Extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder

SHORT REPORT

Thomas Wolfers[a,b,c,*], Jaroslav Rokicki[a,b], Dag Alnæs[b], Ingrid Agartz[b,d,e,f], Seyed Mostafa Kia[c], Tobias Kaufmann[b], Mariam Zabihi[c], Torgeir Moberget[a,b], Ingrid Melle[b], Christian F Beckmann[c,g], Ole A Andreassen[b,d], Andre F Marquand[c,g,h,#], Lars T Westlye[a,b,d,#]

*corresponding author
[#] shared last authors

Affiliations:
[a] Department of Psychology, University of Oslo, Norway
[b] Norwegian Centre for Mental Disorders Research (NORMENT), Division of Mental Health and Addiction, University of Oslo & Oslo University Hospital, Oslo, Norway
[c] Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands
[d] KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Norway
[e] Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway
[f] Centre for psychiatric research, Department of Clinical Neuroscience, Stockholm, Sweden
[g] Department of Cognitive Neuroscience, Radboud University Medical Centre, Nijmegen, the Netherlands
[h] Department of Neuroimaging, Centre for Neuroimaging Sciences, Institute of Psychiatry, King's College London, London, UK

Contact details corresponding author:
Thomas Wolfers
NORMENT, University of Oslo, Oslo, Norway
e-mail: thomas.wolfers@psykologi.uio.no

Abstract: 169
Words in main text (max 1500): 1378
Methods: 855
Tables: 1
Figures: 3
Supplemental Materials: 4

ABSTRACT
Identifying brain processes involved in the risk and development of mental disorders is a major aim. We recently reported substantial inter-individual heterogeneity in brain structural aberrations among patients with schizophrenia and bipolar disorder. Estimating the normative range of voxel-based morphometry (VBM) data among healthy individuals using a Gaussian Process Regression (GPR) enables us to map individual deviations from the healthy range in unseen datasets. Here we aim to replicate our previous results in an independent sample of patients with schizophrenia (n=166), bipolar disorder (n=135) and healthy individuals (n=687). In line with previous findings, our results revealed robust group level differences between patients and healthy controls, yet only a small proportion of patients with schizophrenia or bipolar disorder exhibited extreme negative deviations from the norm in the same brain regions. These direct replications support that group level-differences in brain structure disguise considerable individual differences in brain aberrations, with important implications for the interpretation and generalization of group-level brain imaging findings to the individual with a mental disorder.

INTRODUCTION
Recently, the degree of inter-individual heterogeneity in brain structure was found to be considerably larger than previously anticipated for both schizophrenia and bipolar disorder (Wolfers et al., 2018). As expected, based on the substantial body of literature reporting results from case-control comparisons, patients with schizophrenia or bipolar disorder show evidence of group level deviations from a normative trajectory in brain structure. However, applying normative modeling (Marquand et al., 2016) to chart variation in brain anatomy across individual patients showed highly idiosyncratic patterns of deviation, suggesting that such group effects are inaccurate reflections of the brain aberrations found at the individual level (Wolfers et al., 2018). Of note, a similar high level of heterogeneity has recently also been observed in attention-deficit/ hyperactivity disorder (Wolfers et al., 2019) and autism spectrum disorder (Zabihi et al., 2019).

Given the extant literature on reproducible group-level differences in brain structure between cases and controls (Moberget et al., 2017; Van Erp et al., 2016), our initial findings of substantial heterogeneity within disorders demonstrated that moving beyond the study of group differences is highly beneficial to understand variability within clinical cohorts and may be required to make inferences at the level of the individual. Due to these important implications, we here report an attempt to replicate our initial findings in an independent sample following an identical analytical procedure as in our previous discovery study.

RESULTS
We included 166 patients with a schizophrenia diagnosis, 135 patients with a bipolar disorder diagnosis and 687 healthy individuals (Table 1). As the discovery sample we selected 256 healthy individuals, 163 patients with schizophrenia and 190 with bipolar disorder. All participants were recruited from the same population and catchment area in Oslo, Norway but there was no overlap between the discovery and replication samples. We analyzed our data using a normative modelling approach, which can be understood as a statistical model that maps the population variation in quantitative brain readout to demographic or behavioral variables (Marquand et al., 2016). We depict the spatial representation of the voxel-wise normative model, characterized by

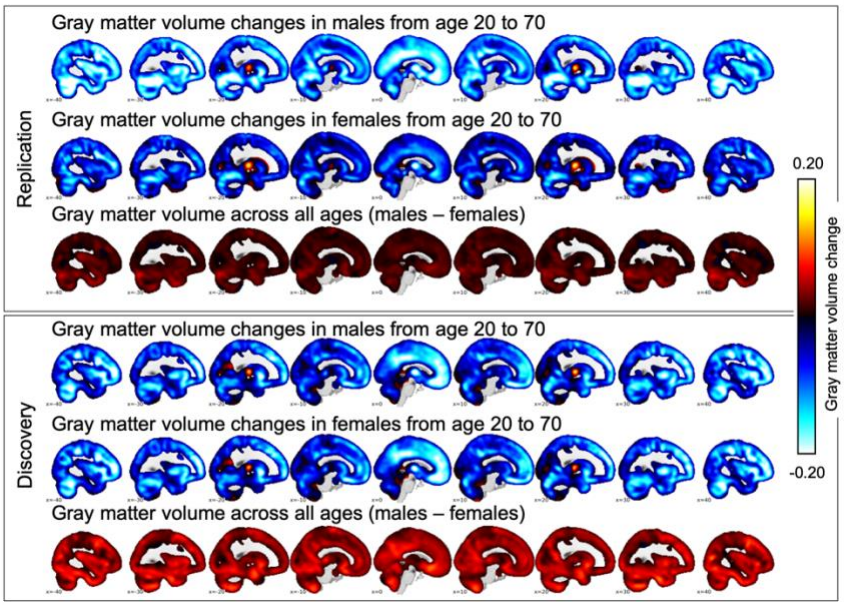


Figure 1: We depict the slope of a linear approximation of the normative model for males (first row in each panel) and females (second row in each panel) as well as the difference between males and females across the entire age range from 20-70 years (third row in each panel). In the lower panel we depict results based on the data reported in Wolfers et al. 2018, Jama Psychiatry. Note: These depictions are based on the forward model of the estimated normative models.

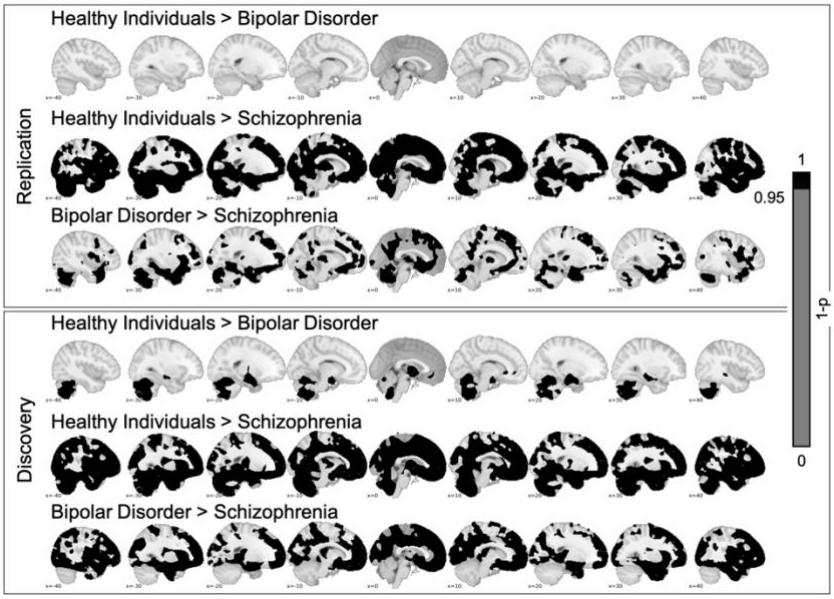


Figure 2: We depict the contrast between healthy individuals, bipolar disorder and schizophrenia. In the lower panel we depict results based on the data reported in Wolfers et al. 2018, Jama Psychiatry. For the PALM derived mean Z-scores see supplementary figure 2. Note: We report one subtracted by multiple comparison corrected p-values.

widespread gray matter decreases from age 20 to 70, with most pronounced age-differences in frontal areas (Figure 1).

Group comparisons

Figure 2 shows the result from pairwise group comparisons, corrected for multiple comparisons using permutation testing. In gray matter, patients with schizophrenia showed stronger mean negative deviations than healthy individuals in frontal, temporal, and cerebellar regions; mean deviations were also more negative than in patients with bipolar disorder and were localized primarily in fontal brain regions. In a sensitivity analysis we matched the age distribution for the controls of the discovery and replication study, with highly consistent findings (Supplementary Figure 1).

Spatial distribution of extreme normative deviations

In line with our previous study, patients with schizophrenia showed a higher percentage of extreme negative deviations across the brain (0.46 ± 0.97% of all voxels) compared to healthy individuals (0.15 ± 0.46%, Mann-Whitney = 36674,

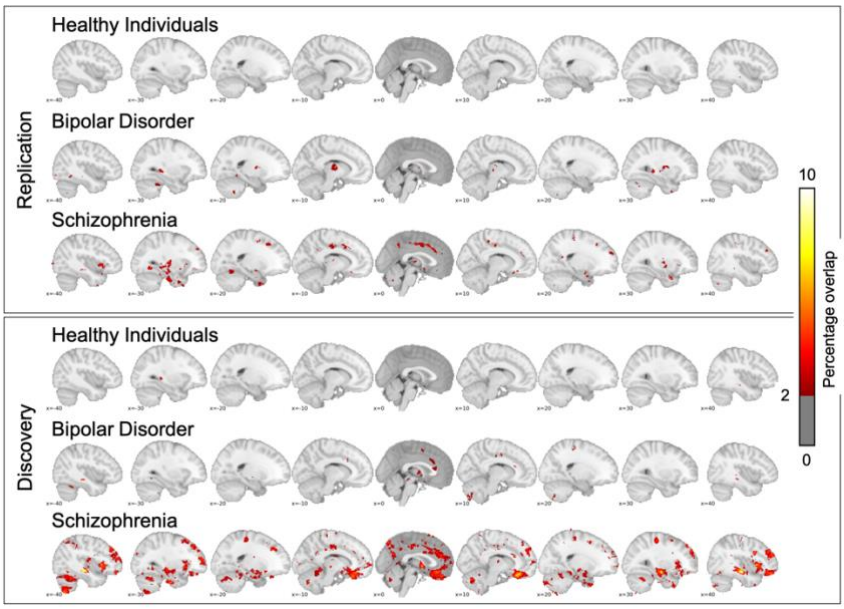


Figure 3: We show extreme negative deviations for healthy individuals, bipolar disorder and schizophrenia. In the lower panel we depict results based on the data reported in Wolfers et al. 2018, Jama Psychiatry. We show that the overlap across studies is comparable with only a few brain regions showing overlap in more than 2% of the individuals. Interestingly, the overlap while consistent is even smaller in the replication study. Note: Extreme negative deviations are defined as Z < -2.6 at the individual level.

p<.001) and individuals with bipolar disorder (0.15 ± 0.34%, Mann-Whitney = 7655, p<.001, Supplementary Table 1). The percentage of extreme positive deviations across participants and groups showed that healthy individuals (1.18 ± 2.23%) differed from patients with schizophrenia (0.81 ± 1.73%; Mann-Whitney = 65349, p <.01) but not from individuals with bipolar disorder *(0.92 ± 1.53%).* This is an exact replication of the previous study (Supplementary Table 1).

Extreme negative deviations in people with schizophrenia were most pronounced in temporal, medial frontal and posterior cingulate regions (Figure 3). In patients with bipolar disorder the overlap was strongest in the thalamus. In line with our previous findings the overlap of extreme negative (Figure 3) and positive deviations (Supplemental Figure 2) is sparse across individuals with the same diagnosis, with peak voxels showing extreme negative overlap is 2% healthy individuals, 4.44% for individuals with bipolar disorder and 5.42% for schizophrenia.

DISCUSSION

Advanced brain imaging technology has allowed for probing the brain functional and structural correlates of complex human traits and mental disorders. While group-level normative deviations in brain structure in patients with a diagnosis of schizophrenia and bipolar disorder are robust and replicable (Figure 2) we observe substantial inter-individual differences in the neuroanatomical distribution of extreme deviations at the individual level (Figure 3, Supplementary Figure 2). These findings replicate our previous study (Wolfers et al., 2018) and are largely in line with evidence of large heterogeneity across mental disorders (Wolfers et al., 2019; Zabihi et al., 2019).

Our results confirm that MRI-based brain structural aberrations in patients with severe mental disorders are highly heterogeneous in terms of their neuroanatomical distribution. These findings are in line with recent evidence of substantial brain structural heterogeneity in patients with schizophrenia (Alnæs et al., 2019) and also comply with accumulating evidence from psychiatric genetics strongly suggesting that mental illnesses are complex and heterogeneous disorders associated with a large number of genetic variants as well as environmental risk factors

(Sullivan and Geschwind, 2019). Along with documented clinical heterogeneity (Insel, 2009) and large interindividual variability in treatment response and outcome (Kapur et al., 2012), our successful replication of considerable neuroanatomical heterogeneity supports the need for statistical approaches that allow for inferences at the level of the individual. Characterizing the magnitude and distribution of brain aberrations in individual patients is key for identifying neuronal correlates of specific symptoms across diagnostic categories and would represent an important step towards increasing the utility of brain imaging in a clinical context.

While the present findings are robust, it must be considered that other data modalities beyond those provided by structural brain imaging may be more able to capture any common pathophysiological processes in patients with schizophrenia or bipolar disorder. Thus, we may observe larger overlaps across individuals with the same mental disorders in other data domains, such as those measuring brain function, cognition or specific behaviors, or relevant biological assays. While this possibility cannot be ruled out the present results indicate that multiple pathophysiological processes and pathways are at play, which is also supported by the large number of identified genetic variants (Ripke et al., 2014; Smoller et al., 2013; Stahl et al., 2019).

Over the last decades it has become increasingly apparent that replication attempts in psychology, psychiatry, neuroscience and related fields frequently fail (Avinun et al., 2018; Dinga et al., 2019; Eklund et al., 2016; Hong et al., 2019; Ioannidis, 2005; Open Science Collaboration*, 2015; Tackett et al., 2019)(Baker and Penny, 2016), which has fueled initiatives promoting reproducible science (Munafò et al., 2017; Poldrack et al., 2017; Schooler, 2014). The neuroimaging field is no exception, and lack of reproducibility in brain imaging studies have been attributed to the high researcher degree of freedom in terms of the many and sometimes arbitrary analytical choices (Eklund et al., 2016). Here we strictly adhered to the analytical protocols as specified in our original study (Wolfers et al., 2018). The entire analytical pipeline is made publicly available to ease replication by independent researchers and to allow for application to different cohorts and disorders.

While we believe that these analytical protocols are appropriate for testing the reproducibility of our original report, the approaches will be improved in future studies and are under continuous development. For instance, in the current study as well as the original study all data was collected on one MRI scanner, ruling out scanner effects as a possible source of confound. As we aim to build large population based normative models and apply those to different cohorts improving methods that deal with noise introduced through multiple sites is an important research priority (Marquand et al., 2019). Moving forward, we will scale up this work towards larger samples covering a wider age range, including neurodevelopment, incorporate different modalities and levels of information, including genetics and biomarkers, and link different experimental designs to the normative modeling framework.

With low reproducibility rates across various scientific disciplines (Baker and Penny, 2016) building confidence through replication is critical. Here, by replicating the findings in our initial report in an independent sample we show that group differences in brain structure between healthy controls and patients with schizophrenia or bipolar disorder mask large individual differences in brain aberrations, with important implications for the generalization of group-level brain MRI findings to the level of the individual patient.

METHODS

Participants

Table 1 summarizes the demographic and clinical information of the current sample and the sample used in the original publication (Wolfers et al., 2018). All participants were recruited as part of the Thematically Organized Psychosis (TOP) study, approved by the Regional Committee for Medical Research Ethics and the Norwegian Data Inspectorate (Doan et al., 2017). Patients were recruited from in- and out-patient clinics in the Oslo area, understood and spoke a Scandinavian language, had no history of severe head trauma, and had an IQ above 70. Patients were assessed by trained physicians or clinical psychologists. Psychiatric diagnosis was established using the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID). Healthy individuals were randomly sampled from national registries and neither they nor their relatives had a psychiatric or alcohol/substance use disorder or cannabis use during the last 3 months. Written informed consent was obtained from all participants.

Table 1: Demographics Table

	Replication			Discovery		
Demographics	**Healthy individuals**	**Bipolar disorder**	**Schizophrenia**	**Healthy individuals**	**Bipolar disorder**	**Schizophrenia**
N	687	135	166	256	190	163
Male (%)	54.87%	36.29%	51.80%	54.70%	41.80%	64.40%
Age (mean ± std)	41 ± 16.0	32 ± 0.8	30 ± 9.6	34 ± 9.5	34 ± 11.3	31 ± 8.7
Years of education (mean ± std)	14.4 ± 2.4	15.0 ± 2.4	13.8 ± 3.0	14.0 ± 2.3	13.6 ± 2.3	12.9 ± 2.6
Symptom scores*						
PANSS negative (mean ± std)	na	10.2 ± 3.2	14.5 ± 5.3	na	10.1 ± 3.5	15.9 ± 6.4
PANSS positive (mean ± std)	na	9.1 ± 2.7	13.9 ± 4.7	na	10.0 ± 3.5	15.1 ± 5.5
PANSS total (mean ± std)	na	44.4 ± 8.3	59 ± 5.8	na	45.5 ± 10.1	63.1 ± 16.9

*Symptom scores have been assessed using PANSS, which is a standard clinical instrument for the quantification of positive and negative psychotic symptoms.

MRI acquisition

Structural scans were obtained on 3 Tesla GE 750 Discovery scanner at Oslo University Hospital using a standard 32-channel head coil. T1-weighted images were acquired using a BRAVO sequence with the following parameters: repetition time (TR) = 8.16ms, echo time (TE) = 3.18ms, inversion time (TI) = 400ms, field of view (FOV) = 256 mm, flip angle (FA) = 12°, matrix = 188 × 256, voxel size = 1 × 1 × 1 mm, 256 sagittal slices.

Estimation of gray matter volume

In the same way as in our previous study, raw T1-weighted MRI volumes were processed using the computational analysis toolbox version 12 (CAT12; http://www.neuro.uni-jena.de/software/), based on statistical parametric mapping version 12 (SPM12). Images were segmented, normalized, and bias-field-corrected using VBM-SPM12 (http://www.fil.ion.ucl.ac.uk/spm, London, UK) (Ashburner and Friston, 2003, 2000), yielding images containing gray and white

matter segments. Prior to the estimation of the normative models, all gray and white matter volumes were smoothed with an 8 mm FWHM Gaussian smoothing kernel and we restricted our analyses to voxels included in the gray matter mask constructed for the previous study.

Normative modeling

As in our previous article, we estimated the normative model using Gaussian Process Regression (GPR) to predict VBM based regional gray matter volumes across the brain from age and sex. To avoid overfitting of the normative models, it is crucial to estimate predictive performance out of sample. Therefore, we estimated the normative range for this model in healthy individuals under 10-fold cross-validation, and then applied one model across all healthy individuals to patients with schizophrenia and bipolar disorder. GPR yields coherent measures of predictive confidence in addition to point estimates. This is important in normative modelling as we need this uncertainty measure to quantify the deviation of each patient from the group mean at each brain locus. Thus, we are able to statistically quantify deviations from the normative model with regional specificity, by computing a Z-score for each voxel reflecting the difference between the predicted and the observed gray matter volume normalized by the uncertainty of the prediction (Marquand et al., 2016).

In line with our previous article, we thresholded the individual normative probability maps at $p<.005$ (i.e. $|Z|>2.6$) and extreme positive and extreme negative deviations from the normative model were defined based on this threshold. All extreme deviations were combined into scores representing the percentage of extreme positively and negatively deviating voxels for each participant, relative to the total number of voxels in the brain mask. We tested for associations between diagnosis and those scores using a non-parametric test corrected for multiple comparisons using the Bonferroni-Holm method (Holm, 1979). To assess the spatial extent of those extreme deviations, we created individualized maps and calculated the voxel-wise overlap between individuals from the same groups. In the main text we report this overlap for healthy individuals, and patients diagnosed with schizophrenia and bipolar disorder. All analyses were performed in python3.6 (www.python.org) and scripts are available on GitHub (https://github.com/RindKind/). Also, in line with our previous article, we fed the normative probability maps into PALM (Winkler et al., 2015) to test for mean differences between groups by means of a general linear model framework and permutation-based inference.

ACKNOWLEDGEMENTS AND FUNDING

The study is supported by the Research Council of Norway (223273, 249795, 298646, 300768, 276082), the South-Eastern Norway Regional Health Authority (2014097, 2015073, 2016083, 2019101), a Wellcome Trust Innovator award ('BRAINCHART', 215698/Z/19/Z), and the European Research Council under the European Union's Horizon 2020 research and Innovation program (ERC StG Grant 802998 and Grant 847776). This work was performed on the TSD (Tjenester for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT) with resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway. TW gratefully acknowledges the Niels Stensen Fellowship for supporting this work.

COMPEPTING INTERESTS

O.A.A. is a consultant to HealthLytix and received speaker's honorarium from Lundbeck. C.B. is shareholder of and director of SBG Neuro. The other authors declare that they have no competing interests.

REFERENCES

Alnæs D, Kaufmann T, Van Der Meer D, Córdova-Palomera A, Rokicki J, Moberget T, Bettella F, Agartz I, Barch DM, Bertolino A, Brandt CL, Cervenka S, Djurovic S, Doan NT, Eisenacher S, Fatouros-Bergman H, Flyckt L, Di Giorgio A, Haatveit B, Jönsson EG, Kirsch P, Lund MJ, Meyer-Lindenberg A, Pergola G, Schwarz E, Smeland OB, Quarto T, Zink M, Andreassen OA, Westlye LT. 2019. Brain Heterogeneity in Schizophrenia and Its Association with Polygenic Risk. *JAMA Psychiatry* **76**:739–748. doi:10.1001/jamapsychiatry.2019.0257

Ashburner J, Friston KJ. 2003. Spatial normalization using basis functions{H}uman {B}rain {F}unction. pp. 1–26.

Ashburner J, Friston KJ. 2000. Voxel-based morphometry - The methods. *Neuroimage* **11**:805–821. doi:10.1006/nimg.2000.0582

Avinun R, Nevo A, Knodt AR, Elliott ML, Hariri AR. 2018. Replication in Imaging Genetics: The Case of Threat-Related Amygdala Reactivity. *Biol Psychiatry* **84**:148–159. doi:10.1016/j.biopsych.2017.11.010

Baker M, Penny D. 2016. Is there a reproducibility crisis? *Nature* **533**:3–5.

Dinga R, Schmaal L, Penninx BWJH, Jose M, Tol V, Veltman DJ, Velzen L Van, Mennes M, Wee NJA Van Der, Marquand AF. 2019. Evaluating the evidence for biotypes of depression: Methodological replication and extension of Drysdale et al. (2017). *NeuroImage Clin* **22**:101796. doi:10.1016/j.nicl.2019.101796

Doan NT, Kaufmann T, Bettella F, Jørgensen KN, Brandt CL, Moberget T, Alnæs D, Douaud G, Duff E, Djurovic S, Melle I, Ueland T, Agartz I, Andreassen OA, Westlye LT. 2017. Distinct multivariate brain morphological patterns and their added predictive value with cognitive and polygenic risk scores in mental disorders. *NeuroImage Clin* **15**:719–731. doi:10.1016/j.nicl.2017.06.014

Eklund A, Nichols TE, Knutsson H. 2016. Cluster failure : Why fMRI inferences for spatial extent have inflated false-positive rates. *PNAS* **113**:7900–7905. doi:10.1073/pnas.1602413113

Holm S. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scand J Stat* **6**:65–70. doi:10.2307/4615733

Hong YW, Yoo Y, Han J, Wager TD, Woo CW. 2019. False-positive neuroimaging: Undisclosed flexibility in testing spatial hypotheses allows presenting anything as a replicated finding. *Neuroimage* **195**:384–395. doi:10.1016/j.neuroimage.2019.03.070

Insel TR. 2009. Disruptive insights in psychiatry: Transforming a clinical discipline. *J Clin Invest* **119**:700–705. doi:10.1172/JCI38832

Ioannidis JPA. 2005. Contradicted and Initially Stronger Effects in Highly Cited Clinical Research. *JAMA - J Am Med Assoc* **294**:218–227.

Kapur S, Phillips AG, Insel TR. 2012. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it. *Mol Psychiatry* **17**:1174–1179. doi:10.1038/mp.2012.105

Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF. 2019. Conceptualizing

mental disorders as deviations from normative functioning. *Mol Psychiatry*. doi:10.1038/s41380-019-0441-1

Marquand AF, Rezek I, Buitelaar J, Beckmann CF. 2016. Understanding heterogeneity in clinical cohorts using normative models: beyond case control studies. *Biol Psychiatry* **80**:552–561. doi:10.1016/j.biopsych.2015.12.023

Moberget T, Doan NT, Alnæs D, Kaufmann T, Córdova-Palomera A, Lagerberg T V., Diedrichsen J, Schwarz E, Zink M, Eisenacher S, Kirsch P, Jönsson EG, Fatouros-Bergman H, Flyckt L, Pergola G, Quarto T, Bertolino A, Barch D, Meyer-Lindenberg A, Agartz I, Andreassen OA, Westlye LT. 2017. Cerebellar volume and cerebellocerebral structural covariance in schizophrenia: a multisite mega-analysis of 983 patients and 1349 healthy controls. *Mol Psychiatry* 1–9. doi:10.1038/mp.2017.106

Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie N, Simonsohn U, Wagenmakers E. 2017. A manifesto for reproducible science Marcus. *Nat Hum Behav* **1**:1–9. doi:10.1038/s41562-016-0021

Open Science Collaboration*. 2015. Estimating the reproducibility of psychological science. *Science (80-)* **349**. doi:10.1126/science.aac4716

Poldrack RA, Baker CI, Durnez J, Gorgolewski KJ, Matthews PM, Munaf MR, Nichols TE, Poline J, Vul E, Yarkoni T. 2017. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat Rev* **18**. doi:10.1038/nrn.2016.167

Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, Lee P, Bulik-Sullivan B, Collier DA, Huang H, Pers TH, Agartz I, Agerbo E, Albus M, Alexander M, Amin F, Bacanu SA, Begemann M, Belliveau RA, Bene J, Bergen SE, Bevilacqua E, Bigdeli TB, Black DW, Bruggeman R, Buccola NG, Buckner RL, Byerley W, Cahn W, Cai G, Campion D, Cantor RM, Carr VJ, Carrera N, Catts S V., Chambert KD, Chan RCK, Chen RYL, Chen EYH, Cheng W, Cheung EFC, Chong SA, Cloninger CR, Cohen D, Cohen N, Cormican P, Craddock N, Crowley JJ, Curtis D, Davidson M, Davis KL, Degenhardt F, Del Favero J, Demontis D, Dikeos D, Dinan T, Djurovic S, Donohoe G, Drapeau E, Duan J, Dudbridge F, Durmishi N, Eichhammer P, Eriksson J, Escott-Price V, Essioux L, Fanous AH, Farrell MS, Frank J, Franke L, Freedman R, Freimer NB, Friedl M, Friedman JI, Fromer M, Genovese G, Georgieva L, Giegling I, Giusti-Rodríguez P, Godard S, Goldstein JI, Golimbet V, Gopal S, Gratten J, De Haan L, Hammer C, Hamshere ML, Hansen M, Hansen T, Haroutunian V, Hartmann AM, Henskens FA, Herms S, Hirschhorn JN, Hoffmann P, Hofman A, Hollegaard M V., Hougaard DM, Ikeda M, Joa I, Julià A, Kahn RS, Kalaydjieva L, Karachanak-Yankova S, Karjalainen J, Kavanagh D, Keller MC, Kennedy JL, Khrunin A, Kim Y, Klovins J, Knowles JA, Konte B, Kucinskas V, Kucinskiene ZA, Kuzelova-Ptackova H, Kähler AK, Laurent C, Keong JLC, Lee SH, Legge SE, Lerer B, Li M, Li T, Liang KY, Lieberman J, Limborska S, Loughland CM, Lubinski J, Lönnqvist J, Macek M, Magnusson PKE, Maher BS, Maier W, Mallet J, Marsal S, Mattheisen M, Mattingsdal M, McCarley RW, McDonald C, McIntosh AM, Meier S, Meijer CJ, Melegh B, Melle I, Mesholam-Gately RI, Metspalu A, Michie PT, Milani L, Milanova V, Mokrab Y, Morris DW, Mors O, Murphy KC, Murray RM, Myin-Germeys I, Müller-Myhsok B, Nelis M, Nenadic I, Nertney DA, Nestadt G, Nicodemus KK, Nikitina-Zake L, Nisenbaum L, Nordin A, O'Callaghan E, O'Dushlaine C, O'Neill FA, Oh SY, Olincy A, Olsen L, Van Os J, Pantelis C, Papadimitriou GN, Papiol S, Parkhomenko E, Pato MT, Paunio T, Pejovic-Milovancevic M, Perkins DO, Pietiläinen O, Pimm J, Pocklington AJ, Powell J, Price A, Pulver AE, Purcell SM,

Quested D, Rasmussen HB, Reichenberg A, Reimers MA, Richards AL, Roffman JL, Roussos P, Ruderfer DM, Salomaa V, Sanders AR, Schall U, Schubert CR, Schulze TG, Schwab SG, Scolnick EM, Scott RJ, Seidman LJ, Shi J, Sigurdsson E, Silagadze T, Silverman JM, Sim K, Slominsky P, Smoller JW, So HC, Spencer CCA, Stahl EA, Stefansson H, Steinberg S, Stogmann E, Straub RE, Strengman E, Strohmaier J, Stroup TS, Subramaniam M, Suvisaari J, Svrakic DM, Szatkiewicz JP, Söderman E, Thirumalai S, Toncheva D, Tosato S, Veijola J, Waddington J, Walsh D, Wang D, Wang Q, Webb BT, Weiser M, Wildenauer DB, Williams NM, Williams S, Witt SH, Wolen AR, Wong EHM, Wormley BK, Xi HS, Zai CC, Zheng X, Zimprich F, Wray NR, Stefansson K, Visscher PM, Adolfsson R, Andreassen OA, Blackwood DHR, Bramon E, Buxbaum JD, Børglum AD, Cichon S, Darvasi A, Domenici E, Ehrenreich H, Esko T, Gejman P V., Gill M, Gurling H, Hultman CM, Iwata N, Jablensky A V., Jönsson EG, Kendler KS, Kirov G, Knight J, Lencz T, Levinson DF, Li QS, Liu J, Malhotra AK, McCarroll SA, McQuillin A, Moran JL, Mortensen PB, Mowry BJ, Nöthen MM, Ophoff RA, Owen MJ, Palotie A, Pato CN, Petryshen TL, Posthuma D, Rietschel M, Riley BP, Rujescu D, Sham PC, Sklar P, St Clair D, Weinberger DR, Wendland JR, Werge T, Daly MJ, Sullivan PF, O'Donovan MC. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**:421–427. doi:10.1038/nature13595

Schooler JW. 2014. Metascience could rescue the 'replication crisis.' *Nature* **515**:9.

Smoller JW, Kendler K, Craddock N, Lee PH, Neale BM, Nurnberger JN, Ripke S, Santangelo S, Sullivan PS, Neale BN, Purcell S, Anney R, Buitelaar J, Fanous A, Faraone SF, Hoogendijk W, Lesch KP, Levinson DL, Perlis RP, Rietschel M, Riley B, Sonuga-Barke E, Schachar R, Schulze TS, Thapar A, Kendler KK, Smoller JS, Neale M, Perlis R, Bender P, Cichon S, Daly MD, Kelsoe J, Lehner T, Levinson D, O'Donovan Mick, Gejman P, Sebat J, Sklar P, Daly M, Devlin B, Sullivan P, O'Donovan Michael. 2013. Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *Lancet* **381**:1371–1379. doi:10.1016/S0140-6736(12)62129-1

Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, Mattheisen M, Wang Y, Coleman JRI, Gaspar HA, de Leeuw CA, Steinberg S, Pavlides JMW, Trzaskowski M, Byrne EM, Pers TH, Holmans PA, Richards AL, Abbott L, Agerbo E, Akil H, Albani D, Alliey-Rodriguez N, Als TD, Anjorin A, Antilla V, Awasthi S, Badner JA, Bækvad-Hansen M, Barchas JD, Bass N, Bauer M, Belliveau R, Bergen SE, Pedersen CB, Bøen E, Boks MP, Boocock J, Budde M, Bunney W, Burmeister M, Bybjerg-Grauholm J, Byerley W, Casas M, Cerrato F, Cervantes P, Chambert K, Charney AW, Chen D, Churchhouse C, Clarke TK, Coryell W, Craig DW, Cruceanu C, Curtis D, Czerski PM, Dale AM, de Jong S, Degenhardt F, Del-Favero J, DePaulo JR, Djurovic S, Dobbyn AL, Dumont A, Elvsåshagen T, Escott-Price V, Fan CC, Fischer SB, Flickinger M, Foroud TM, Forty L, Frank J, Fraser C, Freimer NB, Frisén L, Gade K, Gage D, Garnham J, Giambartolomei C, Pedersen MG, Goldstein J, Gordon SD, Gordon-Smith K, Green EK, Green MJ, Greenwood TA, Grove J, Guan W, Guzman-Parra J, Hamshere ML, Hautzinger M, Heilbronner U, Herms S, Hipolito M, Hoffmann P, Holland D, Huckins L, Jamain S, Johnson JS, Juréus A, Kandaswamy R, Karlsson R, Kennedy JL, Kittel-Schneider S, Knowles JA, Kogevinas M, Koller AC, Kupka R, Lavebratt C, Lawrence J, Lawson WB, Leber M, Lee PH, Levy SE, Li JZ, Liu C, Lucae S, Maaser A, MacIntyre DJ, Mahon PB, Maier W, Martinsson L, McCarroll S, McGuffin P, McInnis MG, McKay JD, Medeiros H, Medland SE, Meng F, Milani L, Montgomery GW, Morris DW, Mühleisen TW, Mullins N, Nguyen H,

Nievergelt CM, Adolfsson AN, Nwulia EA, O'Donovan C, Loohuis LMO, Ori APS, Oruc L, Ösby U, Perlis RH, Perry A, Pfennig A, Potash JB, Purcell SM, Regeer EJ, Reif A, Reinbold CS, Rice JP, Rivas F, Rivera M, Roussos P, Ruderfer DM, Ryu E, Sánchez-Mora C, Schatzberg AF, Scheftner WA, Schork NJ, Shannon Weickert C, Shehktman T, Shilling PD, Sigurdsson E, Slaney C, Smeland OB, Sobell JL, Søholm Hansen C, Spijker AT, St Clair D, Steffens M, Strauss JS, Streit F, Strohmaier J, Szelinger S, Thompson RC, Thorgeirsson TE, Treutlein J, Vedder H, Wang W, Watson SJ, Weickert TW, Witt SH, Xi S, Xu W, Young AH, Zandi P, Zhang P, Zöllner S, Adolfsson R, Agartz I, Alda M, Backlund L, Baune BT, Bellivier F, Berrettini WH, Biernacka JM, Blackwood DHR, Boehnke M, Børglum AD, Corvin A, Craddock N, Daly MJ, Dannlowski U, Esko T, Etain B, Frye M, Fullerton JM, Gershon ES, Gill M, Goes F, Grigoroiu-Serbanescu M, Hauser J, Hougaard DM, Hultman CM, Jones I, Jones LA, Kahn RS, Kirov G, Landén M, Leboyer M, Lewis CM, Li QS, Lissowska J, Martin NG, Mayoral F, McElroy SL, McIntosh AM, McMahon FJ, Melle I, Metspalu A, Mitchell PB, Morken G, Mors O, Mortensen PB, Müller-Myhsok B, Myers RM, Neale BM, Nimgaonkar V, Nordentoft M, Nöthen MM, O'Donovan MC, Oedegaard KJ, Owen MJ, Paciga SA, Pato C, Pato MT, Posthuma D, Ramos-Quiroga JA, Ribasés M, Rietschel M, Rouleau GA, Schalling M, Schofield PR, Schulze TG, Serretti A, Smoller JW, Stefansson H, Stefansson K, Stordal E, Sullivan PF, Turecki G, Vaaler AE, Vieta E, Vincent JB, Werge T, Nurnberger JI, Wray NR, Di Florio A, Edenberg HJ, Cichon S, Ophoff RA, Scott LJ, Andreassen OA, Kelsoe J, Sklar P. 2019. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* **51**:793–803. doi:10.1038/s41588-019-0397-8

Sullivan PF, Geschwind DH. 2019. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* **177**:162–183. doi:10.1016/j.cell.2019.01.015

Tackett JL, Brandes CM, King KM, Markon KE. 2019. Psychology's replication crisis and clinical psychological science. *Annu Rev Clinial Psychol* **15**:579–604.

Van Erp TGM, Hibar DP, Rasmussen JM, Glahn DC, Pearlson GD, Andreassen OA, Agartz I, Westlye LT, Haukvik UK, Dale AM, Melle I, Hartberg CB, Gruber O, Kraemer B, Zilles D, Donohoe G, Kelly S, McDonald C, Morris DW, Cannon DM, Corvin A, Machielsen MWJ, Koenders L, De Haan L, Veltman DJ, Satterthwaite TD, Wolf DH, Gur RCECE, Gur RCECE, Potkin SG, Mathalon DH, Mueller BA, Preda A, Macciardi F, Ehrlich S, Walton E, Hass J, Calhoun VD, Bockholt HJ, Sponheim SR, Shoemaker JM, Van Haren NEM, Pol HEH, Ophoff RA, Kahn RS, Roiz-Santiaez R, Crespo-Facorro B, Wang L, Alpert KI, Jönsson EG, Dimitrova R, Bois C, Whalley HC, McIntosh AM, Lawrie SM, Hashimoto R, Thompson PM, Turner JA. 2016. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol Psychiatry* **21**:547–553. doi:10.1038/mp.2015.63

Winkler AM, Webster MA, Vidaurre D, Nichols TE, Smith SM. 2015. Multi-level block permutation. *Neuroimage* **123**:253–268. doi:10.1016/j.neuroimage.2015.05.092

Wolfers T, Beckmann CF, Hoogman M, Buitelaar JK, Franke B, Marquand AF. 2019. Individual differences v. The average patient: Mapping the heterogeneity in ADHD using normative models. *Psychol Med* 1–10. doi:10.1017/S0033291719000084

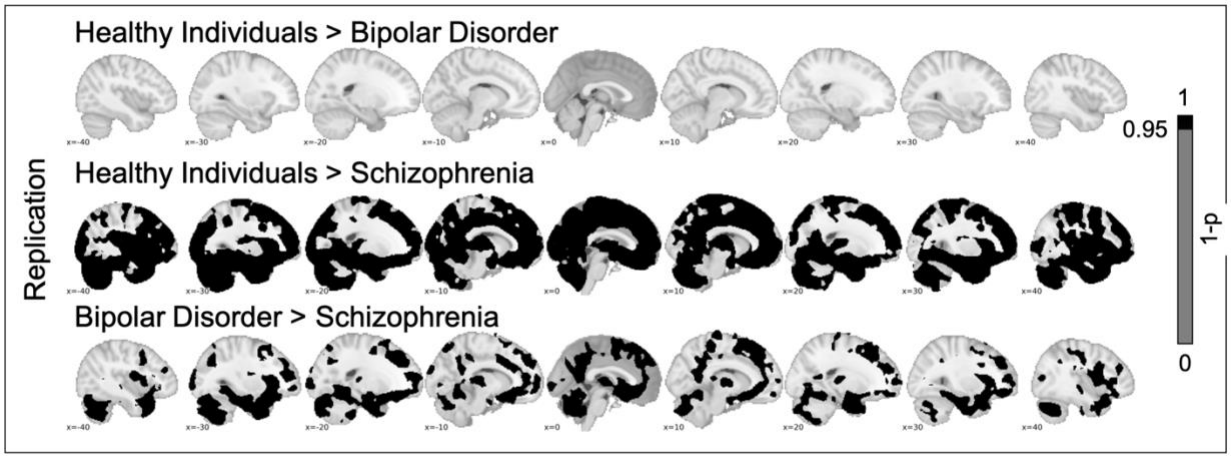
Wolfers T, Doan NT, Kaufmann T, Alnæs D, Moberget T, Agartz I, Buitelaar JK, Ueland T, Melle I, Franke B, Andreassen OA, Beckmann CF, Westlye LT, Marquand AF. 2018. Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models.

JAMA Psychiatry **11**:1146–1155. doi:10.1001/JAMAPSYCHIATRY.2018.2467
Zabihi M, Oldehinkel M, Wolfers T, Frouin V, Goyard D, Loth E, Charman T, Tillmann J, Banaschewski T, Dumas G, Holt R, Baron-Cohen S, Durston S, Bölte S, Murphy D, Ecker C, Buitelaar JK, Beckmann CF, Marquand AF. 2019. Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models. *Biol Psychiatry Cogn Neurosci Neuroimaging* 1–12. doi:10.1016/j.bpsc.2018.11.013

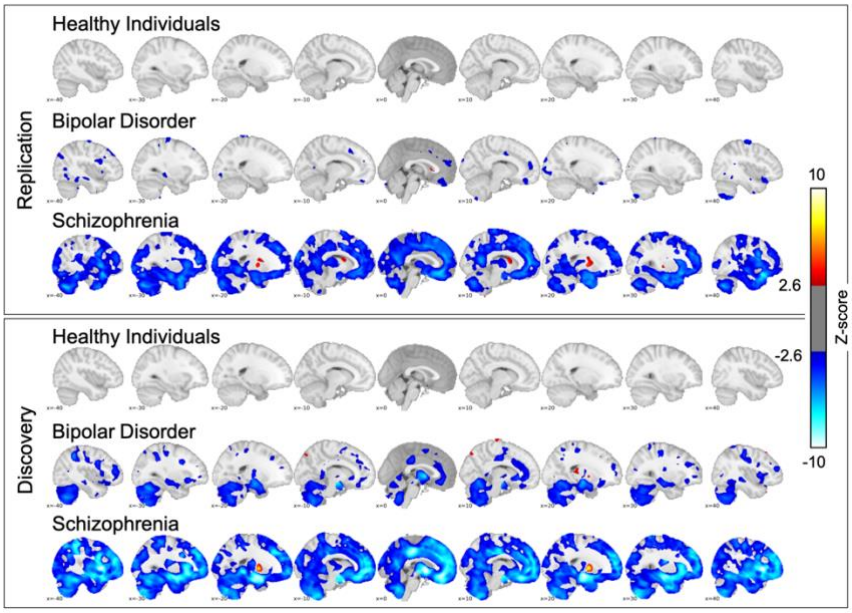
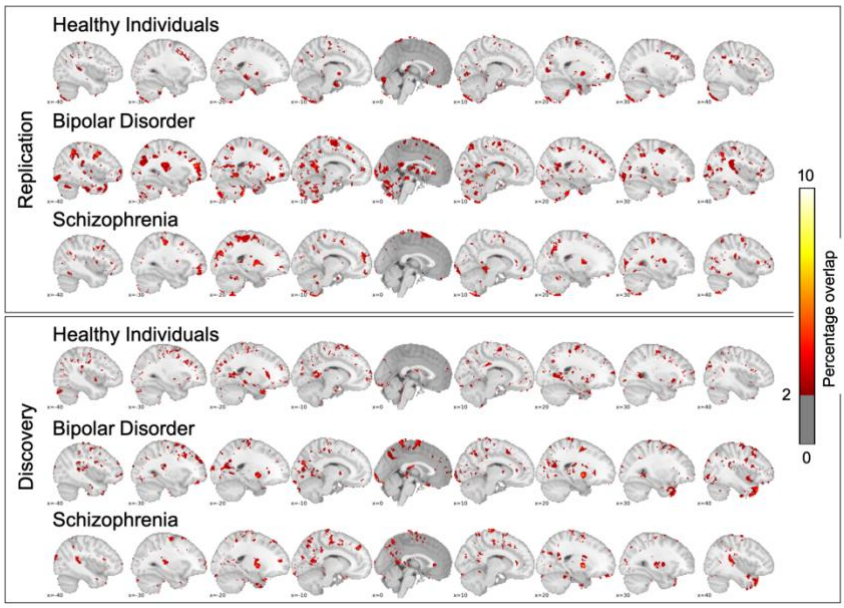
SUPPLMENTARY MATERIALS

Supplementary Table 1: Extreme deviations						
	Replication			**Discovery**		
	Healthy Individuals	**Bipolar Disorder**	**Schizophrenia**	**Healthy Individuals**	**Bipolar Disorder**	**Schizophrenia**
Extreme negative (mean, std)	.15 +- .46	.15+- .34%	.46+- .97%	.23 +- 0.76%	.24 +- .49%	.90 +- 2.16%
Significance	HC = BD **HC < SZ (p < .001)** **BP < SZ (p < .001)**			HC = BD **HC < SZ (p < .001)** **BP < SZ (p < .001)**		
Extreme positive (mean, std)	1.18 +- 2.23%	.92 +- 1.53%	.81 +- 1.73%	1.08 +- 1.76%	.79 +- 1.25%	.78 +- 1.51%
Significance	HC = BD **HC > SZ (p < .01)** BP = SZ			HC = BD **HC > SZ (p < .001)** BP = SZ		
Abbreviations: NA, not applicable; PANSS, Positive and Negative Syndrome Scale.						

Supplementary Figure 1: We depict the contrast between healthy individuals, bipolar disorder and schizophrenia in a subsample of the replication sample with a similar age distribution as in the discovery sample for healthy individuals (mean age 34.89 +- 11.11; males = 50.86%). All contrasts are corrected for multiple comparisons. We show that our results are robust when age and sex in the healthy group more similar to the distribution in the discovery sample. You can compare the mean and standard deviations reported above to the relevant numbers in table 1.



Supplementary Figure 2: In the upper panel we depict the mean Z-scores based on the PALM results for individuals that are healthy individuals or those that hold a diagnosis of bipolar disorder or schizophrenia. In the lower panel we depict results based on the data reported in Wolfers et al. 2018, Jama Psychiatry.



Supplementary Figure 3: *Extreme positive deviations for healthy individuals, bipolar disorder and schizophrenia. We show that the overlap across studies is comparable with only a few brain regions showing overlap in more than 2% of the individuals diagnosed with the same mental disorder.* Peak voxels show extreme positive overlap in 4.66% in healthy individuals, 8.15% in individuals with bipolar disorder and 6.13% in schizophrenia. *In the lower panel we depict results based on the data reported in Wolfers et al. 2018, Jama Psychiatry.*