature in early adolescence followed by the maturation of cortico-cortical connections[19,20]. The second CCA mode was negatively correlated with age (depicted in blue in Fig. 2), connections within and between attentional networks (depicted in blue in Supplementary Fig. S3d–f) and was positively correlated with various subcortical-cortical connections (depicted in red in Supplementary Fig. S3d–f). Again, these results corroborate the aforementioned models of adolescent brain development. In particular, the results of the two CCA modes substantiate the sequential maturation of brain circuits, namely, the fine-tuning of circuits from subcortical-subcortical (early adolescence) to cortico-subcortical (late adolescence) and cortico-cortical (young adulthood)[28]. Furthermore, the sequential maturation of brain circuits might be a risk factor for alcohol use[29], which aligns well with the strong positive correlation between alcohol use and age found in both CCA modes (Fig. 2).

Our connectivity results are also consistent with recent literature suggesting that most psychiatric disorders emerge as a result of impairments within a few core brain circuits and networks[10,17,18]. In particular, the first mode was negatively correlated with depression and connections of the default mode, frontoparietal and limbic networks (Fig. 3); whilst the second mode was negatively correlated with depression and positively correlated with many default mode areas (Fig. 4). These networks underlie core social, executive and affective cognition, respectively, and dysfunctions in these networks might result in specific domains of symptoms (e.g. alterations in default mode network connectivity resulting in impaired self-representation and social functioning)[17]. Interestingly, due to the strong interplay between these networks, the aberrant functioning in any of these could cause impairments of the others. For example, excessive coupling between the limbic and default mode networks could mean that initial dysfunction in the former may propagate to the latter, causing depressive symptoms[10,30,31]. Conversely, a default mode network that can only dominate but cannot reciprocally communicate with the limbic network could prevent positive mood being established by the latter[32].

It is common practice in statistics to estimate a model using the entire dataset, once model selection has been performed using the permutation-based approach as described above. In contrast, in machine learning a dataset is often split into training and test sets (or training/test/validation sets) using cross-validation procedures, such that the model parameters are estimated on training data and the model performance is estimated on a test data. The permutation-based approach may be preferable to cross-validation when the number of samples is not very large, since it avoids the need to split the data into even smaller sets for training and testing. However, cross-validation approaches might be preferable if one wants to measure the robustness and generalisability of

the results. In the present study, we also ran CCA embedded in a multiple hold-out framework (Supplementary Methods and Fig. S6) which was proposed by Monteiro et al.[16]. We found one mode of covariation, which was comparable to the first one found using the permutation framework (Supplementary Fig. S9). The second mode obtained using the permutation framework was not found with the hold-out framework, potentially due to the small sample size and the strictness of the multiple hold-out framework. The most striking finding obtained with this approach was that the distribution of the subjects along the CCA main axis on the test set was very similar to the training set (Supplementary Fig. S8), which means that the CCA mode generalised for the test set.

Finally, we acknowledge limitations to the current study. Methodological limitations relate to the pipeline choice, which includes use of an atlas, full correlation as a connectivity metric and the PCA dimensionality reduction step that might remove a significant amount of signal variability of potential interest. Further work exploring other approaches to estimating resting state connectivity (e.g. Independent Component Analysis (ICA)[8] and partial correlation[33]) and regularised or sparse CCA could be investigated to overcome potential limitations of the current pipeline. We also acknowledge the cross-sectional nature of our analysis. Future studies involving longitudinal samples could investigate how the described brain-behaviour modes of covariation change over time or whether they are predictive of future psychopathology. In addition, although we have used a multiple hold-out framework, we should ideally use an independent replication sample to validate our model. Finally, there are limitations of sample size. Future studies including big datasets, such as the ABCD study[34], will be useful to explore higher variability in general population and potentially find different dimensions of psychopathology or groups of adolescents at risk.

In conclusion, our results demonstrate that identifying brain-behaviour modes of covariation in healthy and depressed young people provides a powerful way of understanding the latent dimensions underlying abnormal mental states and behaviour[35] and brings potential new insights into the mediation of vulnerability to mental health disorders.

Methods

Subjects.   In total, 2406 healthy subjects and 50 subjects clinically diagnosed with depression (diagnosis and referral made by the subject's NHS care service) aged 14 to 24 years were recruited from schools, colleges, National Health Service (NHS) primary care and mental health services, and via direct advertisement in London and Cambridgeshire[12]. This was carried out by the University College London and University of Cambridge NeuroScience in Psychiatry Network (NSPN) research initiative, supported by a strategic award from the Wellcome Trust. A Magnetic Resonance Imaging (MRI) cohort was subsampled from the primary cohort, comprising a healthy cohort of 318 subjects and a depression cohort of 37 subjects. Furthermore, a demographically balanced cohort of 297 subjects was subsampled from the healthy cohort, with approximately 60 subjects in each of five age-defined strata: 14–15, 16–17, 18–19, 20–21, and 22–24 years inclusive.

Of the healthy cohort ($n = 297$), 2 subjects were excluded due to low quality images, 1 was excluded due to gross radiological abnormalities, 4 were excluded due to missing convergence in ME-ICA pre-processing, and 9 were excluded due to excessive motion during the resting-state functional scan (5 subjects with maximum framewise displacement larger than 1.3 mm and 4 subjects with mean framewise displacement larger than 0.3 mm using calculation by Power et al.[36]). Of the depression cohort ($n = 37$), 3 subjects were excluded due to low quality anatomical scans, 1 was excluded due to radiological artefacts, 4 were excluded due to motion-induced Freesurfer reconstruction errors, 1 was excluded due to lack of convergence in ME-ICA pre-processing, 1 was excluded due to extremely low explained variance in ME-ICA pre-processing ($<20\%$) and 2 were excluded due to excessive motion during the resting-state functional scan (applying the same criteria as for the healthy cohort). These exclusion criteria produced a final healthy cohort consisting of 281 subjects (mean age = 19.13, SD = 2.88, 144 females) and a final depression cohort comprising 25 subjects (mean age = 16.80, SD = 1.15, 21 females).

Written informed consent was obtained for all subjects over the age of 16 years. For subjects under the age of 16, written informed assent was obtained for the subject and written informed consent from their parent/legal guardian. The study was ethically approved by the Cambridge Central Research Ethics Committee and conducted in accordance with NHS research governance standards.

Behavioural and demographic data.   Subjects completed self-report questionnaires and cognitive tests as part of the NSPN data acquisition[12]. We used the following subset of these measures at the baseline study visit that assess psychopathological symptoms, personality characteristics, mental well-being and IQ as follows: Antisocial Behaviours Checklist (ABQ); Antisocial Process Screening Device (APSD); Barratt Impulsivity Scale (BIS); Child and Adolescent Dispositions Scale (CADS); Child Trauma Questionnaire (CTQ); Drugs Alcohol and Self-Injury (DASI); Inventory of Callous-Unemotional Traits (ICU); Kessler Psychological Distress Scale (K10); Leyton Obsessional Inventory (LOI); Moods and Feelings Questionnaire (MFQ); Revised Children's Manifest Anxiety Scale (RCMAS); Rosenberg Self-Esteem Scale (SES); Schizotypal Personality Questionnaire (SPQ); Wechsler Abbreviated Scale of Intelligence (WASI); Warwick Edinburgh Mental Wellbeing Scale (WEMWBS).

Finally, we added four demographic variables (age, sex, and socioeconomic deprivation index; hot coding was used for sex resulting in 2 variables; see Supplementary Material for the calculation of the deprivation index) to the items of these questionnaires resulting in a total of 372 variables.

MRI data acquisition.   All MRI data were acquired on three identical 3T whole-body MRI systems (Magnetom TIM Trio; VB17 software version; Siemens Healthcare): two located in Cambridge and one located in London. Between-site reliability and tolerability of all MRI procedures were satisfactorily assessed by a pilot study of five healthy volunteers at each site[37]. Only scans at the baseline visit were included in the current study. Structural MRI scans were acquired using multi-echo acquisition protocol with six equidistant echo times (TE) between 2.2 and 14.7 ms and averaged to form a single image of increased signal-to-noise ratio (SNR)[37]. Apparent longitudinal relaxation rate R1 (R1 = 1/T1w) was calculated using previously developed

models to create quantitative R1 maps[38–40]. Other acquisition parameters included: temporal resolution (TR) of 18.70 ms, spatial resolution 1.0 mm isotropic, field of view (FOV) = 256 × 256, 176 sagittal slices and parallel imaging using GRAPPA factor 2 in anterior-posterior phase-encoding (PE) direction. Resting-state fMRI (rsf-MRI) data were acquired using multi-echo acquisition protocol with three echo times TE = 13, 31, 48 ms, temporal resolution (TR) of 2420 ms, spatial resolution 3.8 mm isotropic with 10% gap, sequential slice acquisition, FOV = 240 × 240 mm, 34 oblique slices; bandwidth 1⁄4 2368 Hz/pixel and matrix size = 64 × 64 × 34.

Structural MRI pre-processing. R1 images were used to perform surface reconstruction of each subject using Freesurfer's *recon-all*[41] (https://surfer.nmr.mgh.harvard.edu/). Freesurfer average subject (fsaverage) was parcellated using a multimodal scheme that subdivides the cortex into 360 bilaterally symmetric regions based on Human Connectome Project (HCP) data[14]. HCP parcellation was transformed from fsaverage space to the cortical surface of each individual subject using Freesurfer's *mri_surf2surf*. In addition, 16 regions were used from the subcortical segmentation of Freesurfer (thalamus-proper, caudate, putamen, pallidum, hippocampus, amygdala, accumbens-area, ventral diencephalon (DC) for each hemisphere).

Resting-state fMRI pre-processing. rsfMRI data were pre-processed using multi-echo independent component analysis[42,43] (ME-ICA). ME-ICA identifies BOLD components that scale linearly with TE and discards remaining components to reduce motion-related artefacts. Only BOLD components were optimally combined to generate the denoised time-series of each voxel. A wavelet filtering was used to focus on the physiologically relevant frequency range of 0.025–0.111 Hz (scales 2 and 3). Functional scans were coregistered with each individual structural R1 images for time-series extraction. Regional time-series were estimated as the average time-series of all the voxels included in each of the 360 cortical and 16 subcortical regions. 28 regions (mostly near the frontal and temporal pole) were excluded due to low regional mean signal in at least one subject (z-score across regions within subject, $z < -1.96$), resulting in a total of 348 retained regions. Functional connectivity was calculated as the pairwise Pearson-correlation between time-series of each possible pair of regions resulting in a total of 60378 brain variables per subject.

Behaviour and demographic data processing. The initial considered questionnaire data comprised 372 item level variables. However, we removed 8 variables for which more than 95% of subjects had the same value. Missing data were replaced by the median of the respective variable across subjects. This resulted in a total of 364 behaviour variables per subject.

Confounds. We identified two main confound measures, whose effect was not of interest and were therefore regressed out from both brain and behavioural data:

1. Mean framewise displacement: a summary statistic quantifying average subject head motion during the rsfMRI acquisitions (using calculation by Power *et al.*[36]).
2. Site: each MRI site (two in Cambridge and one in London) was encoded as a one-hot variable (for example: [0 0 1] for London).

Finally, each brain and behavioural variable was mean-centred and normalised ($\mu = 0$, $\sigma = 1$), and these brain ($\mathbf{X} \in \mathbb{R}^{n \times p}$, where n is the number of subjects and p is the number of brain connectivity variables) and behaviour ($\mathbf{Y} \in \mathbb{R}^{n \times q}$, where n is the number of subjects and q is the number of behavioural variables) matrices were then entered into the CCA analysis.

Canonical correlation analysis. We used Canonical Correlation Analysis (CCA) to investigate modes of covariation between patterns of brain connectivity and behavioural data. CCA seeks maximal correlations between linear combinations of two or more sources of data, e.g. brain connectivity and behavioural data.

To be able to apply CCA directly on this dataset, where the number of variables in both connectivity and behavioural data is greater than the number of subjects, we first reduced the dimensionality of the data using principal component analysis (PCA). In summary, the pre-processed brain and behavioural data matrices, $\mathbf{X} \in \mathbb{R}^{n \times p}$ and $\mathbf{Y} \in \mathbb{R}^{n \times q}$, respectively, were first decomposed into $\mathbf{X}_d \in \mathbb{R}^{n \times d}$ and $\mathbf{Y}_d \in \mathbb{R}^{n \times d}$, where d is the number of PCA components. We chose the optimal number of components based on the permutation approach described below, using 9 different PCA dimensionalities ($d = 5, 10, 25, 50, 75, 100, 125, 200$).

After the dimensionality reduction step, $\mathbf{X}_d$ and $\mathbf{Y}_d$ are fed into CCA, which outputs d components (modes of covariations, or in short, 'CCA modes'). Each CCA mode includes a pair of canonical basis vectors $\mathbf{u} \in \mathbb{R}^{p \times 1}$ and $\mathbf{v} \in \mathbb{R}^{q \times 1}$ indicating the direction of maximum brain-behaviour correlation; as well as a pair of canonical variates $\mathbf{P}_X \in \mathbb{R}^{n \times 1}$ and $\mathbf{P}_Y \in \mathbb{R}^{n \times 1}$ obtained by projecting $\mathbf{X}_d$ and $\mathbf{Y}_d$ onto their corresponding canonical basis vectors. On these canonical variates, subjects are represented by brain and behaviour scores, which describe the subject specific loadings in each CCA mode. Significance for the CCA modes was assessed based on the permutation framework described below. To find brain connections and behavioural variables most strongly associated with the CCA modes, we correlated the pre-processed brain and behavioural variables with the canonical variates $\mathbf{P}_X$ and $\mathbf{P}_Y$, respectively. Then for each significant CCA mode, we have a vector of correlations for the top brain connectivity variables (Figs 3 and 4) and a vector of correlations for the top behavioural variables (Fig. 2).

Permutation framework. We used a permutation-based approach (Supplementary Fig. S1) for choosing the optimal number of PCA components and estimating the Family Wise Error (FWE) corrected *p*-value on the canonical variates. The algorithm proceeds as follows:

1. For a given number of PCA components (e.g. $d = 5$), the reduction step is performed on **X** and **Y** resulting in the reduced data matrices $\mathbf{X}_{n \times 5}$ and $\mathbf{Y}_{n \times 5}$. Then, CCA is applied to this lower-dimensional space to compute the vector with "true" canonical correlations $\mathbf{q}_{5 \times 1}$, with one value per CCA component.

2. The rows of data matrix $\mathbf{Y}_{n \times 5}$ are permuted to obtain $\mathbf{Y}^*_{n \times 5}$ and then CCA is again computed using $\mathbf{X}_{n \times 5}$ and $\mathbf{Y}^*_{n \times 5}$ to obtain the corresponding vector of canonical correlations $\mathbf{q}^*_{5 \times 1}$. This procedure is repeated 10000 times, resulting in a matrix of canonical correlations $\mathbf{Q}^*_{5 \times 10000}$.

3. For each row $i = 1, \ldots, 5$ of $\mathbf{Q}^*_{5 \times 10000}$, a p-value is computed by assessing the number of times the permuted canonical correlations in row i are equal or greater than the first "true" canonical correlation (as canonical correlations are ordered), i.e. the first element of $\mathbf{q}_{5 \times 1}$ (equivalent to a maximum statistics approach). At the end of this procedure, a vector of p-values $\mathbf{p}_{5 \times 1}$ is obtained, one per CCA component. This allows one to estimate the number of significant CCA components accounting for FWE (i.e. any CCA component with $p_{\text{FWE}} < 0.05$ is considered statistically significant). The p-value of the first CCA component (i.e. the first element of $\mathbf{p}_{5 \times 1}$) is used to assess the optimal number of PCA components.

4. Steps 1–3 are repeated for the other PCA dimensionalities ($d = 10, 25, 50, 75, 100, 125, 150, 200$).

The obtained p-values for each PCA dimensionality were corrected for multiple comparisons using *Bonferroni* correction (i.e. $\alpha = 0.05/9 = 0.0056$), which means that only the PCA sets with $p_{corr} \leq 0.0056$ are considered statistically significant. The optimal number of PCA components is chosen based on the lowest p_{corr}.

Data Availability

The datasets analysed during the current study are not publicly available because this was not foreseen when the ethical approval for the study was obtained, but are available from the corresponding author on reasonable request.

References

1. Kessler, R. C. *et al.* Age of onset of mental disorders: a review of recent literature. *Curr. Opin. Psychiatry* **20**, 359–64 (2007).
2. Paus, T., Keshavan, M. & Giedd, J. N. Why do many psychiatric disorders emerge during adolescence? *Nat. Rev. Neurosci.* **9**, 947–957 (2008).
3. Lipari, R. N., Hedden, S. L. & Hughes, A. Substance Use and Mental Health Estimates from the 2013 National Survey on Drug Use and Health: Overview of Findings. In *The CBHSQ Report* (Rockville (MD): Substance Abuse and Mental Health Services Administration (US), 2014).
4. Smith, S. M. *et al.* Functional connectomics from resting-state fMRI. *Trends Cogn. Sci.* **17**, 666–682 (2013).
5. Bassett, D. S. & Bullmore, E. T. Human brain networks in health and disease. *Curr. Opin. Neurol.* **22**, 340–347 (2009).
6. McIntosh, A. R. & Mišić, B. Multivariate statistical analyses for neuroimaging data. *Annu. Rev. Psychol.* **64**, 499–525 (2013).
7. Hotelling, H. Relations between two sets of variates. *Biometrika* **28**, 321–377 (1936).
8. Smith, S. M. *et al.* A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience* **18**, 1565–1567 (2015).
9. Drysdale, A. T. *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine* **23**, 28–38 (2016).
10. Xia, C. H. *et al.* Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat. Commun.* **9**, 3003 (2018).
11. Moser, D. A. *et al.* Multivariate associations among behavioral, clinical, and multimodal imaging phenotypes in patients with psychosis. *JAMA Psychiatry* **75**, 386–395 (2018).
12. Kiddle, B. *et al.* Cohort Profile: The NSPN 2400 Cohort: a developmental sample supporting the Wellcome Trust NeuroScience in Psychiatry Network. *Int. J. Epidemiol* **47**, 18–19g (2018).
13. Fischl, B. *et al.* Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
14. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
15. Witten, D. M. & Tibshirani, R. J. Extensions of Sparse Canonical Correlation Analysis with Applications to Genomic Data. *Stat. Appl. Genet. Mol. Biol* **8**, 29 (2009).
16. Monteiro, J. M., Rao, A., Shawe-Taylor, J. & Mourão-Miranda, J. A multiple hold-out framework for Sparse Partial Least Squares. *J. Neurosci. Methods* **271**, 182–194 (2016).
17. Buckholtz, J. W. & Meyer-Lindenberg, A. Psychopathology and the human connectome: toward a transdiagnostic model of risk for mental illness. *Neuron* **74**, 990–1004 (2012).
18. Menon, V. Large-scale brain networks and psychopathology: A unifying triple network model. *Trends Cogn. Sci.* **15**, 483–506 (2011).
19. Casey, B. J., Jones, R. M. & Hare, T. A. The adolescent brain. *Ann. N. Y. Acad. Sci* **1124**, 111–126 (2008).
20. Casey, B. J. Beyond simple models of self-control to circuit-based accounts of adolescent behavior. *Annu. Rev. Psychol.* **66**, 295–319 (2015).
21. Chekroud, A. M. *et al.* Reevaluating the Efficacy and Predictability of Antidepressant Treatments. *JAMA Psychiatry* **74**, 370 (2017).
22. Labonté, B. *et al.* Sex-specific transcriptional signatures in human depression. *Nat. Med.* **23**, 1102–1111 (2017).
23. Lee, F. S. *et al.* Adolescent mental health–Opportunity and obligation. *Science* **346**, 547–549 (2014).
24. Alterman, A. I. *Substance Abuse and Psychopathology.* (Springer US, 1985).
25. Brent, D. A. Comorbidity of substance abuse and other psychiatric disorders in adolescents. *Am. J. Psychiatry* **146**, 1131–1141 (1989).
26. Castellanos-Ryan, N. & Conrod, P. Personality and substance misuse: evidence for a four-factor model of vulnerability. In *Drug Abuse and Addiction in Medical Illness: Causes, Consequences and Treatment* (eds Verster, J. C., Brady, K., Galanter, M. & Conrod, P.) 1–573 (Springer New York, 2012).
27. Carver, C. S., Sutton, S. K. & Scheier, M. F. Action, Emotion, and Personality: Emerging Conceptual Integration. *Personal. Soc. Psychol. Bull* **26**, 741–751 (2000).
28. Casey, B. J., Galván, A. & Somerville, L. H. Beyond simple models of adolescence to an integrated circuit-based account: A commentary. *Dev. Cogn. Neurosci* **17**, 128–130 (2016).
29. Spear, L. P. Effects of adolescent alcohol consumption on the brain and behaviour. *Nat. Rev. Neurosci.* **19**, 197–214 (2018).
30. Berman, M. G. *et al.* Neural and behavioral effects of interference resolution in depression and rumination. *Cogn. Affect. Behav. Neurosci.* **11**, 85–96 (2011).
31. Cooney, R. E., Joormann, J., Eugène, F., Dennis, E. L. & Gotlib, I. H. Neural correlates of rumination in depression. *Cogn. Affect. Behav. Neurosci.* **10**, 470–478 (2010).
32. Admon, R. & Pizzagalli, D. A. Corticostriatal pathways contribute to the natural time course of positive mood. *Nat. Commun.* **6**, 1–11 (2015).

9

33. Smith, S. M. *et al*. Network modelling methods for FMRI. *Neuroimage* **54**, 875–91 (2011).
34. Volkow, N. D. *et al*. The conception of the ABCD study: From substance use to a broad NIH collaboration. *Dev. Cogn. Neurosci.* **32**, 4–7 (2017).
35. Insel, T. *et al*. Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry* **167**, 748–751 (2010).
36. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142–2154 (2012).
37. Weiskopf, N. *et al*. Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: A multi-center validation. *Front. Neurosci.* **7**, 95 (2013).
38. Helms, G., Dathe, H., Kallenberg, K. & Dechent, P. High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and T1 relaxation obtained from 3D FLASH MRI. *Magn. Reson. Med.* **60**, 1396–1407 (2008).
39. Helms, G., Dathe, H. & Dechent, P. Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation. *Magn. Reson. Med.* **59**, 667–672 (2008).
40. Weiskopf, N. *et al*. Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT). *Neuroimage* **54**, 2116–2124 (2011).
41. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).
42. Kundu, P. *et al*. Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proc. Natl. Acad. Sci.* **110**, 16187–16192 (2013).
43. Kundu, P. *et al*. Robust resting state fMRI processing for studies on typical brain development based on multi-echo EPI acquisition. *Brain Imaging Behav* **9**, 56–73 (2015).
44. Thomas Yeo, B. T. *et al*. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

Acknowledgements

Author Contributions

A.M., F.S.F., M.J.R. and J.M.M. designed the study. M.M., E.T.B., P.F., I.M.G., P.B.J., R.D. and NSPN consortium collected the data. A.M., M.J.R., M.M., G.Z., J.M., R.R.G., P.E.V., M.G.K., F.V. and M.V. contributed to MRI and/or clinical data pre-processing. A.M., F.S.F. and M.J.R. implemented the modelling analysis; J.M.M. supervised the modelling analysis. A.M., F.S.F., M.J.R., M.M., G.Z., L.P., R.A.A. and J.M.M. discussed the interpretation of the results. A.M., F.S.F. and M.J.R. prepared figures and tables. All authors contributed to the manuscript. All authors approved the final version of the manuscript.

Additional Information

Competing Interests: E.T.B. is employed half-time by the University of Cambridge and half-time by GlaxoSmithKline; he holds stock in GlaxoSmithKline.

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

Consortia
NSPN Consortium

Tobias Hauser[2,3], Sharon Neufeld[7], Michelle St Clair[7], Kirstie Whitaker[7,9], Becky Inkster[7], Gita Prabhu[2,3], Cinly Ooi[7], Umar Toseeb[7], Barry Widmer[7], Junaid Bhatti[7], Laura Villis[7], Ayesha Alrumaithi[7], Sarah Birt[7], Aislinn Bowler[3], Kalia Cleridou[3], Hina Dadabhoy[3], Emma Davies[7], Ashlyn Firkins[7], Sian Granville[3], Elizabeth Harding[3], Alexandra Hopkins[2,3], Daniel Isaacs[3], Janchai King[3], Danae Kokorikou[3,11], Christina Maurice[7], Cleo McIntosh[7], Jessica Memarzia[7], Harriet Mills[3], Ciara O'Donnell[7], Sara Pantaleone[3], Jenny Scott[7], Pasco Fearon[11], John Suckling[7], Anne-Laura van Harmelen[7] & Rogier Kievit[2,12]

[12]Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, United Kingdom.