Data supplement for Roberts et al., Longitudinal Changes in Structural Connectivity in Young People at High Genetic Risk for Bipolar Disorder. Am J Psychiatry (doi: 10.1176/appi.ajp.2021.21010047)

SUPPLEMENTARY METHODS

Participants

Participants were involved in an ongoing longitudinal study with annual follow-up evaluations. High Risk participants were recruited from families who had either previously participated in a bipolar disorder (BD) pedigree molecular genetics study, from a specialised BD research clinic, or were otherwise recruited from clinicians, mental health consumer organisations and other forms of publicity. Control participants were recruited via print and electronic media as well as noticeboards in universities and local communities.

Control participants were defined as those who did not have a first-degree relative with either BD I or II, recurrent major depressive disorder, schizophrenia, schizoaffective disorder, recurrent substance abuse or any past psychiatric hospitalisation. Additionally, they did not have a second-degree relative with a history of psychosis or who had been hospitalised for a mood disorder. High Risk and Control participants with a lifetime or current presence of psychiatric symptoms (apart from the occurrence of BD) were not excluded from the study. This ecological approach has been used by similar studies of individuals at high genetic risk for BD to recruit both Control and High Risk cohorts (1).

We benchmarked these longitudinal data against a cohort of gender-matched participants with bipolar disorder for whom a single cross-sectional scan had been acquired (2). Bipolar Disorder participants were drawn from the same study, age-matched to the High Risk and Control cohorts at follow-up. The Bipolar Disorder participants met DSM-IV criteria for either bipolar I or bipolar II disorder. Longitudinal data were not available from this clinical cohort.

Sample selection and characterization

All potential participant families were initially assessed through baseline administration of the Family Interview for Genetic Studies (FIGS) (3) to determine family history of BD and other disorders for eligibility. At least one parent or participant aged over 22 in each family completed the FIGS. Parents were interviewed about their child (aged 12-21) using the Kiddie-Schedule for Affective Disorders and Schizophrenia for School-Aged Children – Present and Lifetime Version (K-SADS-BP) (4). The 12-21-year-olds also completed a K-SADS-BP interview as part of a collaborative High Risk study with a US consortium (1). For all participants aged 22-30 (including BD-probands), the Diagnostic Interview for Genetic Studies (DIGS) was administered to determine diagnoses (5). Consensus DSM-IV diagnoses of BD-probands, High Risk, and Control participants were determined by two independent raters using best estimate methodology (6), drawing from the DIGS, K-SADS-BP, FIGS, and medical records.

To assess current mood state, the Children's Depression Inventory (7) was administered to participants aged 12-21 years. Both the Montgomery–Åsberg Depression Rating Scale (8) and Young Mania Rating Scale (9) were administered to those aged 22–30 years. Intellectual ability was assessed with the Wechsler Abbreviated Scale of Intelligence (10).

For the cross-sectional Bipolar Disorder cohort, proband consensus DSM-IV diagnosis was also determined by two independent raters following Best Estimate methodology using information from the Diagnostic Interview for Genetic Studies (DIGS) Version 4, the Family Interview for Genetic Studies (FIGS) and medical records (when available) (see (2) for full details for the Bipolar Disorder cohort).

Total MRI scans and exclusions

A total of 279 participants completed a baseline diffusion-weighted (dMRI) scan and 217 participants completed a follow-up scan approximately 2 years later. 79 subjects who were excluded (31 Control and 48 High Risk) had a baseline dMRI scan but did not complete a follow-up diffusion-weighted scan. 17 subjects (7 Control and 10 High Risk) were excluded because they did not complete a baseline dMRI scan. Of the 200 subjects with dMRI scans at both time-points (87 Control and 112 High Risk), 1 High Risk subject was removed because there were more than three years between their scans, another High Risk subject was excluded on clinical grounds. Two participants were also removed (1 Control and 1 Hig-Risk) due to image artifact and distortion in

a baseline or follow up scan. One participant (1 Control) was removed due to excessive head movement dMRI scans (mean rms motion > 2). Thus, 195 subjects with usable baseline and follow-up dMRI scans remained (86 Control and 109 High Risk). To minimise possible confounding effects of gender and age, we selected 2 age- and gender-matched samples comprising 86 Control and 97 High Risk subjects, by removing younger Control subjects. There was no effect of time on head motion, pooled across both groups (independent t-test, t=1.70, p=0.092).
Diffusion MRI images were available for 52 participants with bipolar disorder, age-matched to the High Risk and Control cohorts at time of follow-up, meeting the same quality control criteria (Table S4).

Data acquisition, pre-processing and structural network construction

MRI acquisition & pre-processing of DWI

Data acquisition

Diffusion-weighted MRI data were acquired with a 3-T Philips Achieva scanner at Neuroscience Research Australia (NeuRA) in Sydney with an 8-channel head coil. One acquisition of 32 directional DWI (b =1000 s/mm^2, with one non-diffusion-weighed image) was acquired using a single-shot echo planar imaging (EPI) sequence. The imaging parameters were as follows: TR = 7767 ms, TE = 68 ms, 55 slices, slice thickness = 2.5 mm, gap = 0 mm, acquisition matrix size = 96 × 96 (field of view = 240×240×137.5 mm), flip angle = 90°, reconstructed to yield 1 mm × 1 mm × 2.5 mm voxels (where the longer dimension is along the dorsoventral axis).
A T1-weighted image was also acquired in the imaging session but was not used in connectome generation or analysis.
Diffusion-weighted MRI data were pre-processed using functions within MRtrix3 software (11), (https://github.com/MRtrix3/mrtrix3/releases/tag/3.0_RC3), and FSL, which were called from a pre-processing pipeline developed in-house (https://github.com/breakspear/diffusion-pipeline/tree/bipolarlongitudinal). First, the dMRI data were denoised (12), and then corrected for motion and eddy-current induced distortions within FSL *eddy* (version 5.0.11) (13). FSL eddy (using --repol) was also used to detect slice signal outliers due to bulk motion (among other factors), and then corrected for using a non-parametric replacement method (13). Finally, bias-intensity correction was performed using the *dwibiascorrect* (*-fsl* argument) (14).

All diffusion of volumes of each participant underwent a final visual quality assurance. This included inspection of (1) each diffusion weighted volume with *fslview*; (2) the generated brain mask, (3) the mean diffusion volume across the scan, (4) the FA image, and (5) the track density image of the reconstructed streamlines. Finally, to determine the accuracy of the co-registration steps (see below) used in structural connectome construction, each subject-specific parcellation image was overlaid on their FA map.

This quality assurance pipeline is available at

https://github.com/AlistairPerry/dwi2connectome/blob/master/miscscripts/QC_wfslview_basic .

Whole-brain fibre tractography

To provide estimates of local fibre orientations, the relative signal responses of multiple tissue-types (anisotropic white matter and isotropic grey and CSF) were obtained using an unsupervised algorithm (i.e. no anatomical priors) (15, 16). The average white-matter and CSF signal contribution were then partitioned and constrained spherical deconvolution (*lmax* = 6, *msmt_csd*) (17), was used to provide the fibre orientation distribution functions (fODF), reflecting local estimates (i.e. voxel-wise) of the apparent density of fibre direction as a function of angular orientation. The iFOD2 probabilistic streamline algorithm was employed to generate plausible whole-brain fibre propagations by random sampling of the orientation uncertainty inherent in each fODF at points along each candidate path (post bug-fix, c.f. https://github.com/MRtrix3/mrtrix3/issues/1204). Tracking parameters were as follows: step size = 0.2 mm, minimum length = 10 mm, max length = 250 mm, FOD termination threshold = 0.1, curvature constraint = 1 mm radius, with 5 million streamlines per subject initialized from random seeds throughout the brain mask.

Whole-brain structural network construction

To construct a high resolution parcellation, the 90 regions of the AAL atlas were subdivided into 512 subregions of approximately equivalent volume using a random parcellation method (18), identical to that used in our baseline report (Table S1). This parcellation template of 512 regions was co-registered into a standard-space template representing an average FA image (FMRIB58; available within FSL). For each individual, the FMRIB58 1 mm template was co-registered to the subject's FA image (*dof* = 12; cost function = *normmi*, with the latter used as the reference image).

The parcellation template was then transformed into subject space by applying the transformation matrix generated above from co-registering the FA template to the individual's FA image.

The anatomical information of subject-specific parcellations and fibre streamline trajectories were combined to yield whole-brain structural connectivity graphs. A weighted connection, A_{ij}, corresponds to the total number of fibre streamlines which start/terminate within a radial 2 mm distance of voxels located in each of the corresponding parcels *i* and *j*. Due to the greater number of random seeds along longer white matter tracts, fibre densities are typically over-estimated in longer fibre bundles (19). Distance-correction was performed by adjusting each streamline A_{ij} by the physical fibre length between its start and termination point (*i* and *j*) (20).

The diffusion pipeline is available at (https://github.com/breakspear/diffusion-pipeline/tree/bipolarlongitudinal).

To reduce false positive connections, the resulting connectivity matrices were thresholded using a group consistency approach. In brief, weighted edges were rank ordered by their consistency, defined as the coefficient of variation (SD/mean) across subjects. The most consistent edges above a chosen threshold were retained; edges less consistent than this threshold were set to zero (21). This ensures that all subjects have the same edges, differing only in their weight and thus avoiding a composite of weights and zeros across subjects in individual edges. Following our previous cross-sectional study of this cohort (2), structural networks were thresholded with a connection density of 10%.

The sum of the weights of the suprathreshold edges could also, in principle, still differ between groups. However, total connectivity weights in our data did not show a significant effect of group (p=.066), time (p=.085), or time by group interaction (p=.624).

Network controllability

We implemented measures from linear network control theory to understand the dynamic consequences of High Risk structural connectivity differences. Linear control theory assumes that neural dynamics can be approximated by linear, discrete-time models of regional activity.(22, 23) The propagation of neuronal perturbations along structural projections in response to exogenous inputs (24, 25), is given by,

$$x_i(t+1) = \sum_j W_{ij} x_i . x_j(t) + B_i u(t),$$

where $x_i(t)$ denotes the state of network node (brain parcel) *i* at time *t*. *W* denotes the weighted structural connectivity matrix, with element W_{ij} representing the connectivity between nodes *i* and *j* as defined above. The input vector *B* identifies the control nodes in the network (those receiving the perturbation), with the nodes in our study controlled one at a time. The term *u*(*t*) is the energy applied to the set of control nodes *B* at time *t*.

To ensure the linear stability of all individual structural networks, following our previous work (25), we normalized all network edge weights (W_{ij}) through division by a number 10% greater than the largest eigenvalue calculated across all subjects' connectivity matrices *W*.

Average controllability for a control node is defined as the average energy needed to steer the network to any target state in finite time (24). Complimenting the NBS analysis, controllability hence interrogates the outward connectivity of a network to the rest of the connectome (24, 26). Nodes with lower controllability values exert weaker control over brain states (27).

In the present study, average controllability was calculated for all regions identified in our NBS group by time interaction model. That is, (virtual) perturbations were applied to all nodes within the subnetwork of interest and then allowed to propagate outwards through the entire connectome. The total subnetwork controllability was derived by summing the controllability of all nodes in this network. Average controllabilty is simply referred to as "controllability" in the main text, noting that there are other independent controllability metrics (28).

For the code used in the initial controllability studies, see
https://complexsystemsupenn.com/s/controllability_code-smb8.zip (*ave_control.m*)
Network calculations and parcellation data are available at
https://github.com/AlistairPerry/CNHRLongitudinal.

SUPPLEMENTARY RESULTS

Network edge consistency

Networks constructed using consistency-based thresholding showed an approximately exponential relationship between weight and inter-areal distance. Consistent with prior work, consistent edges show a similar distance effect to the entire (unthresholded) network (Figure S2A), although they tend to be stronger than the unthresholded network by distance and lack the hard lower cut-off present in networks thresholded by a weight-based approach (29).

Edges that were consistent in each group at baseline are also amongst those that were consistent in each group at follow-up. For example, 83.1% of the most consistent edges at baseline are also among the most consistent at follow-up (Figure S2B). Moreover, the 10% most consistent edges at baseline are among the 12.2% of most consistent edges at follow-up. Likewise, consistent edges are very similar across groups: 79.6% of the most consistent edges in the High Risk group were among the most consistent in the Control group. The 10% most consistent edges in the High Risk group were among the 12.6% of most consistent edges in the Control group.

The high concordance of consistent edges suggests that consistency-based thresholding could, in principle, be applied separately to each group and at each time point. However, while relatively few in number, the mismatching edges would have a spread of weights in one group but be thresholded uniformly to zero in the other. This would introduce a sharp difference between groups, with zero variance in one group and thus a high likelihood of a false positive effect. For this reason, we applied a single consistency threshold across both groups at both times and focus on the re-weighting of the ensuing edges using NBS.

Comparison with Bipolar Disorder cohort

The weighted average of the interaction subnetwork in the Control cohort at baseline was similar to the Bipolar Disorder network, but increased substantially by follow-up (Figure 2B). In contrast, the weighted average of this network in the High Risk cohort crossed from slightly above to slightly below the Bipolar Disorder network from baseline to follow-up. Over time there was a marginally significant reduction in the subnetwork weights for the High Risk cohort (MD=-.55, p=.051) and a significant increase in the subnetwork weights for the Control cohort (p=.001). There was no significant difference between the Bipolar Disorder subnetwork weights and the follow-up High Risk weights (p=0.421). However, there was a significantly higher weighted average in the Control at follow-up (p=.03) than in the Bipolar Disorder cohort. Bayes-analyses of these results favour a difference in the Bipolar Disorder and Control network weights at follow-up (BF=1.64)

but favour no difference between the High Risk and Bipolar Disorder cohorts (BF=0.25, Table S5). Restricted to the follow-up T2 time point, there was no significant age by group (HR, CN, BD) effect for the weighted average of this subnetwork ($p>0.70$; Figure S6).

To supplement these post-hoc analyses, we also tested for the interaction of age x group across all three groups at follow-up, using NBS to examine all nodes in the structural connectome. There is no significant effect for the interaction of age and group in these follow-up data ($p>0.61$).

Auxiliary Analyses

Subsequent exploratory analyses were undertaken to study putative effects of medication or affective episodes between baseline and follow-up imaging.

Two sets of linear-mixed effects models were run to study the effects of medication. The first set used the base models (as described in the main text) with the addition of a binary predictor for current use of any psychotropic medication (n=20). This comprised of 15 subjects at baseline (3 Control, 12 High Risk) and 10 subjects (1 Control, 9 High Risk) remaining on psychotropic medication at follow-up, with an additional 5 subjects (3 Control, 2 High Risk) using psychotropic medications at follow-up only. Similar to the models examining medication effects, the base models were also re-run with the addition of a binary predictor for current mood episode (n=13) to examine possible confounding effects. This comprised of 4 subjects at baseline scan (2 Control, 2 High Risk) with an additional 8 subjects (4 Control, 4 High Risk) experiencing a current episode at time of follow-up scan. One subject (High Risk) experiencing a mood episode at baseline was also experiencing an episode at follow-up.

Current mood state associations with the group by time network effect (Figure 2) and controllability were explored by including current mood state measures as a predictor in the base models. The Children's Depression Inventory(7) was administered to participants aged 12-21 years, and both the Montgomery–Åsberg Depression Rating Scale (8), and Young Mania Rating Scale (9), were administered to those aged 22–30 years. As mood state was assessed using separate scales for younger (12-21-year-olds) and older (22-30-year-olds) groups, separate models were run for each age group (all other analyses were run with the total sample).

To investigate if the group by time network effect or controllability reflected DSM-IV disorders in our High Risk group, we performed various exploratory sub-group analyses. Firstly, we sub-divided our High Risk group according to the following criteria: i) new onset of any mood episode

(major depressive or manic/hypomanic episode); ii) new onset of manic/hypomanic episode; or iii) new onset of any DSM-IV disorder from baseline to follow-up. To benchmark new against prior episodes, we additionally divided the Hig Risk cohort into iv) those with or without a lifetime mood episode at baseline. Due to the small sample sizes of these High Risk sub-groups we estimated effect sizes for differences and did not report the associated p-values which are likely to be unstable due to finite size effects. We additionally used Bayesian repeated measures ANOVA (https://jasp-stats.org, 1 million samples) to determine the relative evidence in favour of an effect in subgroups (Bayes Factors, BF).

TABLE S1. Additional baseline and follow-up demographic and clinical data

	Control n=86	High Risk n=97	Difference Statistic	p	Difference
Meeting Criteria for Current DSM-IV Episode at Baseline					
Major Depressive Episode [#], *n (%)*	2 (2.3)	2 (2.1)	N/A	*p*=1.00	
Hypo/manic episode, *n (%)*	-	-	N/A		
Meeting Criteria for Current DSM-IV Episode at Follow-Up					
Major Depressive Episode [#], *n (%)*	4 (4.7)	5 (5.2)	N/A	*p*=1.00	
Hypo/manic episode, *n (%)*	-	1 (1.0)	N/A		
Symptom Severity Scales at baseline					
22 to 30 years	**n=43**	**n=39**			
MADRS[$] *mean (SD)*	2.44 (4.21)	3.95 (5.17)	*U* = 732.00	*p*=0.30	
YMRS[$] *mean (SD)*	0.93 (1.10)	0.97 (1.71)	*U* =762.50	*p*=0.44	
12 to 21 years	**n=40**	**n=49**			
CDI[$] *mean (SD)*	5.30 (4.44)	7.31 (7.13)	*U* =896.50	*p*=0.49	
Clinical Characteristics at baseline					
Global Functioning	**n=86**	**n=93**			
GAS/C-GAS[$], *mean (SD)*	91.13 (5.72)	86.31 (9.41)	*U*=2747.50	*p*<.001***	Control > High Risk
Age at First	**n=10**	**n=28**			
Major Depressive Episode, *mean (SD)*	17.90 (3.14)	16.14 (4.42)	*t*=-1.15	*p*= 0.26	
Number of Episodes					
Major Depressive Episode[$], *mean (SD)*	1.80 (1.55)	3.32 (3.56)	*U*=111.00	*p*= .30	
	n=86	**n=97**			
Baseline Psychotropic Medication					
At least one psychotropic medication[#], *n (%)*	3 (3.5)	12 (12)	N/A	*p*= 0.03*	High Risk > Control
Anti-depressants[#], *n (%)*	2 (2.3)	10 (10)	N/A	*p*=0.03*	High Risk > Control
Mood stabilisers, *n (%)*	-	-	-		
Lithium, *n (%)*	-	-	-		
Anti-psychotics, *n (%)*	-	-	-		
Anti-convulsants, *n (%)*	-	-	-		
Other psychotropic medication[#], *n (%)*	-	2	N/A	*p*=.62	
Benzodiazepines, *n (%)*	1 (1.2)	-	N/A		
Stimulants, *n (%)*	-	-	-		

	n=86	n=97		
Follow Up Psychotropic Medication				
At least one psychotropic medication[#], *n (%)*	4 (4.7)	11 (11)	N/A	*p*=0.11
Anti-depressants[#], *n (%)*	4 (4.7)	8 (8.2)	N/A	*p*=0.38
Mood stabilisers, *n (%)*	-	1 (1.0)	N/A	
Lithium, *n (%)*	-	-		
Anti-psychotics, *n (%)*	-	1 (1.0)	N/A	
Anti-convulsants, *n (%)*	-	1 (1.0)	N/A	
Other psychotropic medication[#]	1 (1.2)	1 (1.0)	N/A	*p*=1.00
Benzodiazepines, *n (%)*	-	2 (2.1)	N/A	
Stimulants, *n (%)*	-	-		

Abbreviations: MADRS, Montgomery-Asberg Depression Rating Scale; YMRS, Young Mania Rating Scale; CDI, Children's Depression Inventory: GAS, Global Assessment Scale; C-GAS, Children's Global Assessment Scale.

[$] Variable did not meet assumptions for parametric analysis. Mann-Whitney U was used. [#] Variable did not meet assumptions for parametric analysis. Fisher's exact test was used. [*] $p<.05$; [**] $p<.01$; [***] $p<.001$.

Diagnostic details: Diagnostic confidence rating ranges using the Best Estimate Methodology vary from 1-4, where 1 represents criteria not met for a diagnosis and 4 represents a definite diagnosis. All diagnoses listed here have a confidence rating of 3 or higher. The 183 subjects belonged to 157 families.

TABLE S2. Edge weights of connections for significant NBS derived Time (T2 > T1) subnetwork, adjusting for age at time of scan (t = 3; p<0.024, FWE-corrected)

Network	Node pair	Baseline	Follow Up
Left Middle Cingulum (a) ↔ Left Anterior Cingulum	104 ↔ 162	0.28	0.25
Left Middle Cingulum (b) ↔ Left Anterior Cingulum	139 ↔ 162	0.52	0.53
Left Anterior Cingulum ↔ Left Precuneus (a)	162 ↔ 250	0.16	0.16
Left Anterior Cingulum ↔ Left Precuneus (b)	162 ↔ 252	0.27	0.32
Left Anterior Cingulum ↔ Right Middle Cingulum	162 ↔ 343	0.62	0.58
Left Frontal Superior Medial Gyrus ↔ Right Middle Cingulum (a)	219 ↔ 343	0.19	0.19
Left Anterior Cingulum ↔ Right Middle Cingulum (a)	229 ↔ 343	0.549	0.56
Left Caudate ↔ Right Middle Cingulum (a)	238 ↔ 343	0.22	0.24
Left Anterior Cingulum ↔ Right Middle Cingulum (b)	162 ↔ 360	0.14	0.14
Left Precuneus (b) ↔ Right Precuneus	252 ↔ 449	1.38	1.22
Left Posterior Cingulum ↔ Right Frontal Superior Medial Gyrus	70 ↔ 479	0.03	0.03
Left Frontal Superior Medial Gyrus ↔ Right Frontal Superior Medial Gyrus	223 ↔ 479	10.10	10.39
Left Caudate ↔ Right Frontal Superior Medial Gyrus	238 ↔ 479	0.64	0.78
Left Frontal Superior Medial Gyrus ↔ Right Caudate	223 ↔ 494	1.21	1.29
Left Anterior Cingulum ↔ Right Precuneus (a)	162 ↔ 506	0.11	0.12
Left Anterior Cingulum ↔ Right Precuneus (b)	162 ↔ 508	0.17	0.19

TABLE S3. Edge weights of connections for significant Group by Time NBS subnetwork (t = 3; p=0.007, FWE-corrected)

Network	Node pair	Control (*n* = 86)		High Risk (*n* = 97)	
		Baseline	Follow Up	Baseline	Follow Up
Left Caudate ↔ Left Inferior Frontal Gyrus/ Pars Triangularis	10 ↔ 24	1.48	1.33	1.28	1.58
Left Caudate ↔ Left Fusiform (a)	10 ↔ 81	0.07	0.06	0.07	0.09
Left Fusiform (b) ↔ Left Putamen	16 ↔ 82	0.12	0.10	0.11	0.13
Left Fusiform (b) ↔ Left Inferior Frontal Gyrus / Pars Triangularis (b)	16 ↔ 136	0.05	0.04	0.04	0.06
Left Fusiform (a)↔ Left Inferior Frontal Gyrus / Pars Triangularis (b)	81 ↔ 136	0.06	0.06	0.06	0.08
Left Lingular Gyrus ↔ Left Inferior Frontal Gyrus/ Pars Triangularis[$]	90 ↔ 136	0.05	0.05	0.05	0.07
Left Inferior Frontal Gyrus / Pars Triangularis (b) ↔ Left Middle Frontal Gyrus	136 ↔ 230	23.55	22.09	21.23	24.58
Right Precuneus (b) ↔ Right Middle Occipital Gyrus	306 ↔ 311	4.19	3.79	3.28	4.32
Right Middle Occipital Gyrus ↔ Right Thalamus	311 ↔ 319	1.20	0.90	0.94	0.98
Right Precuneus (b) ↔ Right Cuneus	306 ↔ 417	22.59	19.40	20.58	21.16
Left Fusiform (a)↔ Right Superior Occipital Gyrus	81 ↔ 463	0.31	0.26	0.28	0.32
Right Thalamus ↔ Right Superior Occipital Gyrus	319 ↔ 463	0.90	0.75	0.71	0.79
Right Inferior Occipital Gyrus (a) ↔ Right Inferior Occipital Gyrus (b)	411 ↔ 495	68.62	66.93	67.60	76.29
Right Superior Occipital Gyrus ↔ Right Inferior Occipital Gyrus (b)	463 ↔ 495	3.67	3.34	3.60	4.3

The NBS derived network for group by time interaction. After controlling for age at the time of scan (t = 3.0, p<0.007 FWE corrected) a network that contains 13 of the 14 edges was observed. [$]Edge that was not part of the group by time network when covarying for age at time of scan.

TABLE S4. Demographic and clinical data of the 52 Bipolar Disorder subjects

Demographic data	***Bipolar Disorder (n=52)***
Females, *n (%)*	26 (50.0)
Age (years), *mean (SD)*	24.74 (3.80)
Intelligence Quotient, *mean (SD)*	116.06 (12.41)
Clinical Data	
Bipolar subtype	
Bipolar II, n (%)	25 (*48.0*)
Bipolar I, n (%)	27 (*51.9)*
Any anxiety disorder, *n (%)*	26 (50.0)
Any behavioural disorder, *n (%)*	10 (19.2)
Any substance disorder, *n (%)*	14 (26.9)
Any other disorder, *n (%)*	7 (13.5)
Meeting Criteria for Current DSM-IV Episode	
Major Depressive Episode, *n (%)*	1 (1.9)
Hypo/manic episode, *n (%)*	1 (1.9)
Symptom Severity Scales	
22 to 30 years	**n=41**
MADRS *mean (SD)*	10.11 (9.77)
YMRS *mean (SD)*	4.90 (4.43)
12 to 21 years	**n=11**
CDI *mean (SD)*	20.45 (8.62)
Clinical Characteristics	
Global Functioning	**n=52**
GAS/C-GAS, *mean (SD)*	77.12 (12.34)
	n=48
Age at first Major Depressive Episode, *mean (SD)*	15.31 (3.67)
	n=52
Age at onset of bipolar (years), mean (SD)	*17.61 (4.28)*
Psychotropic Medication	
At least one psychotropic medication, *n (%)*	40 (76.9)
Anti-depressants, *n (%)*	23 (44.2)
Mood stabilisers, *n (%)*	35 (67.3)
Lithium, *n (%)*	14 (26.9)
Anti-psychotics, *n (%)*	16 (30.8)
Anti-convulsants, *n (%)*	30 (57.7)
Other psychotropic medication[#], *n (%)*	6 (11.5)

*One subject is missing an IQ score.

TABLE S5. Comparisons between Control and High Risk connectivity of NBS group by time network at follow-up with Bipolar Disorder patients

Group means		Mean (SD)
Control	8.50 (1.90)	
High Risk	9.63 (2.21)	
Bipolar Disorder	8.78 (2.14)	
Group comparisons		*p*-val (BF_{10})
Control vs Bipolar Disorder	0.03 (1.64)	
High Risk vs Bipolar Disorder	0.42 (0.25)	

BF_{10} = Bayes Factor for evidence of alternative hypothesis

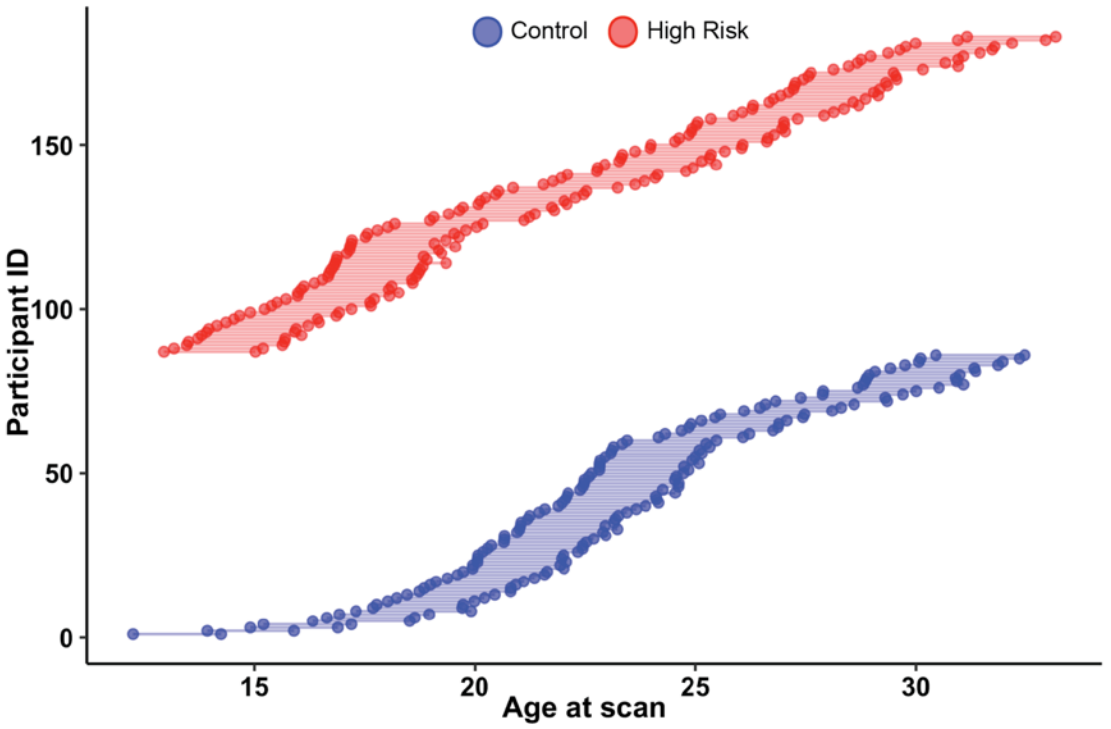
N.B. Both *p*-vals and Bayes Factors are uncorrected for multiple testing

TABLE S6. Effect sizes and Bayes Factors for longitudinal sub-group analyses

		Group x Time NBS (Threshold = 3.0)		Controllability
		Effect sizes (d)	Bayes Factor (BF_{10})	Effect sizes (d)
Original model	All High Risk (n=97)	0.28		0.07
New onset any DSM-IV	High Risk with new onset of any DSM-IV diagnosis, (n=18)	0.38	0.69[%]	0.06
	High Risk minus High Risk with new onset of any DSM-IV diagnosis, (n=79)	0.26		0.07
New mood episode	High Risk with new mood episode, (n=8)	0.52	1.20[$]	0.03
	High Risk minus High Risk with new mood episode, (n=89)	0.27		0.03
Lifetime mood episode at baseline	High Risk with lifetime mood episode at baseline, (n=28)	0.36	0.65[^]	0.26
	High Risk minus High Risk with lifetime mood episode at baseline, (n=69)	0.25		0.02
Converters	High Risk who converted to BD, (n=5)	0.97	2.61[#]	0.31
	High Risk minus High Risk who converted to BD, (n=92)	0.24		0.06
	High Risk subjects without a lifetime mood episode (n=64)	0.23		

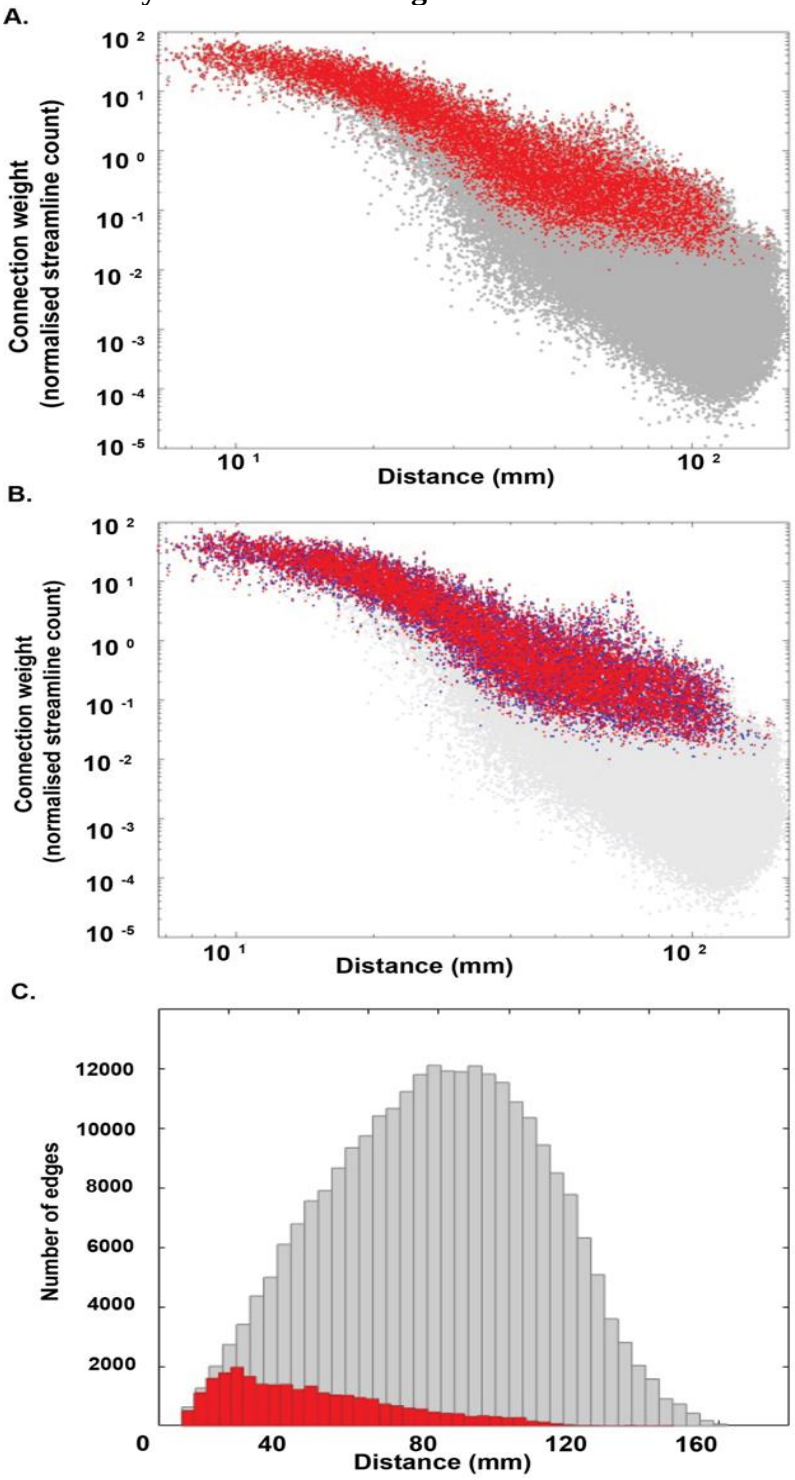
[%]High Risk with new onset of any DSM-IV diagnosis, (n=18) vs High Risk minus High Risk with new onset of any DSM-IV diagnosis, (n=79). [$] High Risk with new mood episode, (n=8) vs High Risk minus High Risk with new mood episode, (n=89). [^] High Risk with any affective diagnosis at baseline, (n=30) vs High Risk minus High Risk with any affective diagnosis at baseline, (n=67). [#] High Risk who converted to BD, (n=5) vs High Risk minus High Risk who converted to BD, (n=92).

FIGURE S1. Interval between the baseline and follow-up imaging scan for all participants



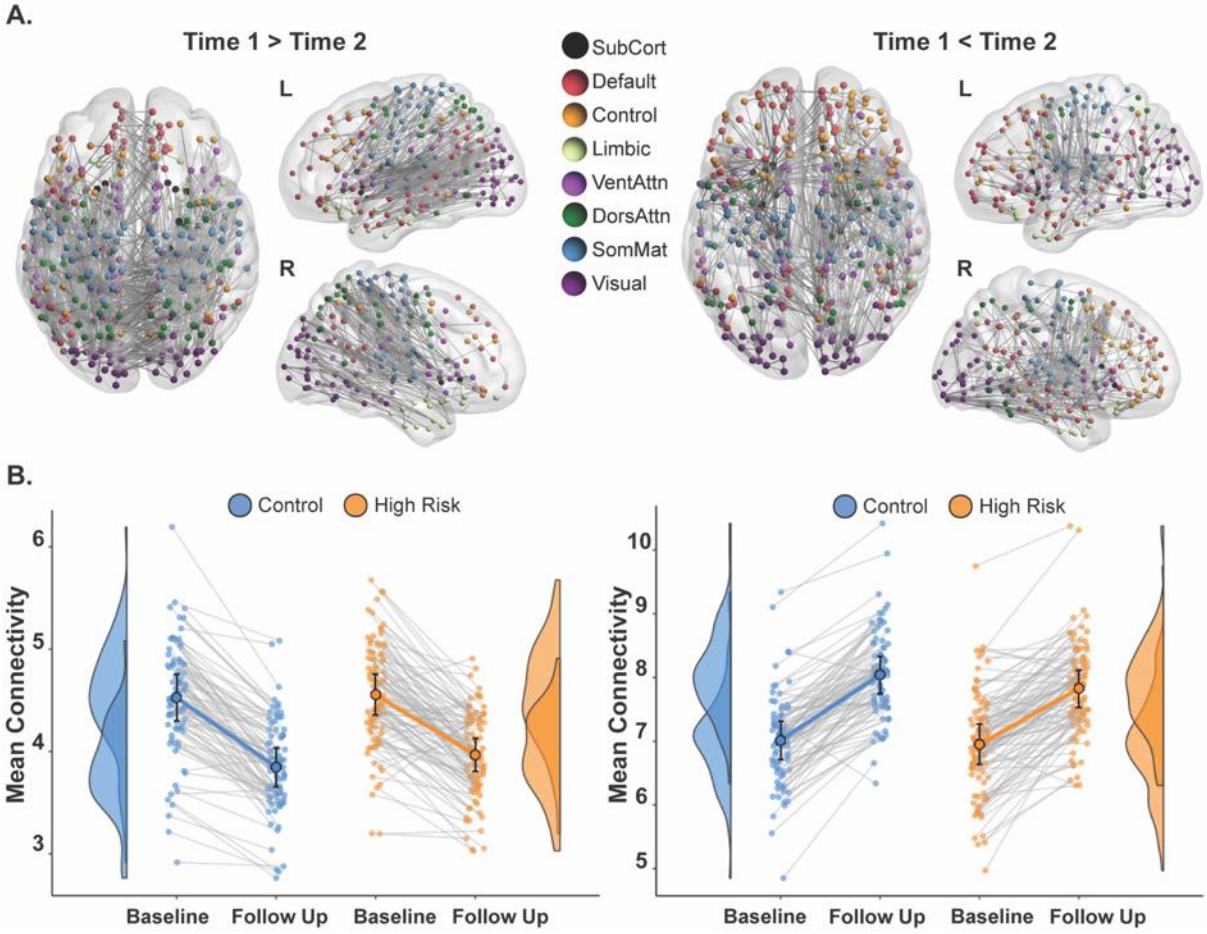
The position along the y-axis for participants in each group is ordered according to their age at baseline imaging scan (ascending order). The mean inter-scan interval did not differ between groups (controls 2.04 +/-0.14 years; High Risk 2.07 +/-0.13 years). However, the control group had a significantly greater positive skew (1.64 versus 1.48, ks-stat = 0.23, *p*=0.014).

FIGURE S2. Consistency of connectome edges across group and imaging time points with consistency-based thresholding



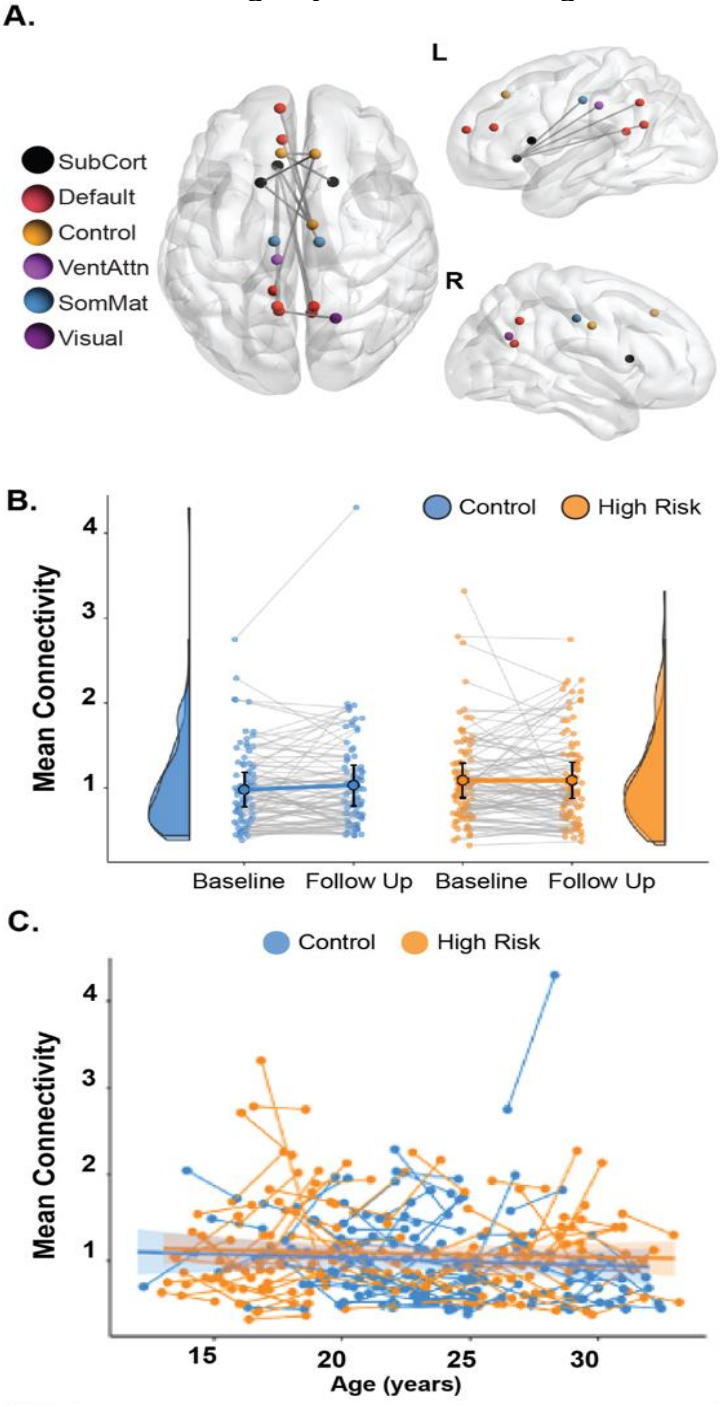
(A) Consistency-based threshold across both groups and both time points in double logarithmic coordinates (red) shows a similar distance-weight relationship to the entire unthresholded connectome (grey). (B) 83.1% of the 10% most consistent edges at baseline (red) are also among the most 10% consistent at follow-up (blue). (C) Consistency-thresholded edges tend to be stronger and somewhat shorter than the full connectome, with a median (mean) length of 39.0 cm (44.1 cm).

FIGURE S3. Longitudinal changes in structural connectivity across both High Risk and Control groups at a liberal test-threshold (*t* = 3; *p* < 0.05, FWE corrected)



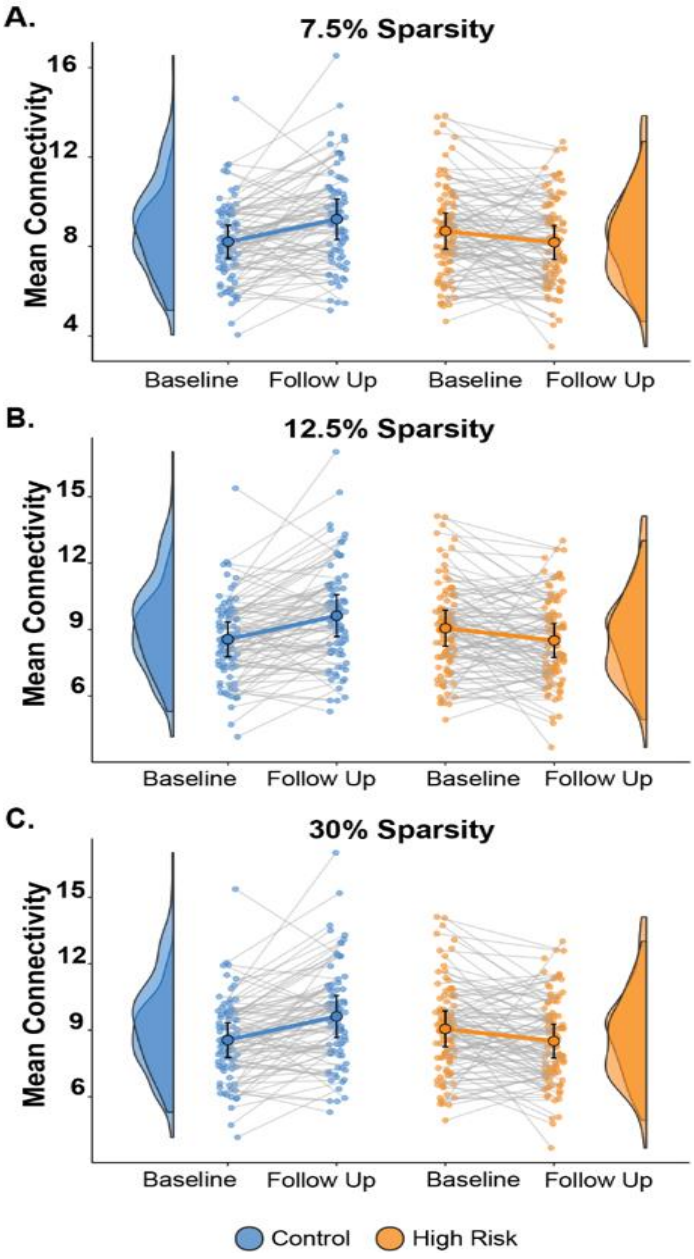
(A) Distribution of edges showing a significant decrease (T1>T2, left) and increase (T2>T1, right) in strength. Colours show coding of nodes according to functional distribution (30); abbreviations: SubCort: sub-cortical, Default: default mode network, VentAttn: Ventral Attention network, DorsAttn: Dorsal Attention network, SomMat: somato-motor network. (B) Corresponding mean connectivity strength of this network in all individuals at both time points and the corresponding group distributions. Circles and bars show mean +/- 95% C.I. Networks with a more conservative threshold of *t*=3.5 are provided in the main text (Figure 1).

FIGURE S4. Network of longitudinal changes in structural connectivity across both High Risk and Control groups, corrected for age at scanning (t = 3; p < 0.024, FWE corrected)



(A) This "age invariant" longitudinal contrast reveals a smaller, discrete network of edges than without age as a covariate (observed in Figure 1 and Figure S3) connecting bilateral cortical midline structures including bilateral middle and anterior cingulum, precuneus, and the caudate. (B) Distribution of network weights across participants showing an increase in strength shared across both groups. (C) Age-weight relationship of this network reveals the subtle effect of a 2-year shift in the age range of both cohorts. Colours show coding of nodes according to functional distribution; abbreviations: SubCort: sub-cortical, Default: default mode network, VentAttn: Ventral Attention network, DorsAttn: Dorsal Attention network, SomMat: somato-motor network.

FIGURE S5. Streamline count of time by group subnetworks (with NBS height threshold of t=3.0) across different connectome sparsity thresholds



Mean streamline counts calculated at network density levels of (A) 7.5%, (B) 10% (the default for the main text), and (C) 12.5%. Circles and bars show mean +/- 95% C.I.

FIGURE S6. Network connectivity strength in the high risk, bipolar and control individuals at baseline follow-up, with corresponding group regression slopes (and confidence intervals)

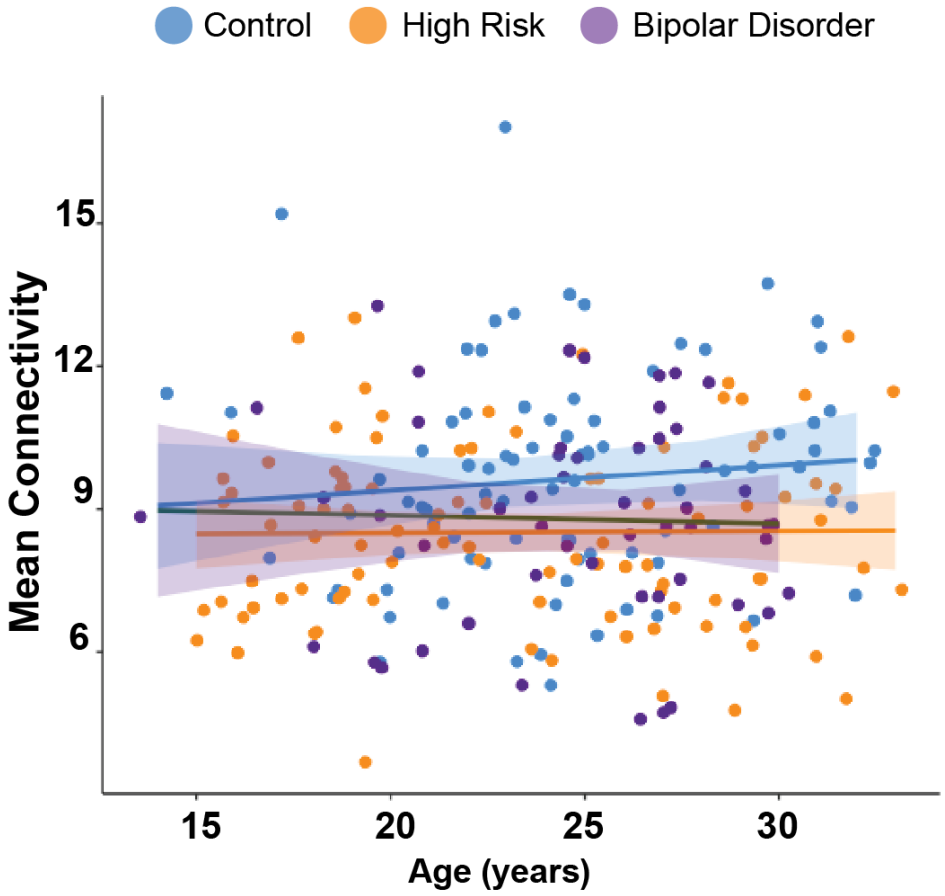
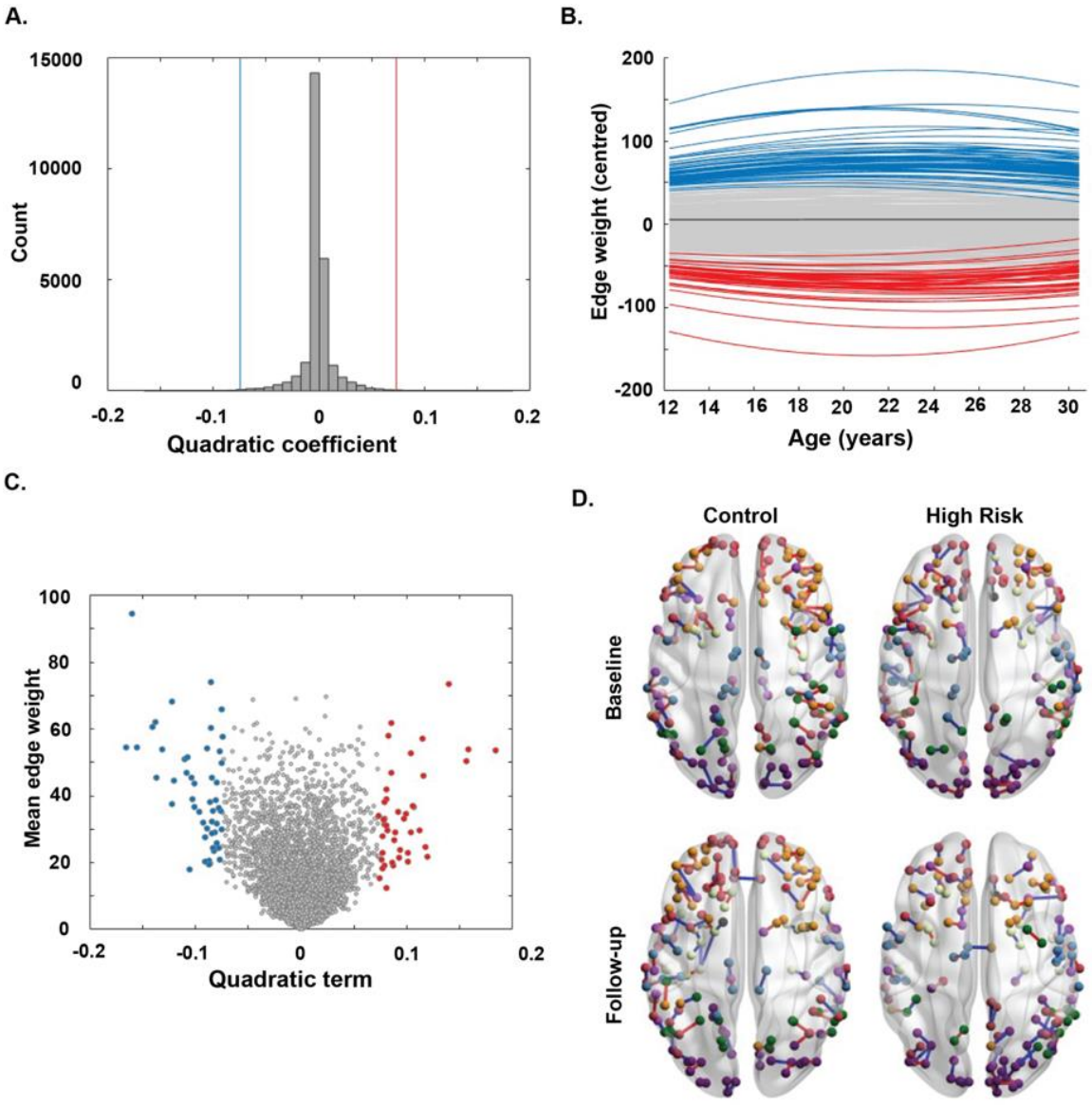


FIGURE S7. Non-linear age effects on connectivity



(A) Histogram of quadratic coefficient term for all edges reflecting that >99% sat within the (corrected) linear, null effect (red and blue lines show the Bonferroni-corrected cut-off). (B) Linear and quadratic trends in the mean (thick black) and individual (thin) edge weights as a function of age across both groups at baseline. Of all 26164 edges (grey), 86 edges possess a negative concavity (quadratic coefficient negative, blue) and 100 possess a positive concavity (red). The mean edge weight (black) sat within the confidence intervals for a purely linear effect. (C) The magnitude (absolute value) of the quadratic term was strongly correlated with the mean edge weight (corr=0.73, $p<0.001$). That is, quadratic edges tend to be strong, and short. (E-G). Highly comparable numbers of effects (numbers, lengths and weights of edges) were present at follow-up and within each group considered separately. Quadratic edges were consistent across time but not across group. The consistency of nonlinearity of edges was more consistent in the control (than the High Risk group (corr 0.60 versus 0.54, $p<0.05$, non-parametric bootstrap test).

REFERENCES

1. Nurnberger JI, McInnis M, Reich W, Kastelic E, Wilcox HC, Glowinski A, et al. A High Risk study of bipolar disorder: childhood clinical phenotypes as precursors of major mood disorders. Arch Gen Psychiatry. 2011;68(10):1012-20.
2. Roberts G, Perry A, Lord A, Frankland A, Leung V, Holmes-Preston E, et al. Structural dysconnectivity of key cognitive and emotional hubs in young people at high genetic risk for bipolar disorder. Mol Psychiatry. 2018;23:413-21.
3. Maxwell ME. Manual for the FIGS. Maryland: Clinical Neurogenetics Branch, National Institute of Mental Health; 1992.
4. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. J Am Acad Child Adolesc Psychiatry. 1997;36(7):980-8.
5. Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, et al. Diagnostic interview for genetic studies: rationale, unique features, and training. Arch Gen Psychiatry. 1994;51(11):849-59.
6. Leckman JF, Sholomskas D, Thompson WD, Belanger A, Weissman MM. Best estimate of lifetime psychiatric diagnosis: a methodological study. Arch Gen Psychiatry. 1982;39(8):879-83.
7. Kovacs M. Children's Depression Inventory (CDI). New York: Multi-Health Systems; 1992.
8. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134(4):382-9.
9. Young R, Biggs J, Ziegler V, Meyer D. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry. 1978;133(5):429-35.
10. Wechsler D. Wechsler Abbreviated Scale of Intelligence. New York: The Psychological Corporation; 1999.
11. Tournier J-D, Smith R, Raffelt D, Tabbara R, Dhollander T, Pietsch M, et al. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. NeuroImage. 2019;202:116-37.
12. Veraart J, Novikov DS, Christiaens D, Ades-Aron B, Sijbers J, Fieremans E. Denoising of diffusion MRI using random matrix theory. NeuroImage. 2016;142:394-406.
13. Andersson JL, Sotiropoulos SN. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. NeuroImage. 2016;125:1063-78.
14. 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. 2001;20(1):45-57.
15. Dhollander T, Raffelt D, Connelly A, editors. Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion MR data without a co-registered T1 image. ISMRM Workshop on Breaking the Barriers of Diffusion MRI; 2016; Lisbon, Portugal.
16. Dhollander T, Raffelt D, Connelly A, editors. Accuracy of response function estimation algorithms for 3-tissue spherical deconvolution of diverse quality diffusion MRI data. 26th International Society of Magnetic Resonance in Medicine; 2018; Paris, France.: 26th International Society of Magnetic Resonance in Medicine.
17. Jeurissen B, Tournier J-D, Dhollander T, Connelly A, Sijbers J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. NeuroImage. 2014;103:411-26.
18. Zalesky A, Fornito A, Harding IH, Cocchi L, Yücel M, Pantelis C, et al. Whole-brain anatomical networks: does the choice of nodes matter? NeuroImage. 2010;50(3):970-83.
19. Smith RE, Tournier J-D, Calamante F, Connelly A. SIFT: Spherical-deconvolution informed filtering of tractograms. NeuroImage. 2013;67:298-312.

20. Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, et al. Mapping the structural core of human cerebral cortex. PLoS Biol. 2008;6(7):1479-93.
21. Roberts JA, Perry A, Roberts G, Mitchell PB, Breakspear M. Consistency-based thresholding of the human connectome. NeuroImage. 2017;145:118-29.
22. Honey CJ, Sporns O, Cammoun L, Gigandet X, Thiran J-P, Meuli R, et al. Predicting human resting-state functional connectivity from structural connectivity. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(6):2035-40.
23. Pasqualetti F, Zampieri S, Bullo F. Controllability metrics, limitations and algorithms for complex networks. J IEEE Transactions on Control of Network Systems. 2014;1(1):40-52.
24. Gu S, Pasqualetti F, Cieslak M, Telesford QK, Alfred BY, Kahn AE, et al. Controllability of structural brain networks. Nature Communications. 2015;6(1):1-10.
25. Jeganathan J, Perry A, Bassett DS, Roberts G, Mitchell PB, Breakspear M. Fronto-limbic dysconnectivity leads to impaired brain network controllability in young people with bipolar disorder and those at high genetic risk. Neuroimage Clin. 2018;19:71-81.
26. Tang E, Pasqualetti F, Bassett DS, editors. The Control of Brain Activity Across Spatial and Temporal Scales. Bulletin of the American Physical Society; 2018; American Physical Society, Los Angeles.
27. Patankar SP, Kim JZ, Pasqualetti F, Bassett DS. Binary mesoscale architecture does not explain controllability of structural brain networks. J arXiv preprint arXiv:06514. 2020.
28. Wu-Yan E, Betzel RF, Tang E, Gu S, Pasqualetti F, Bassett DS. Benchmarking measures of network controllability on canonical graph models. J of Nonlinear Science. 2018:1-39.
29. Roberts JA, Perry A, Lord AR, Roberts G, Mitchell PB, Smith RE, et al. The contribution of geometry to the human connectome. NeuroImage. 2016;124(Part A):379-93.
30. Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol. 2011;106:1125-65.