

# Multimodal population brain imaging in the UK Biobank prospective epidemiological study

Karla L Miller<sup>1</sup>, Fidel Alfaró-Almagro<sup>1</sup>, Neal K Bangerter<sup>2</sup>, David L Thomas<sup>3</sup>, Essa Yacoub<sup>4</sup>, Junqian Xu<sup>5</sup>, Andreas J Bartsch<sup>6</sup>, Saad Jbabdi<sup>1</sup>, Stamatios N Sotiropoulos<sup>1</sup>, Jesper L R Andersson<sup>1</sup>, Ludovica Griffanti<sup>1</sup>, Gwenaëlle Douaud<sup>1</sup>, Thomas W Okell<sup>1</sup>, Peter Weale<sup>7</sup>, Iulius Dragonu<sup>7</sup>, Steve Garratt<sup>8</sup>, Sarah Hudson<sup>8</sup>, Rory Collins<sup>8,9</sup>, Mark Jenkinson<sup>1</sup>, Paul M Matthews<sup>10</sup> & Stephen M Smith<sup>1</sup>

Medical imaging has enormous potential for early disease prediction, but is impeded by the difficulty and expense of acquiring data sets before symptom onset. UK Biobank aims to address this problem directly by acquiring high-quality, consistently acquired imaging data from 100,000 predominantly healthy participants, with health outcomes being tracked over the coming decades. The brain imaging includes structural, diffusion and functional modalities. Along with body and cardiac imaging, genetics, lifestyle measures, biological phenotyping and health records, this imaging is expected to enable discovery of imaging markers of a broad range of diseases at their earliest stages, as well as provide unique insight into disease mechanisms. We describe UK Biobank brain imaging and present results derived from the first 5,000 participants' data release. Although this covers just 5% of the ultimate cohort, it has already yielded a rich range of associations between brain imaging and other measures collected by UK Biobank.

The primary clinical role of brain imaging to date has been in diagnosis and monitoring of disease progression, rather than providing predictive markers for preventative stratification or early therapeutic intervention. The predominant strategy for finding image-based markers of neurological and psychiatric disease has been to identify patients early in the diagnostic process to maximize statistical power in a small cohort (tens to hundreds of subjects). A key factor motivating the use of small, clinically defined cohorts is the expense, time and specialized hardware associated with imaging. This approach has been effective in providing markers of disease progression, but identifying imaging markers of early disease requires measurements at the pre-symptomatic stage. Image-based measures of brain structure and function may evolve in a complex way throughout aging and the progression of neuropathology. Thus, markers with utility in monitoring disease progression post-diagnostically may not manifest pre-symptomatically, and conversely the most sensitive early predictors of disease may have plateaued by the time existing diagnoses become accurate.

Nevertheless, when known risk factors have enabled risk-stratified cohorts, imaging has been able to predict disease before symptom presentation. For example, magnetic resonance imaging (MRI) has demonstrated altered brain activity that is associated with the APOE genotype decades in advance of symptoms associated with Alzheimer's disease<sup>1</sup>, and conversion from mild cognitive impairment to Alzheimer's has been predicted<sup>2</sup>. These studies suggest that the primary

obstacle to identifying early imaging markers is obtaining data in pre-symptomatic cohorts drawn from the general population.

Alternatively, pre-symptomatic cohorts can be assembled using a prospective approach, in which a large number of healthy participants are intensively phenotyped (including imaging) and subsequently monitored for long-term health outcomes. Although this approach is expensive, it is also efficient, as it captures early biomarkers and risk factors for a broad range of diseases. It further becomes possible to discover unexpected interactions between risk factors (such as lifestyle and genetics). To date, the largest brain imaging studies have gathered data on a few thousand subjects. Although this approach has identified associations between imaging and highly prevalent diseases, existing cohorts are still too small to produce sufficient incidence of many diseases if participants are recruited without identifying risk factors.

UK Biobank is a prospective epidemiological resource gathering extensive questionnaires, physical and cognitive measures, and biological samples (including genotyping) in a cohort of 500,000 participants<sup>3</sup>. Participants consent to allow access to their full health records from the UK National Health Service, enabling researchers to relate phenotypic measures to long-term health outcomes. This is particularly powerful as a result of the combination of the number of subjects and the breadth of linked data. Participants were 40–69 years of age at baseline recruitment; this aims to balance the goals of characterizing subjects before disease onset against the delay before

<sup>1</sup>Oxford Centre for Functional MRI of the Brain (FMRIB), University of Oxford, Oxford, UK. <sup>2</sup>Department of Electrical Engineering, Brigham Young University, Provo, Utah, USA. <sup>3</sup>Institute of Neurology, University College London, London, UK. <sup>4</sup>Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, Minnesota, USA. <sup>5</sup>Icahn School of Medicine at Mount Sinai, New York, New York, USA. <sup>6</sup>Department of Neuroradiology, University of Heidelberg, Heidelberg, Germany. <sup>7</sup>Siemens Healthcare UK, Frimley, UK. <sup>8</sup>UK Biobank, Stockport, UK. <sup>9</sup>Nuffield Department of Population Health, University of Oxford, Oxford, UK. <sup>10</sup>Division of Brain Sciences, Department of Medicine, Imperial College London, London, UK. Correspondence should be addressed to K.L.M. (karla@fmrib.ox.ac.uk).

Received 16 March; accepted 25 August; published online 19 September 2016; doi:10.1038/nn.4393

health outcomes accumulate. The cohort is particularly appropriate for the study of age-associated pathology. All data from UK Biobank are available to researchers world-wide on application, with no preferential access for the scientists leading the study. An imaging extension to the existing UK Biobank study was funded in 2016 to scan 100,000 subjects from the existing cohort, aiming to complete by 2022. Imaging includes MRI of the brain, heart and body, low-dose X-ray bone and joint scans, and ultrasound of the carotid arteries. Identification of disease risk factors should increase over time with emerging clinical outcomes. For example, in the imaged cohort, 1,800 participants are expected to develop Alzheimer's disease by 2022, rising to 6,000 by 2027 (diabetes: 8,000 rising to 14,000; stroke: 1,800 to 4,000; Parkinson's: 1,200 to 2,800)<sup>4</sup>.

Here we present example analytic approaches and studies that will be enabled by UK Biobank. The identification of new imaging biomarkers of disease risk could support diagnosis, development of therapeutics and assessment of interventions. The multi-modal, multi-organ imaging enables the study of interactions between organ systems, for example, between cardiovascular health and dementia. The breadth of imaging makes this data set valuable for multi-systemic syndromes such as frailty, accelerated aging characterized by general loss of reserves and poor tolerance to stressors, which indicates increased risk for a range of conditions including dementia<sup>5</sup>. This kind of resource can also evince hypotheses regarding causal mechanisms of disease that could be tested in follow-up interventional studies. Examples include modifiable risk factors, such as the association of obesity with later life cognitive dysfunction<sup>6</sup>, and the ability to study complex interactions of risk factors with lifestyle, environment and genetics. Finally, UK Biobank will enable validation and extension of associations identified by smaller-scale studies, including the testing of hypotheses that combine results from multiple previous studies.

RESULTS

Design rationale and initial imaging phase

The imaging study was designed to achieve the target of 100,000 subjects, each scanned once, over 5–6 years at three dedicated, identical centers operating 7 days per week, each scanning 18 subjects per day (ref. 7). This requirement places tight timing constraints, corresponding to one subject imaged every 36 min (Online Methods). The first imaging center was built to establish feasibility and scanned 10,000 subjects over a 2-year ramp-up period. Two further identical centers are being commissioned, with the three centers being strategically positioned at population hubs: Manchester, Reading and Newcastle.

To capture imaging phenotypes relevant to the widest possible range of diseases and hypotheses, our protocol must deliver data with the broadest predictive power for neuropathology and mental health. We therefore included modalities that drive estimates of anatomical and neuropathological structure (structural MRI), brain activity (functional MRI, or fMRI), and local tissue microstructure (diffusion MRI, dMRI). The resulting imaging protocol (Supplementary Table 1) included three structural modalities, T1-weighted, T2-weighted and susceptibility-weighted MRI (referred to here as T1, T2 and swMRI); dMRI; and both task and resting-state fMRI (tfMRI and rfMRI). Recent advances in MRI acquisition technology<sup>8</sup> enabled high spatial resolution dMRI and fMRI with high angular and temporal resolution, respectively, despite strict time constraints. For example, the protocol acquires dMRI data with 100 diffusion-encoding directions over two shells in just 7 min, enabling advanced model fitting of microstructural parameters that would not have been possible

under these time constraints with previous generation technology. Following optimization of acquisition protocols, streamlining of participant preparation and minimization of scanner dead time (Online Methods), UK Biobank was able to incorporate six neuroimaging modalities in just 36 min.

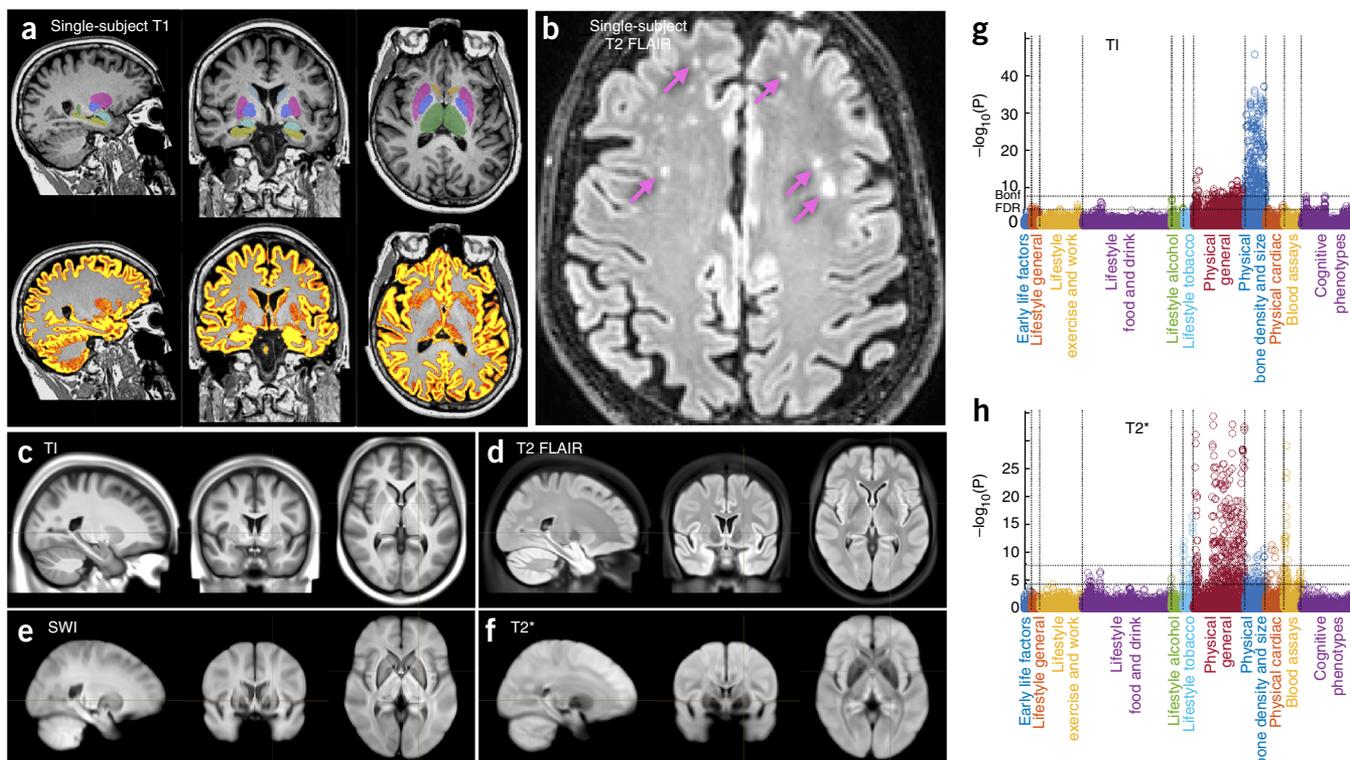
Unlike most of the measurements included in the original UK Biobank resource (for example, alcohol consumption and cognitive test scores), raw imaging data is not a directly useful source of information. In addition to requiring image processing to remove artifacts and align images across modalities and individuals, most useful image phenotypes are derived through complex calculations that combine many voxels and/or images. A fully automated processing pipeline was developed that produces both processed images as well as image-derived phenotypes (IDPs); there are currently 2,501 distinct individual measures of brain structure and function. Example IDPs include the volume of specific brain structures, the strength of connectivity between pairs of brain regions and the estimated dispersion of fibers in a given white-matter tract. IDPs are intended to be useful for non-imaging experts; however, understanding of the confounds and pitfalls of imaging is required to draw appropriate conclusions.

Here we present results from the first data release (<http://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=100>), which includes outputs from the processing pipeline for 5,285 subjects scanned in 2014–2015. As determined by the processing pipeline, 98% of participants' data sets resulted in a usable T1, which is crucial for deriving usable information from the other modalities. Of these, data for the other brain imaging modalities were suitable for processing in the following percentages of subjects: T2 = 97%, swMRI = 93%, dMRI = 95%, tfMRI = 92% and rfMRI = 95%. All modalities were acquired and were usable in 89% of subjects. Results from this data release are illustrated in Figures 1–4, including a multimodal atlas (separate population-average images for each of the modalities, all aligned to each other), available for download and online browsing at <http://www.fmrib.ox.ac.uk/ukbiobank>.

Imaging data, atlases and imaging-derived phenotypes

The three structural modalities (Fig. 1) provide information about different aspects of the brain's tissues, structures and neuropathologies. Data quality at the single-subject level is illustrated in Figure 1a,b. The group-averaged images produced for each modality are included in the initial data release as high-quality atlases (Fig. 1c–f), depicting strong tissue contrast and excellent fidelity of alignment across subjects. The T1 modality (Fig. 1a,c) is the most informative about the basic structure of the brain, including the depiction of the main tissue types (gray and white matter) and gross structure of the brain (main anatomical landmarks). From the T1 data, we derived 25 volumetric IDPs: total tissue volumes (gray, white and ventricular cerebrospinal fluid) and the volumes of subcortical gray matter structures such as thalamus, caudate, putamen, pallidum, hippocampus and amygdala. The T1 data and T1-derived IDPs provide sensitive markers of atrophy (tissue loss), which can be both global (for example, thinning of the cortex in aging)<sup>9</sup> and local (for example, reduction of hippocampal volume in Alzheimer's disease)<sup>10</sup>. The T2 data (Fig. 1b,d) is a fluid-attenuated inversion recovery (FLAIR) acquisition that also depicts basic anatomy, but is valuable primarily for detection of focal 'hyperintensities' (that is, high-signal regions) in white matter. T2 hyperintensities represent white matter lesions that have been associated with a broad range of neuropathological conditions<sup>11</sup> (for example, small vessel ischemic disease), and occur with increasing incidence in aging populations without (or potentially before) manifestation of

© 2016 Nature America, Inc. All rights reserved. npg



**Figure 1** Data from the three structural imaging modalities in UK Biobank brain imaging. (a) Single-subject T1-weighted structural image with minimal pre-processing: removal of intensity inhomogeneity, lower neck areas cropped and the face blanked to protect anonymity. Color overlays show automated modeling of several subcortical structures (above) and segmentation of gray matter (below). (b) Single-subject T2-weighted FLAIR image with the same minimal pre-processing showing hyperintense lesions in the white matter (arrows). (c) Group-average ( $n \approx 4,500$ ) T1 atlas; all subjects' data were aligned together (see Online Methods for processing details) and averaged, achieving high-quality alignment, with clear delineation of deep gray structures and good agreement of major sulcal folding patterns despite wide variation in these features across subjects. (d) Group-average T2 FLAIR atlas. (e) Group-average atlas derived from SWI processing of swMRI phase and magnitude images. (f) Group-average T2\* atlas, also derived from the swMRI data. (g) Manhattan plot (a layout common in genetic studies) relating all 25 IDPs from the T1 data to 1,100 non-brain-imaging variables extracted from the UK Biobank database, with the latter arranged into major variable groups along the x axis (with these groups separated by vertical dotted lines). For each of these 1,100 variables, the significance of the cross-subject univariate correlation with each of the IDPs is plotted vertically, in units of  $-\log_{10}(P_{\text{uncorrected}})$ . The dotted horizontal lines indicate thresholds corresponding to multiple comparison correction using FDR (lower line, corresponding to  $P_{\text{uncorrected}} = 3.8 \times 10^{-5}$ ) and Bonferroni correction (upper line,  $P_{\text{uncorrected}} = 1.8 \times 10^{-8}$ ) across the 2.8 million tests involving correlations of all modalities' IDPs against all 1,100 non-imaging measures. Effects such as age, sex and head size are regressed out of all data before computing the correlations. As an indication of the corresponding range of effect sizes, the maximum  $r^2$  (fractional variance of either variable explained by the other) is calculated, as well as the minimum  $r^2$  across all tests passing the Bonferroni correction. Here, the maximum  $r^2 = 0.045$  and the minimum  $r^2 = 0.0058$ . (h) Plot relating all 14 T2\* IDPs to 1,100 non-imaging variables. Maximum  $r^2 = 0.034$ , minimum  $r^2 = 0.0063$ . Marked Bonferroni and FDR multiple comparison threshold levels are presented as in g.

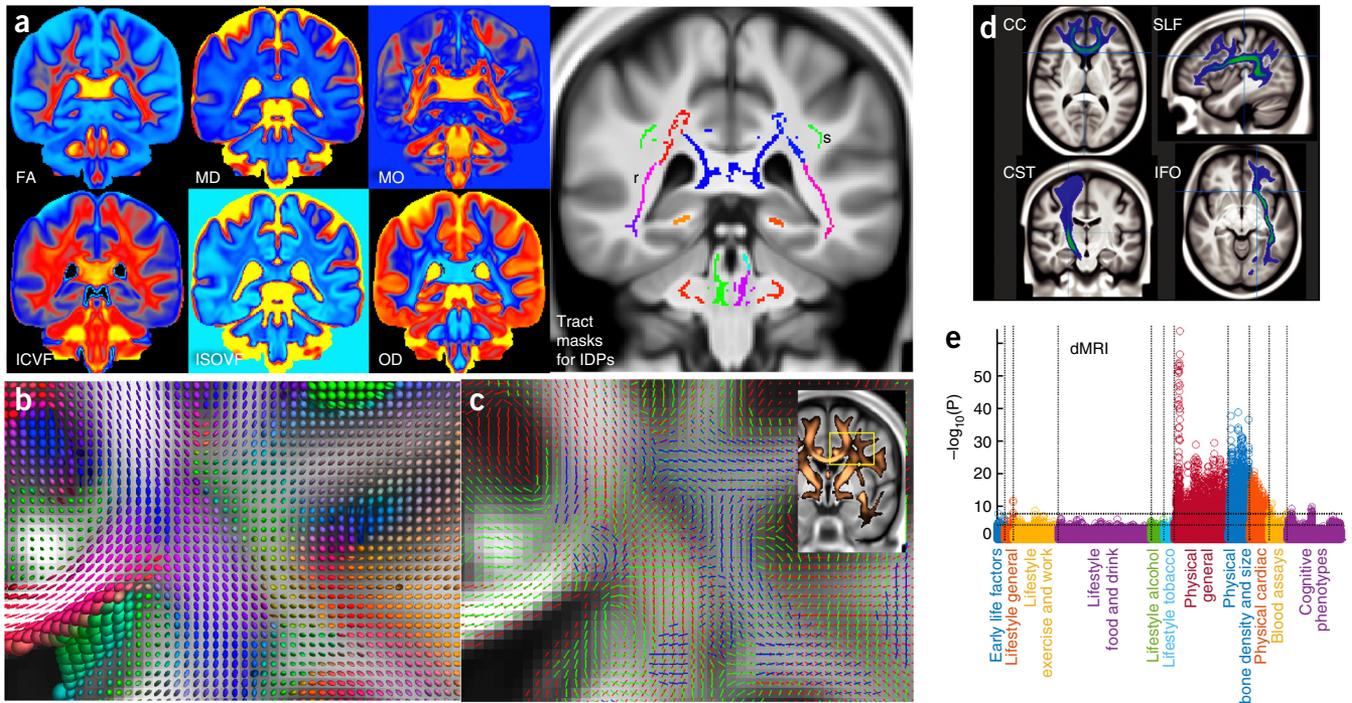
neurological symptoms. IDPs relating to the volume of these white matter lesions will be included in future data releases. swMRI is a flexible modality that can be processed in multiple ways, each sensitive to different clinically relevant properties. The first data release includes T2\* signal decay times and enhancement of venous vasculature using susceptibility-weighted image (SWI) filtering<sup>12</sup> (Fig. 1e,f). swMRI IDPs in the current data release are the median T2\* in each of 14 major subcortical gray matter structures, for example, reflecting increased iron deposition associated with neurodegeneration<sup>13</sup>.

dMRI (Fig. 2) reflects the random diffusion of water molecules, which is affected by the microscopic structure of tissue<sup>14</sup>, enabling us to infer the local density of cellular compartments in tissue (for example, neurites). In addition, axon bundles in white matter create an orientation dependence of water movement as a result of hindrance of diffusion perpendicular to the long axis of white matter tracts, an effect that can be tracked from voxel to voxel (tractography) to derive long-range white matter pathways. Three complementary diffusion models were fit to the signal in each voxel: (i) the diffusion tensor

model<sup>15</sup>, describing the signal phenomenologically as resulting from a three-dimensional ellipsoid profile of water displacement; (ii) the neurite orientation dispersion and density imaging (NODDI) model<sup>16</sup>, estimating microstructural properties (for example, neurites versus extracellular space); and (iii) the ball and sticks model<sup>17</sup>, estimating the orientation of multiple fiber populations in a voxel for tractography. We extracted 675 IDPs by averaging parameters estimated by the first two models over 75 different white matter tract regions on the basis of both subject-specific tractography<sup>18</sup> and population-average white matter masks<sup>19</sup>.

fMRI reflects neural activity indirectly, measuring dynamic changes in blood oxygenation and flow resulting from changes in neural metabolic demand<sup>20</sup>. The task deployed in tfMRI (Fig. 3) involved matching shapes and emotionally negative faces<sup>21</sup> and was chosen to engage a range of neural systems, from low-level sensory and motor to perceptual (for example, fusiform) and emotional (for example, amygdala) areas. The 16 tfMRI IDPs quantitate the strength of brain activity changes for specific aspects of the task in regions defined





**Figure 2** The diffusion MRI data in UK Biobank. **(a)** Group-average ( $n = 4,500$ ) atlases from six distinct dMRI modeling outputs, each sensitive to different aspects of the white matter microarchitecture. The atlases shown are: FA, MD (mean diffusivity) and MO (tensor mode); and ICVF (intra-cellular volume fraction), ISOVF (isotropic or free water volume fraction) and OD (orientation dispersion index) from the NODDI microstructural modeling. Also shown are several group-average white matter masks used to generate IDPs (for example, pink (r) are retrolenticular tracts in the internal capsules; upper green (s) are the superior longitudinal fasciculi). **(b)** Tensor ellipsoids depicting the group-averaged tensor fit at each voxel for the region shown in the inset in **c**. The shapes of the ellipsoids indicate the strength of water diffusion along three principal directions; long thin tensors indicate single dominant fiber bundles, whereas more spherical tensors (within white matter) generally imply regions of crossing fibers (seen more explicitly modeled in corresponding parts of **c**). **(c)** Group-averaged multiple fiber orientation atlases, showing up to three fiber bundles per voxel. Red shows the strongest fiber direction, green the second and blue the third. Each fiber bundle is only shown where the modeling estimates that population to have greater than 5% voxel occupancy. Inset shows the thresholded mean FA image (copper) overlaid on the T1, with the region shown in detail in **b** and **c**. **(d)** Four example group-average white matter tract atlases estimated by probabilistic tractography fed from the within-voxel fiber modeling: corpus callosum (genu), superior longitudinal fasciculus, corticospinal tract and inferior fronto-occipital fasciculus. **(e)** Plot relating all 675 dMRI IDPs (nine distinct dMRI modeling outputs from tensor and NODDI models  $\times$  75 tract masks) to 1,100 non-imaging variables (see **Fig. 1g** for details). Maximum  $r^2 = 0.057$ , minimum  $r^2$  (passing Bonferroni) = 0.0065. Dotted horizontal lines (multiple comparison thresholds) are described in **Figure 1g**.

© 2016 Nature America, Inc. All rights reserved.



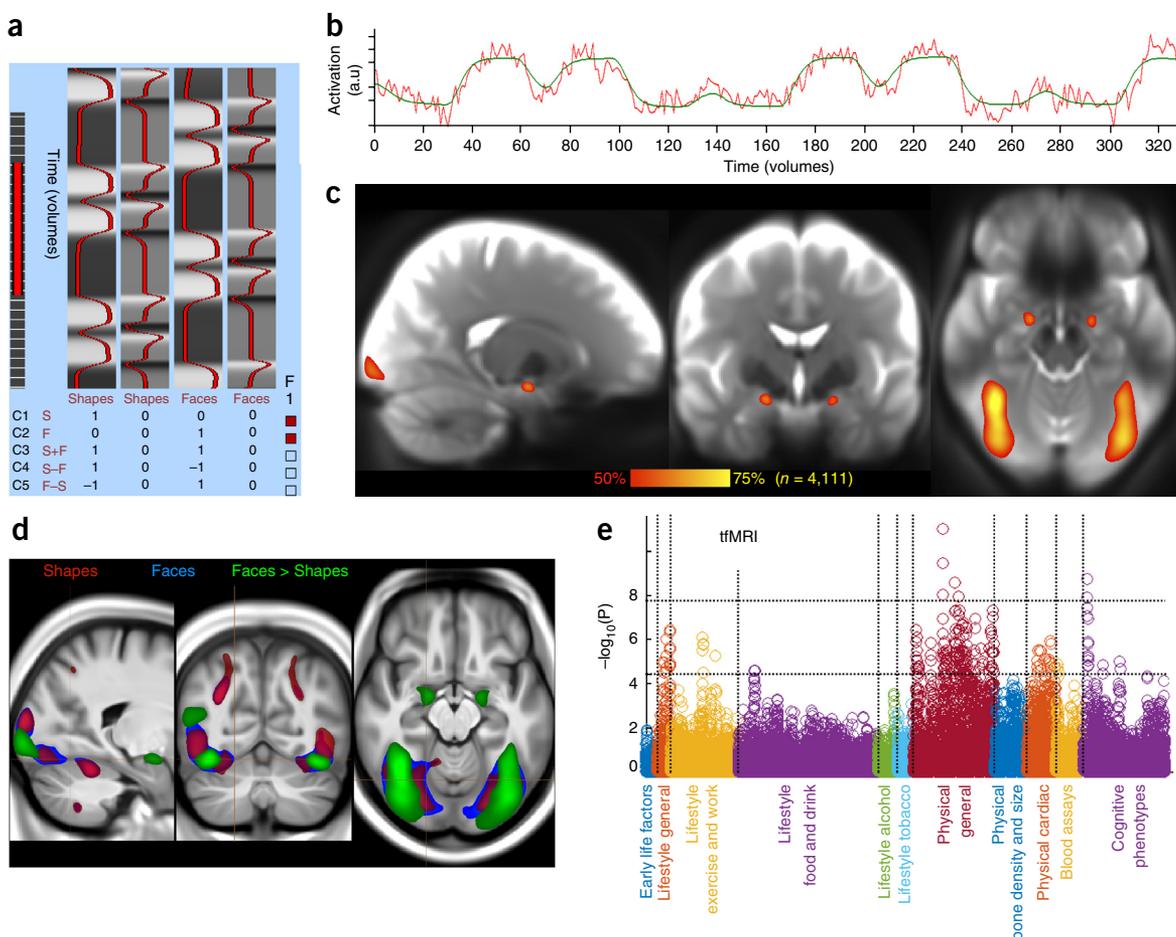
using the group-averaged activation maps shown across three task conditions. Resting-state fMRI (**Fig. 4**) identifies connected brain regions on the basis of common fluctuations in activity over time in the absence of an explicit task<sup>22</sup>. Sets of voxels that co-fluctuate most strongly correspond to brain regions, referred to as network ‘nodes’; different nodes may have weaker co-fluctuations, indicating a connection between them, or a network ‘edge’. The group analysis of the rfMRI data generated two atlases of these functional networks: a low-dimensional decomposition of the brain into 21 functional subdivisions and a higher dimensional parcellation into 55 subdivisions. IDPs represent edge connectivity strengths and node fluctuation amplitudes (**Fig. 4**).

**Voxel-wise associations with aging**

IDPs reduce raw data into a compact set of biologically meaningful measures, with current measures condensing ~2GB of raw data per subject into 2,501 IDPs, but such summary measures can lose valuable information. For example, once aligned to common coordinate systems, images can be analyzed for cross-subject variation at the voxel level to provide a more spatially detailed exploration than can be achieved via IDPs. However, this requires greater imaging expertise and computational resources, as well as often leading to lower statis-

tical power (as a result of the greatly increased number of multiple comparisons and the higher noise in voxel-wise measures compared with regional averages).

**Figure 5** presents voxel-wise correlations of age with several parameters modeled from the dMRI data (along the centers of the main white matter tracts), as well as normalized T2 FLAIR intensity in the white matter. Fractional anisotropy (FA), a sensitive, but nonspecific, marker of white matter integrity, predominantly demonstrated the established reduction of FA with aging (**Fig. 5a,g**). However, some voxels exhibited the opposite, with FA increasing with aging, which may reflect degradation of secondary fibers or reduced fiber dispersion<sup>23</sup>; notably, none of the FA-based IDPs exhibited this significant positive correlation, demonstrating that averaging across tracts can sacrifice richness of information. The tensor mode<sup>24</sup> (**Fig. 5b**), which primarily describes whether a voxel contains one versus multiple tracts, was even more sensitive, with highly significant positive correlations in certain association fiber areas and posterior corpus callosum, which is likely the same effect seen as FA increases<sup>23</sup>. We further observed an increase in free water with aging (**Fig. 5d**); the strongest increase, in the fornix, was likely a result of an increase of the fraction of cerebrospinal fluid in voxels spanning this thin tract as it atrophies. Finally, we calculated voxel-wise cross-subject correlation



**Figure 3** The task fMRI data in UK Biobank. **(a)** The task paradigm temporal model (time running vertically) depicting the periods of the two task types (shapes and faces); for more information on this paradigm view, see <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FEAT/UserGuide>. **(b)** Example fitted activation regression model versus time-series data (time running horizontally) for the voxel most strongly responding to the ‘faces > shapes’ contrast in a single subject ( $Z = 12.3$ ). **(c)** Percentage of subjects passing simple voxel-wise activation thresholding ( $Z > 1.96$ ) for the same contrast. Note the reliable focal activation in left and right amygdala. The underlying image is the group-averaged raw fMRI image. **(d)** Group-averaged activation for the three contrasts of most interest, overlaid on the group-average T1 atlas (fixed-effects group average,  $Z > 100$ , voxelwise  $P_{\text{corrected}} < 10^{-30}$ ). **(e)** Plot relating the 16 tfMRI IDPs to 1,100 non-imaging variables (see **Fig. 1g** for details). Maximum  $r^2 = 0.018$ , minimum  $r^2$  (passing Bonferroni) = 0.0062. Dotted horizontal lines (multiple comparison thresholds) are described in **Figure 1g**.

of age with T2 images. This analysis identified peri-ventricular areas, which are most susceptible to white matter hyperintensities known to be associated with aging (**Fig. 5e**).

A further example voxelwise analysis is shown in **Supplementary Figure 1**, in which we used the rfMRI data to investigate aging effects in the default-mode resting-state network<sup>25</sup>. This also provides a demonstration that group analyses do not degrade with increasingly large subject numbers (for example, as a result of alignment issues), as we used group sizes from 15 to 5,000. With increasing subject numbers, background noise was suppressed without increase in spatial blurring, and localized estimates of age-dependence stabilized, with statistical significance rising indefinitely.

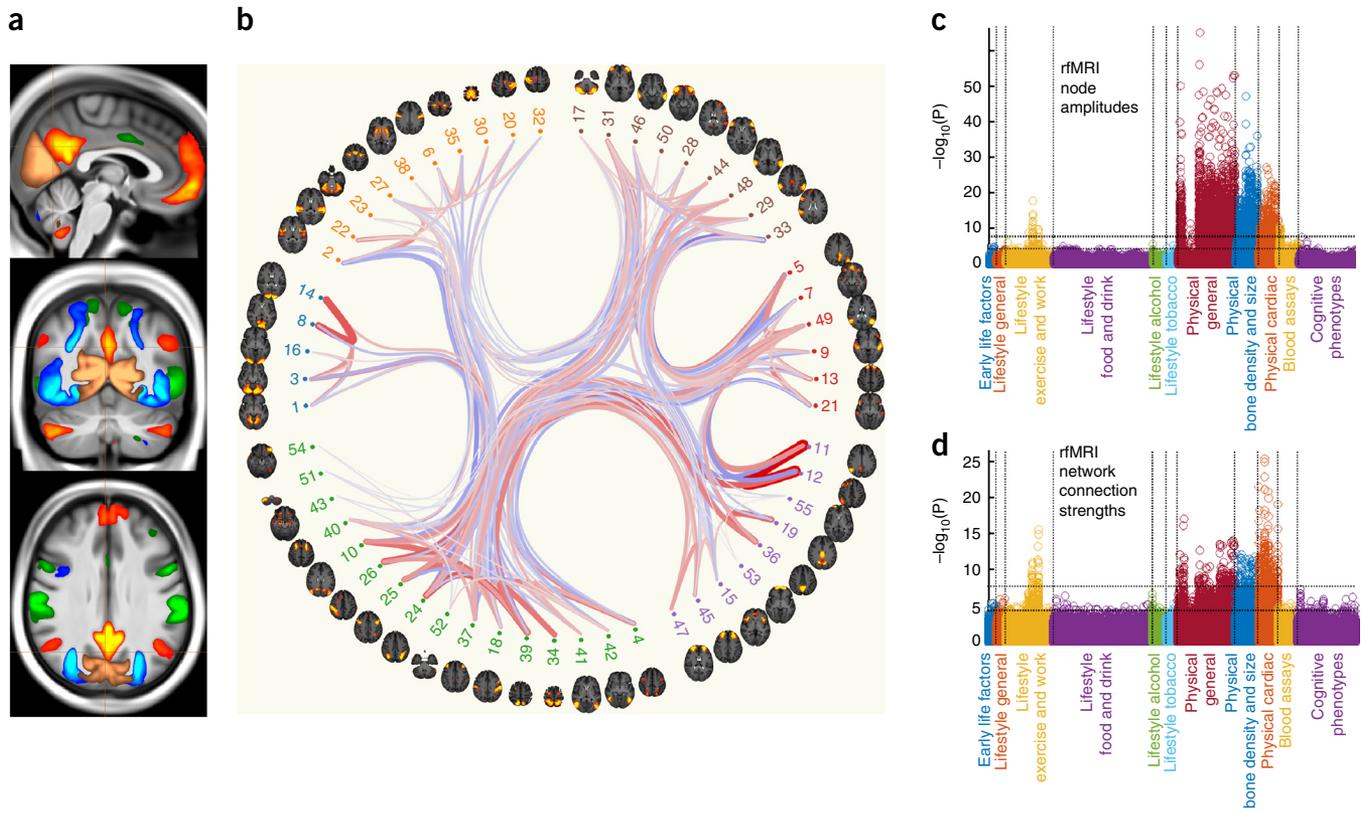
#### Pairwise associations between brain IDPs and other measures

We conducted simple univariate association analyses to illustrate the richness of relationships between IDPs and other available variables, as well as the statistical power afforded by ~5,000 subjects. We individually correlated all 2,501 brain IDPs with 1,100 other Biobank variables; the latter were broadly grouped into 11 categories (**Figs. 1–4** and **6**, and **Supplementary Fig. 2**). Even after false discovery rate (FDR)

multiple comparison correction for these 2.8 million correlations, 57 of the 66 combinations of brain modalities and non-brain-imaging categories showed significant associations. Some variable categories exhibited large numbers of associations with IDPs (for example, height and weight), whereas others (for example, cognitive measures and alcohol and tobacco intake) had more focused associations.

The above associations were estimated after adjusting all variables for age, sex, age-sex interaction, head motion and head size (de-confounding). Some factors can unambiguously be considered a confound to be removed (for example, head motion, which can corrupt IDPs, but also correlates with disease and aging<sup>26</sup>). For other factors (for example, age), the appropriateness of de-confounding depends on the question being asked and needs to be taken into consideration when interpreting associations (see Discussion). The relationship between the correlations estimated with versus without de-confounding (**Fig. 6d**) revealed that, in almost all cases, the strength of association was reduced by de-confounding, and in some cases was almost entirely removed (horizontal cloud around  $y = 0$ ).

We considered associations between cognitive tests and brain IDPs, including potential age interactions, in greater detail. Sex, head motion



**Figure 4** The resting-state fMRI data in UK Biobank. **(a)** Example group-average resting-state network (RSN) atlases from the low-dimensional group-average decomposition showing 4 of 21 estimated functional brain networks, including the default mode network (red-yellow), dorsal attention network (green), primary visual (copper) and higher level visual (dorsal and ventral streams, blue). The three slices shown are (top to bottom) sagittal, coronal and axial. **(b)** The 55 non-artifact components from a higher dimensional parcellation of the brain (axial views). These are shown as displayed by the connectome browser ([http://www.fmrib.ox.ac.uk/ukbiobank/netjs\\_d100](http://www.fmrib.ox.ac.uk/ukbiobank/netjs_d100)), which allows interactive investigation of individual connections in the group-averaged functional network modeling. The 55 brain regions (network nodes) are clustered into groups according to their average population connectivity, and the strongest individual connections are shown (positive in red, anticorrelations in blue). **(c)** Plot relating the 76 rfMRI 'node amplitude' IDPs to 1,100 non-imaging variables (see **Fig. 1g** for details). Maximum  $r^2 = 0.065$ , minimum  $r^2$  (passing Bonferroni) = 0.0059. **(d)** Plot relating the 1,695 rfMRI 'functional connectivity' IDPs to 1,100 non-imaging variables. Maximum  $r^2 = 0.032$ , minimum  $r^2 = 0.0059$ . Dotted horizontal lines (multiple comparison thresholds) in **c** and **d** are described in **Figure 1g**.

© 2016 Nature America, Inc. All rights reserved.



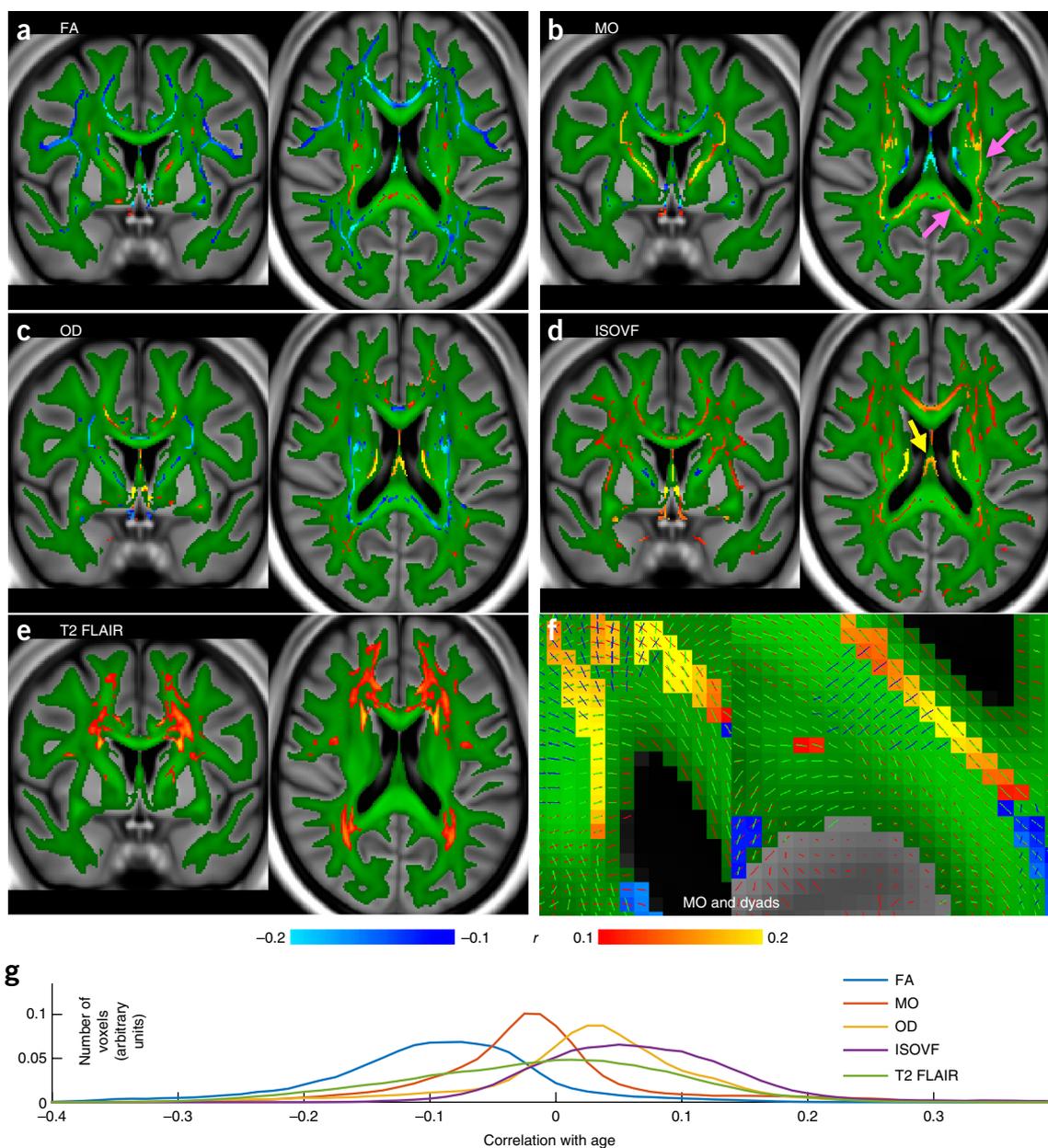
and head size were regressed out of all data before computing correlations (Online Methods). **Figure 6b** shows Bonferroni-significant ( $P_{\text{uncorrected}} < 1.8 \times 10^{-8}$ ) associations with brain IDPs, both with and without adjusting for age. The task-fMRI versus fluid intelligence associations were unchanged by adjusting for age, whereas all other cognition-IDP correlations were approximately doubled, being significantly stronger ( $P_{\text{corrected}} < 0.005$ ) without age adjustment.

In the symbol digit substitution test, participants replaced symbols with numbers using a substitution key. Strong IDP associations were found with two scores: the number of symbol digit matches made correctly and the number of symbol digit matches attempted in the time allowed (because subjects rarely made mistakes, these two scores are highly correlated,  $r = 0.97$ ). These scores correlated negatively with measures of water diffusivity in the corona radiata and superior thalamic radiation, and with FA in the posterior fornix (consistent with previous studies<sup>27</sup>, which may reflect variations in tract thickness<sup>28</sup>). Finally, there was a significant association with thalamus volume (right thalamic volume significant, left thalamic volume close to significance with  $r = 0.10$ ), consistent with previous findings<sup>29</sup>. These negative associations likely reflect lower cognitive performance with aging and pathology (increased diffusivity and atrophy).

In the reaction time test, subjects confirmed whether two abstract symbols matched as quickly as possible. The mean time to correctly identify matches was found to correlate inversely with left putamen volume (right putamen had similar correlation,  $r = -0.06$ , but was below significance). These negative associations are consistent with previous findings<sup>30</sup> and indicate that increased volume correlated positively with cognitive speed (and negatively with reaction time).

The fluid intelligence score reports how many numerical, logic and syntactic questions subjects were able to answer in 2 min. This was negatively correlated with the strength of gray matter activation in the simple shapes matching task in tfMRI, with no age interaction. The shapes matching task incurs low cognitive demand, and it is plausible that higher intelligence requires less neural activity for this task, a mechanism that has previously been ascribed to minimization of cognitive workload<sup>31</sup>.

All cognitive scores reported above involve processing speed as a significant factor, consistent with previous studies<sup>27</sup>. However, the observation that different test scores do not all correlate identically with each other or with the same brain IDPs suggests that there is not a single (speed-related) cognitive factor involved. The increases in association strengths when not controlling for age suggest that age-related cognitive decline is a major source of

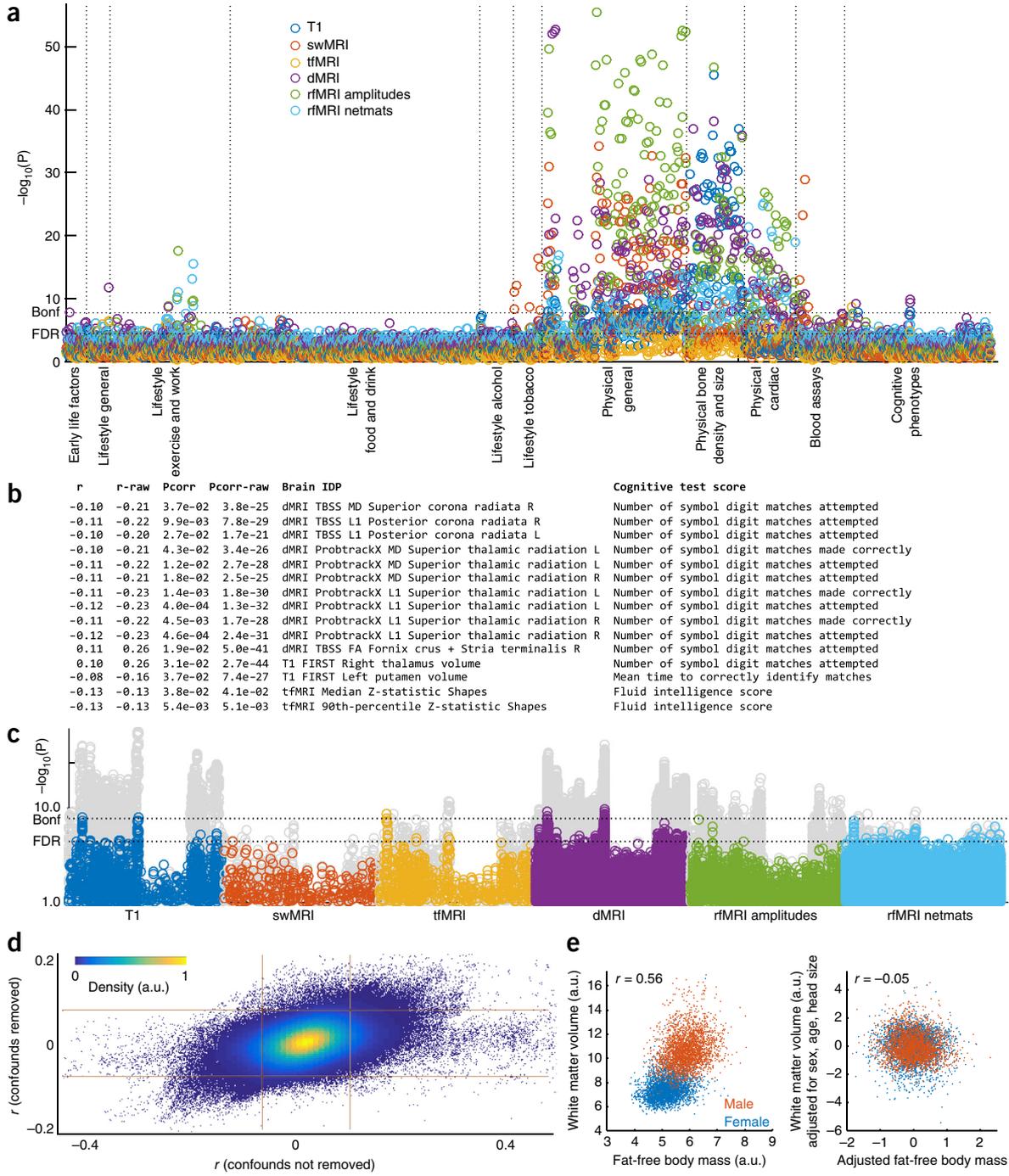


**Figure 5** Voxel-wise correlations of participants' age against several white matter measures from the dMRI and T2 FLAIR data. (a) Voxel-wise (cross-subject) correlation of FA versus age. Group-average FA in white matter is shown in green, overlaid onto the group-average T1. (b) Correlation of MO versus age, using the same color scheme. Nearby areas of MO increase are shown in greater detail in f, which also shows the distinct primary fiber directions. (c) Correlation of OD versus age, including a reduction in dispersion in posterior corpus callosum. (d) Correlation of ISOVF versus age, showing increases in freely diffusing water with age in a broad range of tracts. (e) Voxel-wise correlation of T2 FLAIR intensity showing increased intensity with aging in white matter. For a–e, blue and red-yellow show negative and positive Pearson correlation with age, respectively ( $P_{\text{corrected}} < 0.05$ , with Bonferroni correction across voxels resulting in significance at  $r = 0.1$  (dMRI  $n = 3,722$ ; T2 FLAIR  $n = 3,781$ ). (g) Histograms (across voxels) of the voxel-wise age correlation of the correlation maps shown above, with correlation value on the x axis. FA and MO largely decreased with age, whereas OD and ISOVF largely increased.

cross-subject variability for these IDP-cognition associations<sup>28</sup>. Plotting all IDP-cognitive associations (Fig. 6c) revealed that a large number of non-age-adjusted associations were stronger than the results after age adjustment; below, we show how interpretation of such results can be aided further through multivariate analyses. These age interactions provide an early indication that UK Biobank should provide cognitive biomarkers of clinical relevance as health outcomes accumulate.

### Multivariate associations: modes of population variation

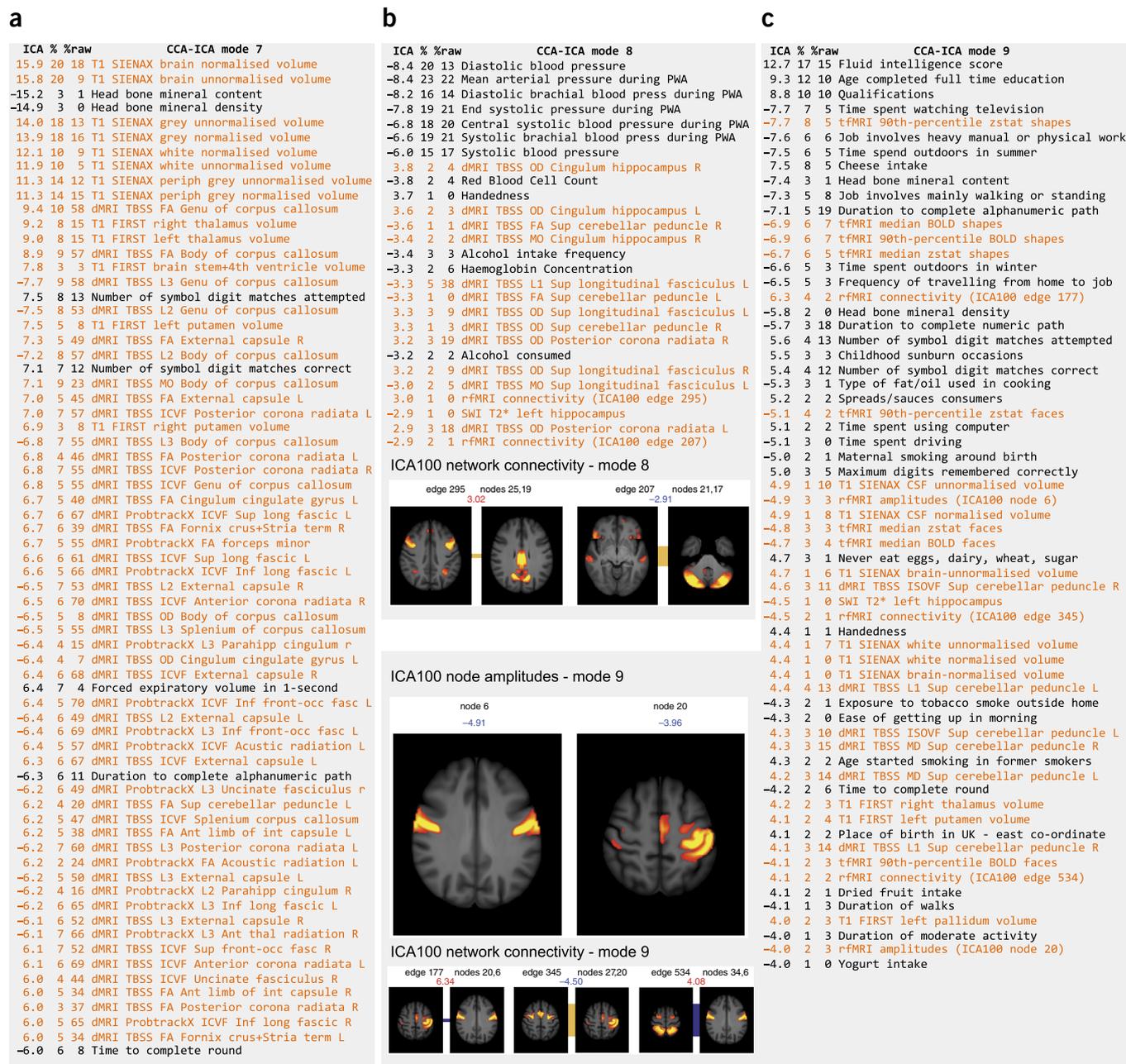
We conducted multivariate analyses using canonical correlation analysis (CCA)<sup>32</sup> combined with independent component analysis (ICA)<sup>33</sup>; Figs. 7 and 8, Supplementary Figs. 3–7, and Online Methods). This analysis identifies ‘modes’ of population covariation linking IDPs to non-imaging measures. Each mode consists of one linear combination of IDPs and a separate combination of non-imaging measures that have a highly similar variation across subjects. The strength of



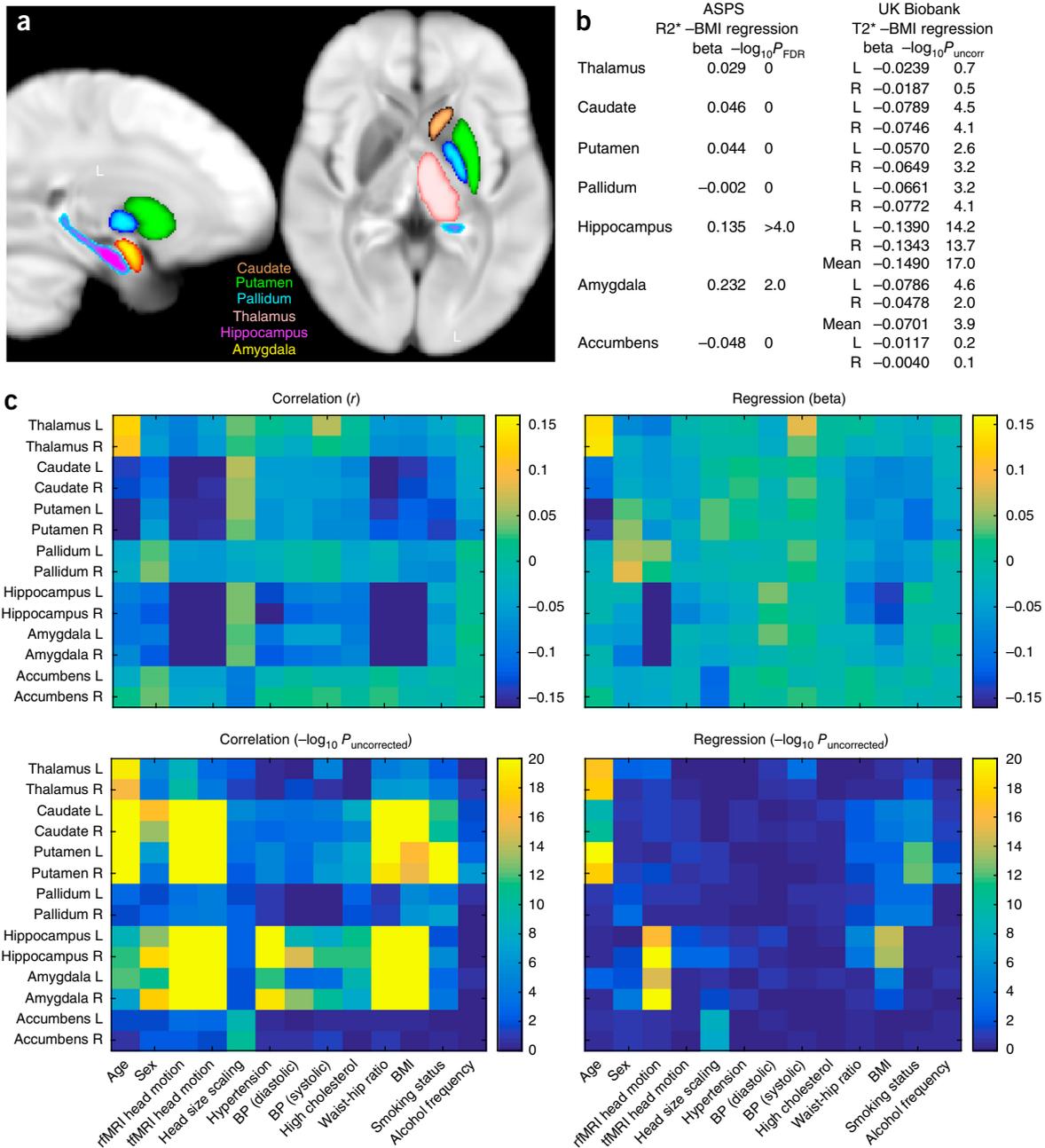
**Figure 6** Visualization of 2.8 million univariate cross-subject association tests between 2,501 IDPs and 1,100 other variables in the UK Biobank database. **(a)** Manhattan plot showing, for each of the 1,100 non-brain-imaging variables, the statistically strongest association of that variable with each distinct imaging sub-modality’s IDPs (that is, six results plotted for each x axis position, each with a color indicating a brain imaging modality; this plot differs from the other Manhattan plots, which show correlations with all IDPs). Whereas the Manhattan plots in **Figures 1–4** showed associations for each brain imaging modality separately, all associations are depicted in a single plot. **(b)** List of all IDP–cognitive score associations passing Bonferroni correction for multiple comparisons ( $P_{corrected} < 0.05$ ;  $P_{uncorrected} < 1.8 \times 10^{-8}$ ). The first column lists the age-adjusted correlation coefficient, and the second shows the unadjusted correlation, both being correlations between a specific brain IDP (fifth column) and a cognitive test score (sixth column). The UK Biobank cognitive tests carried out included fluid intelligence, prospective memory, reaction time (shape pairs matching), memorized pairs matching, trail making (symbol ordering), symbol digit substitution, and numeric memory. **(c)** IDP associations with the cognitive phenotype variables (the full set of 174 cognitive variables, repeated for each brain imaging modality). Shown behind, in gray, are the same associations without adjustment for age, with a large number of stronger associations. Dotted horizontal lines (multiple comparison thresholds) in **a** and **c** are described in **Figure 1g**. **(d)** Scatterplot showing the relationship between adjusted correlations and those obtained without first regressing out the confound variables (each point is a pairing of one IDP with one non-brain-imaging variable, 2.8 million points). The grid lines indicate Bonferroni-corrected significance level (as described in **Fig. 1**). **(e)** Example association between unadjusted white matter volume and fat-free body mass is high ( $r = 0.56$ ) when pooling across the sexes. After adjusting for several variables (including sex), the correlation falls almost to zero.

involvement of a variable in a given mode is dictated by the variable weight (Fig. 7). Multiple population modes may be identified, provided that they describe different (independent) cross-subject variation, meaning that the implied association between a given pair of variables can vary from mode to mode.

From the current UK Biobank release, we identified nine modes that were highly significant ( $P_{\text{corrected}} < 0.002$ , no further modes significant at  $P_{\text{corrected}} < 0.05$ ). Similar methodology using Human Connectome Project (HCP) data previously identified a single statistically significant mode of population covariation in 461 young healthy



**Figure 7** Details of three modes from the doubly-multivariate CCA-ICA analyses across all IDPs and non-brain-imaging variables. IDPs are listed in orange and non-brain-imaging variables in black. The lists show the variables most strongly associated with each mode; where multiple very similar (and highly correlated) non-imaging variables are found, only the most significant is listed here for brevity. The first column shows the weight (strength and sign) of a given variable in the ICA mode, the second shows the (cross-subject) percentage variance of the data explained by this mode, and the third column shows the percentage variance explained in the data without the confounds first regressed out. (a) Mode 7 links measures of bone density, brain structure/tissue volumes and cognitive tests. (b) Mode 8 links measures of blood pressure and alcohol intake to IDPs from the diffusion and functional connectivity data; two functional network connections strongly involved are displayed, with the population mean connection indicated by the bar connecting the two nodes forming the connection (red indicates positive mean correlation, blue negative, and the width of the bar indicates the connection strength). The group-ICA maps are thresholded at  $Z > 5$ , and the colored text is the ICA weight shown in the table list. (c) Mode 9 includes a wide range of imaging and non-imaging variables; as well as showing three strong functional network connections, we also show two functional nodes whose resting fluctuation amplitude is associated with this mode.



**Figure 8** Hypothesis-driven study of age, BMI and smoking associations with subcortical T2\*. (a) UK Biobank population-average map of T2\*, overlaid with the main subcortical structures being investigated. The T2\* IDPs reflect individuals' median T2\* values in these regions. The relatively low T2\* in putamen and pallidum likely reflects greater iron content. (b) BMI regression betas from multiple regressions of R2\* (from the ASPS study) and T2\* (from UK Biobank) against relevant covariates (see c). All variables are standardized so that beta values can be interpreted as (partial) correlation coefficients. R2\* significance is reported as FDR-corrected *P*. T2\* significance is reported as  $-\log_{10}P_{\text{uncorrected}}$  with the more conservative Bonferroni correction (for  $P_{\text{corrected}} = 0.05$ ) resulting in a threshold here of 3.6. (c) Full set of univariate and multiple regression betas and significance values for all brain regions tested and all model covariates. The regression results are much sparser, reflecting the higher associational specificity obtained by reporting unique variance explained.

adults<sup>8,34</sup>. Our ability to identify more modes than in the HCP data set could be a result of the tenfold increase in the number of subjects, the larger range of imaging modalities and non-imaging variables, and the older ages of subjects.

Although these modes are not guaranteed to reflect biological processes, in practice ICA often produces such interpretability<sup>35</sup>. Of the nine modes, some reflected dominant physical factors (for exam-

ple, body size or heart rate), whereas others linked rich subsets of non-imaging measures to IDPs. Modes 7–9 are displayed in **Figure 7**, and modes 1–6 are overviewed in **Supplementary Figures 3–7**. The relationships of these multivariate associations to potential confounds and variables of interest, including some clinical outcomes, are shown in **Supplementary Figure 8**. Modes 1, 2, 4, 5, 7 and 8 were strongly associated with aging, whereas 3, 6 and 9 were not.

Mode 7 primarily linked bone density measures and cognitive scores to brain structure and dMRI measures. There is extensive literature linking volume and diffusivity measures to cognition, but a relationship between these measures and bone density has not, to our knowledge, been reported. This link could reflect variations in physical properties of noninterest that are not fully accounted for by de-confounding. However, correlations between low bone density and accelerated cognitive decline have been reported<sup>36</sup>, including association of bone density with Alzheimer's disease<sup>37</sup>.

Mode 9 exhibited the most complex population pattern (Fig. 7c). The most strongly involved non-imaging measures were intelligence, education levels and occupational factors; in addition, some physical and dietary measures were involved that may reflect socio-economic status as a latent factor (for example, cheese intake or time spent outdoors in winter). Associated brain IDPs included task fMRI (with a negative weight, consistent with the sign of univariate associations), followed by a range of functional and structural IDPs.

There was some overlap between modes 7 and 9 in terms of cognitive tasks (for example, symbol digit matches), bone density and T1-based brain volumes. However, the fact that CCA-ICA separated modes 7 and 9 indicates that they constitute distinct biophysical patterns of variation across subjects; for example, mode 7 correlated with age, whereas mode 9 did not. The broader range of non-imaging measures involved in mode 9, and the ability to interpret many of them in terms of positive or negative life factors, is reminiscent of the single mode previously reported from HCP data<sup>8,34</sup>. That mode resembled the well-established observation of strong correlations in subject performance across a broad range of cognitive and behavioral tests (the general intelligence *g*-factor), but also included demographic and life factors. However, the correspondence between the mode 9 reported here and the previous HCP mode is not perfect. This may be a result of key differences in the HCP and Biobank data sets, including different non-imaging measures, the use of only rfMRI in the HCP analysis, the different cohort profiles (for example, age range) and the ability to separate more modes from the larger Biobank cohort.

### Illustrative hypothesis-driven study

The Austrian Stroke Prevention Study (ASPS) recently reported associations between aging, smoking and body mass index (BMI) with gray matter T2\* in 314 participants (38–82 years)<sup>38</sup>, likely reflecting iron accumulation in local tissue<sup>13</sup>. We sought to replicate several of their key findings as a demonstration of a hypothesis-led investigation. The ASPS reported R2\*, which is the reciprocal of the T2\* value that we estimated in UK Biobank; thus, we expected T2\* associations with opposite signs to those reported by ASPS. The main results from ASPS in deep gray matter structures were that BMI was generally the strongest determinant of R2\* and was significantly related to R2\* in amygdala (beta = 0.23,  $P_{\text{FDR}} = 0.009$ ) and hippocampus (beta = 0.14,  $P_{\text{FDR}} < 0.0001$ ). Further associations with R2\* (averaged across subcortical structures) were found for age (beta = 0.03,  $P_{\text{FDR}} = 0.027$ ) and recent smoking level (beta = 0.02,  $P_{\text{FDR}} = 0.001$ ). No equivalent associations were found for sex or hypertension. The ASPS conducted univariate correlations and multiple regressions to identify both shared and unique variance in the associations, using FDR correction. On the basis of these results, we hypothesize negative association between T2\* in subcortical structures with BMI, age and smoking.

We conducted similar analyses, applying univariate correlations and multiple regressions against a similar set of covariates to ASPS (Fig. 8). The regressions used the  $n = 4,891$  subjects with complete data in all IDPs and covariates. To maximize the complementarity of information content between the univariate correlations and multiple

regressions, we applied no adjustments for factors such as age and sex in correlations, whereas we included these factors as confound covariates in the multivariate regressions. We applied Bonferroni multiple comparisons correction across covariates and brain regions, resulting (for  $P_{\text{corrected}} < 0.05$ ) in a  $-\log_{10}P_{\text{uncorrected}}$  threshold of 3.6.

Our results were highly concordant with ASPS. BMI was significantly associated with T2\* in amygdala (averaged across left and right: beta = -0.07,  $-\log_{10}P_{\text{uncorrected}} = 3.9$ ) and hippocampus (beta = -0.15,  $-\log_{10}P_{\text{uncorrected}} = 17.0$ ; for comparison, FDR correction would result in  $P_{\text{FDR}} < 10^{-10}$ ). Individual subcortical BMI associations are shown in Figure 8b. In accordance with our hypothesis, the signs of regression betas are universally negative with T2\* from UK Biobank data. Associations with T2\* were found for age in thalamus, caudate and putamen (Fig. 8c) and for smoking status in caudate, putamen and right pallidum (beta ranging from -0.03 to -0.1). Association of T2\* with sex was only found in right amygdala, and no association was found for hypertension.

The increased specificity of multiple regression is notable for many of the tests, for example, a significant univariate association of T2\* with cholesterol disappeared after controlling for the other covariates. Similarly, for T2\* in hippocampus and amygdala, many of the associations with age, sex, BMI and other factors became much weaker after controlling for all variables, particularly the amount of head motion. Despite the fact that this motion was recorded from the functional data (not the T2\* data), it is likely a general indicator of head motion, and these results illustrate why interpretation of imaging associations requires care. For example, BMI could be predictive of head motion (for example, comfort in the scanner) while also potentially relating to biophysical parameters of deeper interest.

The BMI and smoking associations with T2\* are found in distinct subcortical structures. Notably, this distinction is reflected in the CCA-ICA results, where these associations appear in separate population modes. The association of T2\* in caudate and putamen with smoking (and more weakly with alcohol; Fig. 8c) was highly concordant with CCA-ICA mode 5 (Supplementary Fig. 4b), and was associated with aging (Fig. 8c and Supplementary Fig. 8). The association of T2\* in hippocampus and amygdala with BMI was highly concordant with CCA-ICA mode 3 (Supplementary Fig. 3c), a distinct mode of population covariation that was not associated with aging (in either analysis). Neither mode includes cognitive test scores, suggesting that, although these associations clearly relate to biological processes, they may be only indirectly linked to cognitive health.

## DISCUSSION

### Challenges of population imaging

UK Biobank data is openly available to researchers, including non-imaging experts. However, imaging data is considerably more complex than most of the existing UK Biobank measures. Extensive post-processing is required to align images across subjects and remove artifacts. Moreover, information is usually encoded across multiple voxels, requiring further processing to extract relevant features. Even with carefully prepared IDPs, meaningful interpretation requires care because MRI is generally an indirect measure of the biology of interest. Apparent structural atrophy can be susceptible to misinterpretation<sup>39</sup>, fMRI signals can reflect vascular properties rather than neural activity<sup>40</sup>, and dMRI is sensitive to many aspects of tissue microstructure<sup>14</sup>. A final challenge is that data sizes have become extremely large, requiring 'big data' techniques; the brain imaging data in UK Biobank will ultimately surpass 0.2 PB even without data inflation during post-processing.

Large cohorts face the further challenge that statistically significant associations are identified even when their explanatory power is

small. In our data set, significance was reached at a correlation of just  $r \approx 0.1$ , that is, 1% of population variance explained<sup>41</sup> even with multiple comparison correction. Large genome-wide association studies (GWAS) face this challenge, where it is accepted that small effect sizes can be meaningful, particularly when multiple factors combine to create a large effect. However, in GWAS, genetic variants can be interpreted as causal factors (whether direct or indirect<sup>42</sup>), whereas apparent associations across IDPs and non-imaging phenotypes could result from a shared latent (non-measured) cause. For example, education level could result in a dietary factor associating with a brain IDP, despite no direct causal connection between diet and IDP. This danger is inflated with larger subject numbers, but may be mitigated by the rich life factor and biological variables that can be controlled for or used to match subgroups. Population variances explained in the pairwise associations reached maxima of around 5% (**Supplementary Fig. 2**), but these were higher with the multivariate analyses (up to 20–50% variance explained in the most highly involved variables in population modes), partly reflecting increased sensitivity gained when appropriately combining across related variables.

The importance of accounting for relevant confounds is exemplified in **Figure 6e**, which displays a strong apparent association between total white matter volume and fat-free body mass (one scatter point per subject) without de-confounding. In fact, this association is largely driven by the average differences in body mass and head size between sexes and disappears after adjusting for sex, age and head size. This is an example of Simpson's paradox<sup>43</sup>, in which suboptimal pooling across variables (here, sex) results in a misleading association. Other pitfalls include failing to consider study population selection bias<sup>44</sup> and inappropriate de-confounding of variables that are caused by (and not feeding into) the variables of interest<sup>45</sup>. Although there is no guarantee that UK Biobank is an unbiased sample of the full population, that does not imply that studies using subsets of the data have to retain any biases (although it is again still possible for bias to arise<sup>44</sup>); one important aspect of study design will be the method of subselection of Biobank subjects to feed into an analysis. In the case of focused hypothesis testing, it is likely that carefully selected subgroups of subjects should be used. For example, once a group of subjects is identified with a clinical diagnosis, it is likely that optimal sensitivity and interpretability will require a control subgroup that is matched over many relevant factors (for example, sex, age and relevant life factors not appearing in the predictive model).

Future studies might seek to find causalities between variables, for example, using structural equation modeling, Bayes Nets or nonlinear/non-Gaussian methods<sup>46</sup>. The dangers of inferring causalities from observational data sets such as UK Biobank are well known; the inclusion of genotype and other 'instrumental' measures enable analyses such as Mendelian randomization, although important caveats must be considered<sup>42</sup>. The safest way to confirm causal results discovered from such observational data sets is to use such results to form hypotheses for new focused interventional studies.

The mapping of disease associations and population patterns (for example, learned from UK Biobank data) onto individuals will be an important long-term goal. For example, population distributions of imaging measures and health outcomes can be learned and used to form patient-specific prior distributions to combine with measures from a new patient. Although this might not provide statistical certainty for a diagnosis or interventional recommendation, it should allow single-patient imaging to be used in a similar way to current state-of-the-art patient-tuned genetic testing.

### Data analysis in population imaging

Our analyses demonstrate some of the possibilities offered by the UK Biobank resource. Focused association studies may select just two variables to investigate, such as one IDP correlated against one life factor, genetic marker, physical assay or health outcome. More complex analyses could model a larger number of variables simultaneously, for example, looking to predict health outcome from multiple linear regression against several predictor variables. Nonlinear methods (for example, penalized regression or data-driven feature selection)<sup>22</sup> could enable use of much larger number of predictor variables. A further extension could identify nonlinear interactions between predictor variables, for example, considering an imaging measure, a life factor and an interaction term between the two as three distinct predictors. An even more complex analysis might predict multiple outcome variables, looking for 'doubly multivariate' associations between two or more sets of variables; the CCA-ICA analyses presented above are an example of this. Finally, imaging measures may in some cases be more sensitive or specific than clinical symptoms<sup>47</sup>, thereby providing proxies for healthcare outcomes and/or enabling clustering of patients that is more predictive of prognosis or therapeutic response<sup>48</sup>.

Pairwise correlation analyses result in simple outcomes that require an understanding of the caveats in imaging-derived measures. Data-driven multivariate analyses identifying associations between sets of variables have complementary benefits, including improved sensitivity to biological processes and a streamlined set of results compared with millions of univariate associations. Furthermore, multivariate analyses can separate distinct biological processes with opposing relationships between variables. For example, our CCA-ICA analysis revealed one aging-related process that involved changes in heart rate and fMRI measures (mode 4), whereas another aging-related process related blood pressure and white matter microstructure (mode 8). A simple correlational analysis would show associations between all of these factors, including even those that appeared in separate modes (for example, fMRI and white matter changes). In addition, as with multiple regression, simultaneous identification of multiple modes of association reduces the unexplained residual variance (effectively data 'de-noising').

Multivariate analyses of multi-modal data such as UK Biobank enable discovery of (potentially complex) clinical phenotypes. This is a powerful alternative to diagnostic categories that rely on clinical symptoms that do not map cleanly onto underlying disease mechanisms. For many complex diseases, the discovery of distinct mechanisms and/or subdiseases that are currently conflated may be unlikely to occur solely through symptom-based investigations. Discovering relevant population axes and subgroups on the basis of imaging, genetics and other objective markers may therefore be expected to increase our understanding of the etiology and pathogenesis of a wide variety of diseases. For example, this concept is at the heart of the recently proposed Research Domain Criteria (RDoC) in psychiatry<sup>48</sup>.

### Population imaging landscape

In the early 2000s, several ambitious studies built cohorts consisting of thousands of subjects. Several recent brain imaging studies are aiming to image tens of thousands of subjects, including the Maastricht Study ( $n = 10,000$ )<sup>49</sup>, the German National Cohort ( $n = 30,000$ )<sup>50</sup> and the Rhineland Study ( $n = 30,000$ ). In addition to having even larger numbers, UK Biobank will benefit from the breadth of organ systems imaged, the highly multi-modal brain protocol and the existing rich phenotyping. A longitudinal component is planned for a subset of the UK Biobank imaging participants ( $n = 10,000$ ), as in the Rhineland

and German National Cohort studies. Most of these studies use identical MRI scanners at a small number of dedicated sites, with the goal of maximizing data homogeneity within the study. A future challenge to further leveraging these large data sets is to develop analysis tools that can harmonize data across these studies for combined analyses, where there could be considerable benefit in focusing on harmonization of a few very large cohorts.

Even with just 5% of the eventual cohort size, our results demonstrate the statistical benefits that are conferred by large numbers. However, the primary rationale for the size of the study is not to boost statistical power across 100,000 subjects, but rather to provide prospective imaging data that are suitable for discovering early markers and risk factors for as broad a set of diseases as possible. For some rare diseases with few established risk factors, this approach is uniquely suited to discovery of pre-symptomatic markers; for example, 50–100 imaging participants are expected to develop sporadic amyotrophic lateral sclerosis (ALS) by 2027. This rich imaging addition to the ongoing UK Biobank study will provide scientists with insights into the causes of brain disease, provide markers with predictive power for therapeutic interventions and advance noninvasive imaging-based screening for preventative healthcare.

## METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

We would like to acknowledge the valuable contributions of members of the UK Biobank Imaging Working Group and the UK Biobank coordinating center. We are very grateful for additional input into the imaging protocol and image processing pipelines from M. Chappell, S. Clare, E. Duff, D. Flitney, M. Hernandez Fernandez, H. Johansen-Berg, P. McCarthy, J. Miller, D. Mortimer, J. Price, G. Salimi-Khorshidi, E. Vallee, D. Vidaurre, M. Webster, A. Winkler, A. Young, E. Auerbach, S. Moeller, K. Ugurbil, D. Alexander, N. Fox, E. Kaden, S. Ourselin, G. Zhang, A. Daducci, T. Stoecker, D. Barch, N. Bloom, G. Burgess, M. Glasser, M. Harms, D. Nolan, B. Fischl, D. Greve, J. Polimeni, T. Nichols, A. Murphy, G. Parker, F. Barkhof, C. Beckmann, M. Mennes, M. Vernooij, N. Weiskopf, C. Rorden and J. Wardlaw. We are grateful for the provision of simultaneous multi-slice (multiband) pulse sequence and reconstruction algorithms from the Center for Magnetic Resonance Research, University of Minnesota. Finally, we are extremely grateful to all UK Biobank study participants, who generously donated their time to make this resource possible. UK Biobank (including the imaging enhancement) has been generously supported by the UK Medical Research Council and the Wellcome Trust. K.L.M. and S.M.S. receive further support from the Wellcome Trust. P.M.M. acknowledges support from the Edmund J Safra Foundation and Lily Safra, the Imperial College Healthcare Trust Biomedical Research Centre, and the Medical Research Council.

## AUTHOR CONTRIBUTIONS

K.L.M., R.C., P.M.M. and S.M.S. provided the overall scientific strategy for UK Biobank brain imaging. K.L.M., N.K.B., D.L.T., E.Y., J.X., A.J.B., S.J., S.N.S., J.L.R.A., M.J., P.M.M. and S.M.S. developed acquisition protocols. N.K.B., K.L.M., T.W.O., P.W., I.D., S.G. and S.H. implemented the imaging protocol at the dedicated imaging center. F.A.-A., K.L.M., S.J., S.N.S., J.L.R.A., L.G., G.D., M.J. and S.M.S. developed post-processing pipelines and IDP calculation. K.L.M. and S.M.S. carried out the univariate and multivariate analyses and prepared figures. K.L.M. and S.M.S. wrote the manuscript, which was edited by all of the authors.

## COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Filippini, N. *et al.* Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele. *Proc. Natl. Acad. Sci. USA* **106**, 7209–7214 (2009).
- Douaud, G. *et al.* Brain microstructure reveals early abnormalities more than two years prior to clinical progression from mild cognitive impairment to Alzheimer's disease. *J. Neurosci.* **33**, 2147–2155 (2013).
- Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
- Allen, N. *et al.* UK Biobank: current status and what it means for epidemiology. *Health Policy and Technology* **1**, 123–126 (2012).
- Wick, J.Y. Understanding frailty in the geriatric population. *Consult Pharm.* **26**, 634–645 (2011).
- Stanek, K.M. *et al.* Obesity is associated with reduced white matter integrity in otherwise healthy adults. *Obesity (Silver Spring)* **19**, 500–504 (2011).
- Petersen, S.E. *et al.* Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank - rationale, challenges and approaches. *J. Cardiovasc. Magn. Reson.* **15**, 46 (2013).
- Van Essen, D.C. *et al.* The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62–79 (2013).
- Good, C.D. *et al.* A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* **14**, 21–36 (2001).
- Karas, G.B. *et al.* A comprehensive study of gray matter loss in patients with Alzheimer's disease using optimized voxel-based morphometry. *Neuroimage* **18**, 895–907 (2003).
- Debette, S. & Markus, H.S. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *Br. Med. J.* **341**, c3666 (2010).
- Haacke, E.M., Xu, Y., Cheng, Y.-C.N. & Reichenbach, J.R. Susceptibility weighted imaging (SWI). *Magn. Reson. Med.* **52**, 612–618 (2004).
- Duyn, J. MR susceptibility imaging. *J. Magn. Reson.* **229**, 198–207 (2013).
- Beaulieu, C. The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR Biomed.* **15**, 435–455 (2002).
- Basser, P.J., Mattiello, J. & LeBihan, D. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson. B.* **103**, 247–254 (1994).
- Zhang, H., Schneider, T., Wheeler-Kingshott, C.A. & Alexander, D.C. NODDI: practical *in vivo* neurite orientation dispersion and density imaging of the human brain. *Neuroimage* **61**, 1000–1016 (2012).
- Jbabdi, S., Sotiropoulos, S.N., Savio, A.M., Graña, M. & Behrens, T.E. Model-based analysis of multishell diffusion MR data for tractography: how to get over fitting problems. *Magn. Reson. Med.* **68**, 1846–1855 (2012).
- de Groot, M. *et al.* Improving alignment in Tract-based spatial statistics: evaluation and optimization of image registration. *Neuroimage* **76**, 400–411 (2013).
- Wakana, S. *et al.* Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage* **36**, 630–644 (2007).
- Logothetis, N.K. What we can do and what we cannot do with fMRI. *Nature* **453**, 869–878 (2008).
- Hariri, A.R., Tessitore, A., Mattay, V.S., Fera, F. & Weinberger, D.R. The amygdala response to emotional stimuli: a comparison of faces and scenes. *Neuroimage* **17**, 317–323 (2002).
- Smith, S.M. *et al.* Functional connectomics from resting-state fMRI. *Trends Cogn. Sci.* **17**, 666–682 (2013).
- Douaud, G. *et al.* DTI measures in crossing-fibre areas: increased diffusion anisotropy reveals early white matter alteration in MCI and mild Alzheimer's disease. *Neuroimage* **55**, 880–890 (2011).
- Ennis, D.B. & Kindlmann, G. Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images. *Magn. Reson. Med.* **55**, 136–146 (2006).
- Raichle, M.E. *et al.* A default mode of brain function. *Proc. Natl. Acad. Sci. USA* **98**, 676–682 (2001).
- Satterthwaite, T.D. *et al.* An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* **64**, 240–256 (2013).
- Kuznetsova, K. *et al.* Brain white matter structure and information processing speed in healthy older age. *Brain Struct. Funct.* **221**, 3223–3235 (2016).
- Madden, D.J. *et al.* Diffusion tensor imaging of cerebral white matter integrity in cognitive aging. *Biochim. Biophys. Acta* **1822**, 386–400 (2012).
- Van Der Werf, Y.D. *et al.* Thalamic volume predicts performance on tests of cognitive speed and decreases in healthy aging. A magnetic resonance imaging-based volumetric analysis. *Brain Res. Cogn. Brain Res.* **11**, 377–385 (2001).
- Fjell, A.M. & Walhovd, K.B. Structural brain changes in aging: courses, causes and cognitive consequences. *Rev. Neurosci.* **21**, 187–221 (2010).
- Reichle, E.D., Carpenter, P.A. & Just, M.A. The neural bases of strategy and skill in sentence-picture verification. *Cognit. Psychol.* **40**, 261–295 (2000).
- Hotelling, H. Relations between two sets of variates. *Biometrika* **28**, 321–377 (1936).
- Hyvärinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **10**, 626–634 (1999).
- Smith, S.M. *et al.* A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.* **18**, 1565–1567 (2015).
- Smith, S.M. *et al.* Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. USA* **106**, 13040–13045 (2009).



36. Yaffe, K., Browner, W., Cauley, J., Launer, L. & Harris, T. Association between bone mineral density and cognitive decline in older women. *J. Am. Geriatr. Soc.* **47**, 1176–1182 (1999).
37. Tan, Z.S. *et al.* Bone mineral density and the risk of Alzheimer disease. *Arch. Neurol.* **62**, 107–111 (2005).
38. Pirpamer, L. *et al.* Determinants of iron accumulation in the normal aging brain. *Neurobiol. Aging* **43**, 149–155 (2016).
39. Hamidi, M., Drevets, W.C. & Price, J.L. Glial reduction in amygdala in major depressive disorder is due to oligodendrocytes. *Biol. Psychiatry* **55**, 563–569 (2004).
40. D'Esposito, M., Deouell, L.Y. & Gazzaley, A. Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. *Nat. Rev. Neurosci.* **4**, 863–872 (2003).
41. Kirk, R. Practical significance: a concept whose time has come. *Educ. Psychol. Meas.* **56**, 746–759 (1996).
42. Gage, S.H., Davey Smith, G., Ware, J.J., Flint, J. & Munafò, M.R. G = E: what GWAS can tell us about the environment. *PLoS Genet.* **12**, e1005765 (2016).
43. Simpson, E. The interpretation of interaction in contingency tables. *J. R. Stat. Soc. B* **13**, 238–241 (1951).
44. Swanson, J.M. The UK Biobank and selection bias. *Lancet* **380**, 110 (2012).
45. Berkson, J. Limitations of the application of fourfold table analysis to hospital data. *Biometrics* **2**, 47–53 (1946).
46. Pearl, J. *Causality: Models, Reasoning and Inference* (Cambridge University Press, 2009).
47. Duff, E.P. *et al.* Learning to identify CNS drug action and efficacy using multistudy fMRI data. *Sci. Transl. Med.* **7**, 274ra16 (2015).
48. 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).
49. Schram, M.T. *et al.* The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *Eur. J. Epidemiol.* **29**, 439–451 (2014).
50. German National Cohort (GNC) Consortium. The German National Cohort: aims, study design and organization. *Eur. J. Epidemiol.* **29**, 371–382 (2014).

## COMPETING FINANCIAL INTERESTS

P.W. and I.D. are employees of Siemens Healthcare UK, the vendor of MRI scanners for UK Biobank, selected under a competitive bidding process.

## EDITORIAL SUMMARY

**AOP:** The UK Biobank combines detailed phenotyping and genotyping with tracking of long-term health outcomes in a large cohort. This study describes the recently launched brain-imaging component that will ultimately scan 100,000 individuals. Results from the first 5,000 subjects are reported, including thousands of associations, population modes and hypothesis-driven results.

## ONLINE METHODS

**Protocol overview.** Imaging protocols were designed by the UK Biobank Imaging Working Group (<http://www.ukbiobank.ac.uk/expert-working-groups>), in consultation with a large number of brain imaging experts (listed in the acknowledgments). MRI provides many imaging modalities offering complementary information. As part of this consultancy, a number of modalities were determined to be infeasible or lower priority for a range of reasons. Considerations included time constraints, generalizability, feasibility of automated analysis, and existence of robust, well-tested acquisition methods. The advisory network therefore decided not to include quantitative relaxometry, MR spectroscopy or angiography. Arterial spin labeling is currently being piloted, as described below.

To maximize data compatibility, three dedicated imaging centers will have identical scanners with fixed platforms (that is, no major software or hardware updates throughout the study). Each center is equipped with a 3T Siemens Skyra (software platform VD13), 1.5T Siemens Aera (VD13), carotid ultrasound and dual energy X-ray absorptiometry (DEXA). Brain imaging is being conducted on the 3T system using a 32-channel receive head coil.

Key acquisition parameters for each modality are summarized in **Supplementary Table 1**, grouped according to primary modality categories (structural MRI, dMRI and fMRI). Order of acquisition was optimized in consideration of subject compliance, assuming subject motion might increase over the scan (favoring early acquisition of the T1 due to its central importance; for example, the processing pipeline cannot run without the T1) and subject wakefulness might decrease (favoring early acquisition of fMRI). The order is: (1) T1, (2) resting fMRI, (3) task fMRI, (4) T2 FLAIR, (5) dMRI, (6) swMRI.

Further protocol details are available at <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=2367> and further description of post-processing pipelines and data outputs included in the first data release are available at [http://biobank.ctsu.ox.ac.uk/crystal/docs/brain\\_mri.pdf](http://biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf). All software used in these pipelines is freely available<sup>51,52</sup> and full pipeline processing scripts will shortly be publicly available.

The processing pipeline used for the initial data release was primarily based on tools from FSL (the FMRIB Software Library<sup>51</sup>), but it will be gradually expanded to utilize a broader range of methods and software, where this will increase the quality, robustness and scope of IDPs generated. For example, one high priority is to adapt the Human Connectome Project pipelines<sup>53</sup> to provide cortical surface modeling.

The intention is that non-imaging experts will be able to use the IDPs directly without having to become expert in the complexities of data processing, although we encourage engagement with imaging experts in light of the numerous and subtle caveats and confounds associated with interpreting these data.

Data access requests from all academic or commercial researchers (with no exclusive or preferential access) are processed by the UK Biobank's Research Access Administration Team and approved relatively rapidly provided that they fulfill UK Biobank's aims of supporting health research in the public interest. Researchers' institutions then sign a Material Transfer Agreement agreeing not to attempt to identify any participant, and to return any derived data (for example, new IDPs) to UK Biobank, to be made available to other approved researchers after an agreed 'embargo' period (to allow findings to be published or IP protected by the researchers). Thus, while the first set of IDPs described here from internally commissioned research is being made available immediately, the range of IDPs is expected to grow rapidly as additional contributions from the wider user community are added.

**Protocol considerations.** Design of the brain imaging protocol was conducted through broad consultation with neuroimaging experts and required careful balance of a range of considerations, often specifically relating to the high throughput nature of UK Biobank. In setting up the pilot protocol, the primary challenge was to achieve the target of one participant scanned every 36 min without serious compromise to data quality compared to research protocols that might conventionally require up to an hour of scan time. Despite these tight time constraints, we aimed to include as many MRI modalities as possible, to take advantage of the full richness of information that can be provided by MRI. Here, we highlight the primary considerations that required a different approach from more conventional imaging studies.

- With each additional minute of scanning per subject effectively costing an additional ~£1million, there is enormous value associated with seemingly

small efficiency savings. We recovered several minutes of scan time by systematically minimizing the overheads associated with subject placement, scan prescription, and calibration measurements. For example, corrections to the static magnetic field homogeneity (shimming) and strict enforcement of a single shim calibration harvested 2 min (changing system defaults to improve and accelerate shimming), which is equivalent to the scan time associated with some of the included modalities.

- Tight imaging FOVs (fields of view, the physical size of the imaged volume) are in general favorable to reduce scan time; however, these restrictions exclude subjects with larger heads or brains. For UK Biobank, even a 'conservative' FOV that includes 99% of the population will exclude 1,000 participants. As detailed statistics on brain size (as distinct from head size) were not available in the literature, we conducted a study of population brain size<sup>54</sup> that (in conjunction with optimal slice angling) enables our FOVs to target 99.9%.
- It is critically important that all analyses are automated. This translates to an additional role for certain imaging modalities beyond their intrinsic information content. Thus, although we considered methods for reducing scan time for T1-weighted structural scans while retaining coverage and resolution (for example, elliptical sampling with consequent image blurring), this was deemed an unacceptable risk given the central role of the T1 to cross-subject and cross-modal alignment for most processing pipelines, including that implemented here for the initial data release.
- The EPI (echo-planar imaging) acquisitions for fMRI and dMRI result in significant image distortion that creates local misalignment in certain brain regions. Correction of this requires measurement of the magnetic field inhomogeneities that cause distortion. Two types of measurements are possible: a non-EPI gradient-echo acquisition with two echoes (conventional field map) or two EPI-based spin echo acquisitions with opposite phase encode direction<sup>55</sup>. We chose the latter, which can be incorporated into the dMRI protocol as additional b = 0 scans to reduce acquisition time (total acquisition time ~30 s).
- To provide data with as rich and broad a range of applications as possible, we include imaging modalities that are not yet widely used in clinical practice, such as fMRI and dMRI. These modalities have demonstrated mechanistic and biological insights, and will hopefully see greater clinical take-up in the future, in part because of projects such as UK Biobank. We took advantage of recent advances in acquisition, largely developed as part of the Human Connectome Project, to obtain research quality data in limited time. Specifically, simultaneous multi-slice (or multiband, MB) acquisitions<sup>56–59</sup> that enable rapid fMRI and dMRI without sacrificing statistical robustness or directions/b-values (ref. 60), respectively. Without these accelerations, a seven-minute dMRI scan of the same spatial resolution would have been limited to ~32 directions and a single shell, precluding NODDI<sup>16</sup> and other more advanced biological modeling.
- After early piloting, a clinical T2/PD-weighted acquisition was removed from the protocol. This decision reflected the limited relevance to UK Biobank goals (given the inclusion of the higher-quality and more biologically informative T2 FLAIR) and the value in recovering this scan time (just over 1 min).
- One shortcoming of the current protocol is the lack of a direct measure of neurovascular health. We are piloting a protocol change to include a 2-min perfusion scan (using arterial spin labeling). This would require reducing task fMRI to 2 min; while this is an extremely short task, early analyses (using truncated copies of existing initial tfMRI data sets) predict that it will be sufficiently robust.
- A major ethical question in studies of this nature relates to identification and handling of incidental findings of previously unknown pathology. The procedure to be followed in UK Biobank has been considered in great depth with major external ethical, legal and clinical radiology bodies, and with the funders and their external review group. An assessment of different approaches to the identification of incidental findings and the impact of their feedback on participants and the health service has been conducted as part of the pilot phase of UK Biobank's imaging project, and will be published separately. Based on its results and the deliberative process undertaken with external experts, the UK Biobank protocol for dealing with incidental findings does not involve the routine review of all scans for potential pathology

by radiologists. Instead, if a radiographer incidentally identifies evidence of potentially serious pathology (that is, likely to threaten life span, quality of life or major body functions) during the imaging process then a formal radiologist review is undertaken and, if it is confirmed as potentially serious, feedback is given to the participant and their doctor.

- Informed consent is obtained from all UK Biobank participants; ethical procedures are controlled by a dedicated Ethics and Guidance Council (<http://www.ukbiobank.ac.uk/ethics>) that has developed with UK Biobank an Ethics and Governance Framework (given in full at <http://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf>), with IRB approval also obtained from the North West Multi-center Research Ethics Committee.
- Subjects are excluded from scanning according to fairly standard MRI safety/quality criteria, such as exclusions for metal implants, recent surgery, or health conditions directly problematic for MRI scanning, such as problems hearing, breathing or extreme claustrophobia.
- Once the second and third imaging centers are complete and running, UK Biobank will use phantom objects and traveling volunteers to confirm quality and consistency across sites.

**Structural imaging.** The T1 structural protocol is acquired at 1mm isotropic resolution using a three-dimensional (3D) MPRAGE acquisition, with inversion and repetition times optimized for maximal contrast. The superior-inferior field-of-view is large (256 mm), at little cost, in order to include reasonable amounts of neck/mouth, as those areas will be of interest to some researchers (for example, in the study of sleep apnea). Pre-processing of this modality included removal of the face (which was deemed important to subject anonymization for the standard data dissemination), brain extraction (removal of non-brain tissues from the image), linear alignment to the standard MNI152 brain template<sup>61</sup> and nonlinear warping to this template<sup>62</sup> to maximize correspondence across individuals in light of significant cross-subject variation in brain structure. These alignments are used throughout the majority of the processing pipeline for other modalities.

T1 images are further analyzed to estimate volumes of a range of tissues and structures in each subject, which may reflect atrophy due to age and disease, as well as normal variation due to (for example) use-dependent plasticity. Images are segmented into tissue types (gray matter, white matter and cerebrospinal fluid)<sup>63</sup>. Cortical gray matter volume is estimated, comparing the segmented gray matter to an atlas reference (where the external skull surface is used to normalize for head size)<sup>64</sup>. Sub-cortical volumes are estimated<sup>65</sup>, using population priors on shape and intensity variation across subjects. T1-based IDPs are generated for the volumes of major tissue types of the whole brain and for specific structures (primarily sub-cortically).

Too much reliance on spatial registration could limit the usefulness or accuracy of some IDPs. This is in part why many of the IDPs are in fact generated from within-subject analyses that do not depend on exact voxelwise spatial alignment to standard space (or between subjects): for example, 283 of the 715 structural and diffusion IDPs do not rely on exact spatial alignment and are carried out in the original space of each subjects' data.

The T2 protocol uses a fluid-attenuated inversion recovery (FLAIR) contrast with the 3D SPACE optimized readout<sup>66</sup>. This shows strong contrast for white matter hyperintensities. An automated pipeline for delineating these hyperintensities is currently being developed and future data releases will include IDPs reflecting the lesion 'load'.

The swMRI scan uses a 3D gradient echo acquisition at 0.8x0.8x3mm resolution, acquiring two echo times (TE = 9.4 and 20 ms). Anisotropic voxels can enhance certain contrast mechanisms, particularly for vascular conspicuity due to through-plane dephasing effects, but are less ideal for other susceptibility-based processing. Ultimately, however, this choice was motivated by the desire for whole brain coverage in the face of very limited scan time (2.5 min). Signal decay times (T2\*) are estimated from the magnitude images at the two TEs, and the generated IDPs are the median T2\* estimated within the various subcortical regions delineated from the T1 processing. Venograms are generated through nonlinear filtering of the magnitude and phase images<sup>12</sup>, which produces enhanced conspicuity of medium and large veins. Automated segmentation of microbleeds and venograms would provide significant value, but to our knowledge robust tools for this are not yet available; future pipeline versions can hope to include such analyses. Future work will also consider whether this data will

support quantitative susceptibility mapping, which would provide further information on tissue constituents as discussed in the main text.

**Diffusion imaging.** Diffusion data is acquired with two b-values (b = 1,000 and 2,000 s/mm<sup>2</sup>) at 2-mm spatial resolution, with multiband acceleration factor of 3 (three slices are acquired simultaneously instead of just one). For each diffusion-weighted shell, 50 distinct diffusion-encoding directions were acquired (covering 100 distinct directions over the two b-values). The diffusion preparation is a standard (monopolar) Stejskal-Tanner pulse sequence. This enables higher SNR due to a shorter echo time (TE = 92 ms) than a twice-refocused (bipolar) sequence at the expense of stronger eddy current distortions, which are removed using the Eddy tool<sup>67</sup> (which also corrects for static field distortion and motion<sup>68</sup>).

Both diffusion tensor and NODDI models are fit voxel-wise, and IDPs of the various model outputs are extracted from a set of white matter tracts. Tensor fits utilize the b = 1000 s/mm<sup>2</sup> data, producing maps including fractional anisotropy, tensor mode and mean diffusivity. The NODDI<sup>16</sup> model is fit using the AMICO (Accelerated Microstructure Imaging via Convex Optimization) tool<sup>52</sup>, with outputs including intra-cellular volume fraction (which is often interpreted to reflect neurite density) and orientation dispersion (a measure of within-voxel disorganization). For tractography, a parametric approach is first used to estimate fiber orientations. The generalized ball & stick model is fit to the multi-shell data, estimating up to three crossing fiber orientations per voxel<sup>17,69</sup>. Tractography is then performed in a probabilistic manner to estimate white matter pathways using the voxel-wise orientations.

Cross-subject alignment of white matter pathways is critical for extracting meaningful IDPs; here, two complementary approaches are used. The first used tract-based spatial statistics (TBSS)<sup>18,70</sup>, in which a standard-space white matter skeleton is mapped to each subject using a high-dimensional warp, after which ROIs are defined as the intersection of the skeleton with standard-space masks for 48 tracts<sup>71</sup> (see the JHU ICBM-DTI-81 white-matter labels atlas described at <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases> for definitions of the tract regions and names). The second approach utilizes subject-specific probabilistic diffusion tractography run using standard-space protocols to identify 27 tracts<sup>18</sup>; in this case, the output IDPs are weighted by the tractography output to emphasize values in regions that can most confidently be attributed to the tract of interest. Currently, no structural connectivity estimates from the diffusion tractography are provided as IDPs, but the probabilistic maps are available and future work will generate measures similar to those provided for resting-state fMRI.

**Functional MRI.** Task and resting-state fMRI use the same acquisition parameters, with 2.4-mm spatial resolution and TR = 0.735 s, with multiband acceleration factor of 8. A 'single-band' reference image (without the multiband excitation, exciting each slice independently) is acquired that has higher tissue-type image contrast; this is used as the target for motion correction and alignment. For both data sets, the raw data are corrected for motion<sup>72</sup> and distortion<sup>55</sup> and high-pass filtered to remove temporal drift.

The task scan used the Hariri faces/shapes 'emotion' task<sup>21,73</sup>, as implemented in the HCP<sup>22</sup>, but with shorter overall duration and hence fewer total stimulus block repeats. The participants are presented with blocks of face or shape trials and asked to decide which of two faces (or shapes) presented on the bottom of the screen match the face (or shape) at the top of the screen. The faces have either angry or fearful expressions. The ePrime stimulus script is available for download (<http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=1462>). Task-induced activation is modeled with FEAT, including auto-correlation correction<sup>74</sup>, using five activation contrasts. Of these, the three activation contrasts of most interest (shapes, faces and faces>shapes) are used to generate output measures, including two IDPs for the faces-shapes task (one including all voxels above a group-level fixed-effects Z > 120, and one including only the amygdala regions above threshold). IDPs corresponding to both percent signal change and statistical significance (Z statistics) are generated.

During resting-state scans, subjects are instructed to keep their eyes fixated on a crosshair, relax and 'think of nothing in particular'. Resting-state networks are identified using ICA (independent component analysis<sup>33,75</sup>), which identifies components within the data that are spatially independent (where a component comprises a spatial map and a single associated time course). Following the pre-processing described above, resting-state fMRI data for each subject is further 'cleaned' using an ICA-based algorithm for automatically identifying and

removing structured artifacts<sup>76</sup>. This data is fed into group-level ICA (including an initial group-level dimensionality reduction<sup>77</sup>), which is used to parcellate the data set into sets of 25 and (separately) 100 spatially independent components. Where a small (<30) number of components is estimated<sup>78</sup>, it is common to consider each component as a separate ‘network’ in its own right; each component will often include several non-contiguous regions, all having the same time course (according to the model). If a higher number of components is estimated<sup>79</sup>, these are more likely to be smaller regions (parcels), which can then be considered as nodes for use in network analysis<sup>80</sup>, where the spatial maps are used to define subject-specific time courses (the first stage of dual regression<sup>1</sup>). These time courses are used to estimate the size of signal fluctuation in each node, as well as to estimate connectivity between pairs of nodes using L2-regularised partial correlation<sup>81</sup>. The connectivity estimates are provided as IDPs at both parcellation dimensionalities (25 and 100 nodes); after removal of group-ICA components considered to be artifactual (that is, relating either to scanning artifacts, or to non-neuronal biophysical processes such as cardiac fluctuations and head motion), this results in 21 and (respectively) 55 nodes left for forming the IDPs such as network matrices (functional connectivities between pairs of nodes).

**Quality control.** To date, raw data and pipeline outputs have been manually checked for gross problems of quality and robustness, with problematic data tagged and removed from pipeline outputs; see main text for results on proportions of usable data in the different modalities. However several quality-related IDPs are automatically generated by the pipeline (for example, number of outlier slices in the dMRI data, and measures of signal-to-noise ratio in the various modalities), and these can be used to help automatically identify problematic data. An expanded set of such quality measures is being produced, in addition to an automated machine learning system for flagging problematic data on the basis of the many IDPs and quality measures; future versions of the pipeline and data releases will benefit from the results of these ongoing developments.

**Statistics.** The two sections below describe the statistical analysis carried out using IDPs and non-brain-imaging measures. As described below, univariate statistics were primarily carried out using Pearson correlation (see details below regarding Gaussian-distribution normalization and linear removal of confound effects) and multivariate statistics were carried out using a combination of canonical correlation analysis and independent component analysis (with permutation testing used to identify the significant number of components estimable). As discussed in the main text, the primary rationale for the size of the study is not to boost statistical power across 100,000 subjects, but rather to provide prospective imaging data suitable for discovering early markers and risk factors for as broad a set of diseases as possible, both rare and highly prevalent. Hence while calculations have been made to estimate the expected numbers of subjects developing different diseases over coming years (see introductory section of main text), no statistical methods were used to pre-determine sample sizes for any one specific disease, given that individual disease sample sizes are not prospectively controlled, and given the very broad expected set of future tests between different imaging measures and different diseases that will be ultimately applied from this prospective long-term resource. Details on significance testing and multiple comparison corrections are included in the two sections below.

A **Supplementary Methods Checklist** is available.

**Simple associations between brain IDPs and other measures.** We report simple correlation analyses between each of the 2,501 brain IDPs and each of 1,100 other variables extracted from the UK Biobank database (these other variables are mostly not derived from imaging, though some do come from the non-brain imaging modalities); for the list of general classes of these variables, see **Figure 6a**, and for many examples of individual variables, see the lists associated with the CCA-ICA modes presented in **Figure 7** and **Supplementary Figures**. The initial set of variables extracted from the UK Biobank database was automatically reduced to those (1,100 variables) containing sufficient numbers of valid (non-missing) data entries, using very similar selection rules to those applied in the recent CCA-based analysis of Human Connectome Project data<sup>34</sup>. Some variables are defined (in the UK Biobank database) such that the numerical encoding is the inverse of what one might naturally assume - for example in the variable ‘Qualifications’, higher numbers refer to lower levels of educational qualifications. In such cases we have inverted the sign of the ICA weightings printed in

the figures, for ease of interpretation. Further, some variables are categorical, with no clear quantitative meaning to the values (for example, ‘Transport type to work’); where we find an apparent association, this can be considered to be indicative of a real association (one might think of the analysis therefore as an over-conservative poor implementation of an ANOVA), but interpretation of the sign of the association clearly needs care. The analysis used data from the first 5,430 subjects scanned and having usable imaging data: age range 44–78 years (IQR 56–68 years); 53% of subjects were female.

Eight confound variables are generated: age, age<sup>2</sup>, sex, age × sex, age<sup>2</sup> × sex, average head motion during fmMRI, average head motion during rfMRI and head size. To enforce Gaussianity, all confound variables, IDPs and non-IDP variables are first passed through a rank-based inverse Gaussian transformation; this improves the robustness of correlations (for example, to avoid undue influence of potential outlier values). The confounds are then regressed out of all IDPs and non-IDP variables to reduce the risk of finding nonmeaningful associations. For example, head motion corrupts imaging data in complex ways<sup>26</sup>, and also correlates with some diseases and with aging ( $r = 0.15$  in this data); hence, if not adjusted for, uninteresting associations would likely arise. However, some measures may have both biologically interesting associations with IDPs, and also act as imaging confounds. For example, abnormal heart rate or blood pressure could alter the fMRI signal through disrupted cerebral auto-regulation (independent of any changes to neural activity)<sup>40</sup>, but cardiovascular pathology could also be related to neurological pathology. Similarly, overall brain size and gray matter thickness IDPs are sensitive simple markers of aging and disease; however, these properties can also affect other IDPs by changing the mixture of tissue types in an imaging voxel, creating an apparent age/disease dependence that is driven by the volume of tissue rather than the properties of a given tissue type (such as fMRI activation or white matter microstructural properties). It is therefore important to interpret apparent associations carefully.

The full set of 2.8 million ( $2,501 \times 1,100$ ) Pearson correlations is then estimated and corrected for multiple comparisons. Bonferroni correction, which is likely to be somewhat conservative in such situations, due to non-independence across variables tested, resulted in  $P_{\text{corrected}} < 0.05$  being equivalent to requiring  $P_{\text{uncorrected}} < 1.8 \times 10^{-8}$ . An alternative popular approach for multiple comparison correction is false discovery rate (FDR)<sup>82</sup>; we use the more conservative FDR option (making no assumption of variable dependencies<sup>83</sup>), resulting here in requiring  $P_{\text{uncorrected}} < 3.8 \times 10^{-5}$ . These two threshold levels are shown with dotted lines in all Manhattan plots in the main figures.

**Multivariate associations between brain IDPs and other measures.** In the example multivariate analyses shown in **Figures 7** and **8**, canonical correlation analysis (CCA<sup>32</sup>) combined with independent component analysis (ICA<sup>33</sup>) is used to identify several ‘modes’ of population covariation which link multiple brain IDPs to sets of other Biobank variables. This is very similar to the methodology used recently to identify a single mode of population covariation between imaging measures and many behavioral and lifestyle measures in data from 461 subjects in the Human Connectome Project<sup>8,34</sup>.

IDP and non-IDP variables are prepared as for the univariate correlation analyses described above, resulting in a brain-IDP matrix of size  $5,034 \times 2,501$  (subjects × IDPs) and a non-IDP matrix of size  $5,034 \times 1,100$  (subjects × non-IDP variables). The intention is to feed these into CCA in order to identify population modes linking multiple variables from both matrices. However, in order to avoid an over-determined (rank deficient) CCA solution, we first compress both matrices along the respective phenotype dimension to 200 columns (that is, much smaller than the numbers of subjects). This was done by separately reducing each matrix to the top 200 subject-eigenvectors using PCA. To achieve this while avoiding the problem of missing data, we use the approach detailed recently<sup>34</sup> of estimating first a pseudo-covariance matrix ignoring missing data, projecting this onto the nearest valid (positive definite) covariance matrix, and then carry out an eigenvalue decomposition. The two resulting (IDP and non-IDP) matrices of size  $5,034 \times 200$  are then fed into standard CCA (‘*canoncorr*’ in Matlab), resulting in 200 CCA modes being estimated. The CCA aims to identify symmetric linear relations between the two sets of variables. Each significant CCA mode identifies a linear combination of IDPs and a linear combination of non-IDPs, where the variation in mode strength across subjects is maximally correlated. That is, CCA finds modes that relate sets of brain measures to sets of subjects’ non-brain-imaging measures; for a graphical illustration of this approach see Supplementary Information in ref. 34.

Permutation testing is then applied to estimate (family-wise-error, multiple-comparison-corrected)  $P$  values for the CCA modes estimated. Nine modes are found to be significant ( $P_{\text{corrected}} < 0.002$ , with all later modes having  $P_{\text{corrected}} > 0.05$ ). Because CCA can in general only unambiguously estimate distinct modes up to an orthogonal rotation amongst them (by direct analogy to PCA), we identify an unambiguous unmixing of the modes using ICA to optimize the final set of modes reported. Because we expect meaningful population modes to be much more structured (for example, sparser) in the cross-variable dimension than in the cross-subject dimension, we calculate ICA components that are statistically independent from each other in the cross-variable dimension. In order to take full advantage of the numbers of variables originally prepared, we first multiply the nine CCA subject-weight vectors into the original IDP and non-IDP data matrices (after concatenating these across variables), resulting in nine CCA variable-weight vectors of length  $2,501 + 1,100 = 3601$ . These nine vectors are then fed into FastICA<sup>33</sup> to estimate nine population data sources having maximal statistical independence. This general approach (CCA, followed by concatenation of CCA weight vectors, followed by ICA) is similar to that proposed by Sui<sup>84</sup>, except that we return to the full feature space (as described above) for the ICA stage, rather than staying in the PCA-reduced space. The ICA result is extremely robust, with split-half (cross-subjects) reproducibility across the 9 ICA components of  $r > 0.89$ . Interestingly, 5 of these ICA modes (including modes 7, 8 and 9; Fig. 7) are virtually unchanged if the de-confounding step was omitted (correlation of variable-weights vectors:  $r > 0.8$ ).

**Data, code and results availability.** As described above, all source data (including raw and processed brain imaging data, derived IDPs, and non-imaging measures) is available from UK Biobank via their standard data access procedure (see <http://www.ukbiobank.ac.uk/register-apply>).

The image processing pipeline will be made publicly available in early 2017 from <http://www.fmrib.ox.ac.uk/ukbiobank> - this is the pipeline used to process the raw imaging data and generate IDPs, and hence is not needed in order to replicate the results of this paper, which could be achieved by accessing IDPs as described above, and then using the IDP analysis code described below.

The Matlab code for the univariate and multivariate tests described in this paper, and the results of those tests (all univariate correlations and multivariate weight vectors) are available from <http://www.fmrib.ox.ac.uk/ukbiobank>; this online resource will be updated as more subjects' data and more IDPs become available.

Higher resolution supplementary figures are available in the PDF version of the supplementary information online.

51. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W. & Smith, S.M. FSL. *Neuroimage* **62**, 782–790 (2012).
52. Daducci, A. *et al.* Accelerated Microstructure Imaging via Convex Optimization (AMICO) from diffusion MRI data. *Neuroimage* **105**, 32–44 (2015).
53. Glasser, M.F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).
54. Mennes, M. *et al.* Optimizing full-brain coverage in human brain MRI through population distributions of brain size. *Neuroimage* **98**, 513–520 (2014).
55. Andersson, J.L., Skare, S. & Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* **20**, 870–888 (2003).
56. Uğurbil, K. *et al.* Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Neuroimage* **80**, 80–104 (2013).
57. Larkman, D.J. *et al.* Use of multicoil arrays for separation of signal from multiple slices simultaneously excited. *J. Magn. Reson. Imaging* **13**, 313–317 (2001).
58. Moeller, S. *et al.* Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn. Reson. Med.* **63**, 1144–1153 (2010).
59. Setsompop, K. *et al.* Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magn. Reson. Med.* **67**, 1210–1224 (2012).
60. Feinberg, D. *et al.* Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PLoS One* **5**, e15710 (2010).
61. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
62. Andersson, J., Jenkinson, M. & Smith, S. Non-linear registration aka spatial normalization. in *FMRIB Technical Report* (Oxford University, 2007).
63. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).
64. Smith, S.M. *et al.* Longitudinal and cross-sectional analysis of atrophy in Alzheimer's disease: cross-validation of BSI, SIENA and SIENAX. *Neuroimage* **36**, 1200–1206 (2007).
65. Patenaude, B., Smith, S.M., Kennedy, D.N. & Jenkinson, M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* **56**, 907–922 (2011).
66. Mugler, J.P. III. Optimized three-dimensional fast-spin-echo MRI. *J. Magn. Reson. Imaging* **39**, 745–767 (2014).
67. Andersson, J.L. & Sotiropoulos, S.N. Non-parametric representation and prediction of single- and multi-shell diffusion-weighted MRI data using Gaussian processes. *Neuroimage* **122**, 166–176 (2015).
68. Andersson, J.L. & Sotiropoulos, S.N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* **125**, 1063–1078 (2016).
69. Hernández, M. *et al.* Accelerating fibre orientation estimation from diffusion weighted magnetic resonance imaging using GPUs. *PLoS One* **8**, e61892 (2013).
70. Smith, S.M. *et al.* Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* **31**, 1487–1505 (2006).
71. Wakana, S., Jiang, H., Nagae-Poetscher, L.M., van Zijl, P.C.M. & Mori, S. Fiber tract-based atlas of human white matter anatomy. *Radiology* **230**, 77–87 (2004).
72. Bannister, P.R., Brady, J.M. & Jenkinson, M. Integrating temporal information with a non-rigid method of motion correction for functional magnetic resonance images. *Image Vis. Comput.* **25**, 311–320 (2007).
73. Barch, D.M. *et al.* Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189 (2013).
74. Woolrich, M.W., Ripley, B.D., Brady, M. & Smith, S.M. Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage* **14**, 1370–1386 (2001).
75. Beckmann, C.F. & Smith, S.M. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* **23**, 137–152 (2004).
76. Salimi-Khorshidi, G. *et al.* Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* **90**, 449–468 (2014).
77. Smith, S.M., Hyvärinen, A., Varoquaux, G., Miller, K.L. & Beckmann, C.F. Group-PCA for very large fMRI datasets. *Neuroimage* **101**, 738–749 (2014).
78. Kiviniemi, V., Kantola, J.-H., Jauhiainen, J., Hyvärinen, A. & Tervonen, O. Independent component analysis of nondeterministic fMRI signal sources. *Neuroimage* **19**, 253–260 (2003).
79. Kiviniemi, V. *et al.* Functional segmentation of the brain cortex using high model order group PICA. *Hum. Brain Mapp.* **30**, 3865–3886 (2009).
80. Smith, S.M. The future of fMRI connectivity. *Neuroimage* **62**, 1257–1266 (2012).
81. Smith, S.M. *et al.* Network modelling methods for fMRI. *Neuroimage* **54**, 875–891 (2011).
82. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
83. Genovese, C.R., Lazar, N.A. & Nichols, T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* **15**, 870–878 (2002).
84. Sui, J. *et al.* A CCA+ICA based model for multi-task brain imaging data fusion and its application to schizophrenia. *Neuroimage* **51**, 123–134 (2010).