Dissecting heterogeneity in cortical thickness abnormalities in major depressive disorder: a large-scale ENIGMA MDD normative modelling study

Dr. Bayer, J.M.M [1, 2, 3, 4 +], Dr. van Velzen, L.S [2, 3], Dr. Pozzi, E [2, 3], Prof Davey, C[5], A/Prof Han, L.K.M[6], Dr. Bauduin, S.E.E.C[7], Dr. Bauer, J[8], Dr. Benedetti, F[9], Dr. Berger, K[10], Dr. Bonnekoh, L.M[1112], Dr. Brosch, K[13], Dr. Bülow, R[14], Dr. Couvy-Duchesne, B[15,16,17,18] Dr. Cullen, K.R[19], Dr. Dannlowski, U[11], Dr. Dima, D[20, 21], Dr. Dohm, K[11], Dr. Evans, J.W[22], Dr. Fu, C.H.Y[23], Dr. Fuentes-Claramonte, P[24], Dr. Godlewska, B.R[25, 26], Dr. Goltermann, J[11], Dr. Gonul, A[27], Dr. Goya-Maldonado, R[28], Dr. Grabe, H.J[29], Dr. Groenewold, N.A[30], Dr. Grotegerd, D[11], Dr. Gruber, O[31], Dr. Hahn, T[11], Dr. Hall, G.B[32], Dr. Hamilton, J[33], Dr. Harrison, B.J[5], Dr. Hatton, S.N[34], Dr. Hermesdorf, M[35], Dr. Hickie, I.B[34], Dr. Ho, T.C[36], Dr. Jahanshad, N[37], Dr. Jansen, A[13], Dr. Jamieson, A.J[5,] Dr. Kamishikiryo, T[38], Dr. Kircher, T[13], Dr. Klimes-Dougan, B[19], Dr. Krämer, B[31], Dr. Kraus, A[11], Dr. Krug, A[39], Dr. Leehr, E.J[11], Dr. Leenings, R[11], Dr. Li, M[40], Prof. McIntosh, A[41], Dr. Medland, S.E[42], Dr. Meinert, S[11, 43], Dr. Melloni, E[9], Dr. Mwangi, B[44], Dr. Nenadić, I[13], Dr. Okada, G[45], Dr. Oudega, M.[6, 46, 47, 48,] Dr. Portella, M.J[49], Dr. Rodríguez, E[24], Dr. Romaniuk, L[41], Dr. Rosa, P.G.[50], Dr. Sacchet, M.D[51], Dr. Salvador, R[24], Dr. Sämann, P.G[52], Dr. Shinzato, H[53], Dr. Sim, K[54,55,56] Dr. Simulionyte, E[57], Dr. Soares, J.C[44], Dr. Stein, D.J[58], Dr. Stein, F[13], Dr. Stolicyn, A[41], Dr. Straube, B[13, 59], Dr. Strike, L.T[60], Teutenberg, L[13], Dr. Thomas-Odenthal,F[13], Thomopoulos, S.I[61], Usemann, P[13], Dr. van der Wee, N.J.A[62], Dr. Völzke, H[61], Dr. Wagenmakers[63,46], M. ,Dr. Walter, M[64, 6566,67] Prof. Whalley, H.C[41], Dr. Whittle, S[2,3], Dr. Winter, N.R[11], Dr. Wittfeld, K[68], Dr. Wu, M[44], Prof. Yang, T.T[69], Dr. Zarate, C.A[70], Dr. Zunta-Soares, G.B[44], Prof. Thompson, P.M[71], Prof Veltman, D.J[72], Prof Marquand, A.F*[1,4], Prof. Schmaal, L*[2,3]

+ corresponding author

* joint last author

Contact information corresponding author:
Johanna.bayer@radboudumc.nl

[1] Donders Institute for Brain, Cognition and Behaviour, the Netherlands

[2] Centre for Youth Mental Health, University of Melbourne, Melbourne, VIC, Australia Australia

[3] Orygen, Parkville, VIC

[4] Radboudumc, Nijmegen, the Netherlands

[5] Department of Psychiatry, The University of Melbourne, Australia

[6] Amsterdam UMC, location Vrije Universiteit Amsterdam, Department of Psychiatry, Amsterdam Neuroscience, De Boelelaan 1117, Amsterdam, the Netherlands

[7] Leiden University Medical Center, Leiden, the Netherlands

[8] University Clinic for Radiology, University of Muenster, Germany

[9] Division of Neuroscience, IRCCS Sar Raffaele Scientific Institute, Milan, Italy

[10] Institute of Epidemiology and Social Medicine, University of Münster, Germany

[11] Institute for Translational Psychiatry, University of Münster, Münster, Germany

[12] Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy University of Münster, Germany

[13] Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Germany

[14] University Medicine Greifswald Institute for Radiology and Neuroradiology, Germany

[15] Institute for Molecular Bioscience, the University of Queensland, St Lucia, QLD, Australia

[16] Sorbonne University, Paris Brain Institute - ICM, France

[17] CNRS, Inria, Inserm, AP-HP, France

[18] Hôpital de la Pitié Salpêtrière, F-75013, Paris, France

[19] University of Minnesota, Minneapolis, Minnesota, USA

[20] Department of Psychology, School of Health and Psychological Sciences, City, University of London, London

[21] UK Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

[22] NIMH/NIH, USA

[23] Centre for Affective Disorders, King's College London Department of Psychology, University of East London

[24] FIDMAG Germanes Hospitalaries Research Foundation, Barcelona, Spain CIBERSAM, ISCIII, Madrid, Spain

[25] Departement of Psychiatry, University of Oxford, UK

[26] Oxford Health NHS Foundation Trust, Oxford, UK

[27] Ege University School of Medicine Department of Psychiatry, Turkey

[28] Laboratory of Systems Neuroscience and Imaging in Psychiatry (SNIP-Lab), Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany

[29] University Medicine Greifswald, Germany

[30] Department of Psychiatry & Mental Health, Neuroscience Institute, University of Cape Town, Cape Town, South Africa

[31] Section for Experimental Psychopathology and Neuroimaging, Department of General Psychiatry, Heidelberg University, Germany

[32] Dept of Psychology, Neuroscience & Behaviour, McMaster University, Ontario, USA.

[33] Department of Biological and Medical Psychology; University of Bergen; Bergen, Norway

[34] University of Sydney, Australia

[35] Institute of Epidemiology and Social Medicine, University of Münster, Germany

[36] Department of Psychology, University of California, Los Angeles Brain Research Institute and Interdepartmental Graduate Program in Neuroscience, University of California, Los Angeles, USA

[37] Mark and Mary Stevens Neuroimaging and Informatics Institute Department of Biomedical Engineering, USA

[38] Hiroshima University

[39] Department of Psychiatry and Psychotherapy, University Hospital Bonn, Germany

[40] Department of Psychiatry and Psychotherapy, Jena University Hospital, Jena, Germany

[41] Division of Psychiatry, Centre for Clinical Brain Science, University of Edinburgh

[42] Queensland Institute of Medical Research, Queensland, Australia

[43] Institute for Translational Neuroscience, University of Münster, Münster, Germany

[44] Center of Excellence on Mood Disorders, Louis A. Faillace, MD, Department of Psychiatry and Behavioral Sciences, McGovern Medical School, the University of Texas Health Science Cetner at Houston, USA

[45] Department of Psychiatry and Neurosciences, Graduate School of Biomedical and Health Sciences, Hiroshima University, Japan

[46] GGZinGeest, Specialized Mental Health Care, Amsterdam, the Netherlands

[47] Amsterdam Neuroscience, Mood Anxiety Psychosis Sleep & Stress program, Amsterdam, the Netherlands.

[48] Amsterdam Public Health, Mental Health program, Amsterdam, the Netherlands

[49] Institut d'Investigació Biomèdica Sant Pau (IIB Sant Pau) Universitat Autònoma de Barcelona (UAB) Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Spain

[50] PLaboratory of Psychiatric Neuroimaging (LIM-21), Departamento e Instituto de Psiquiatria, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, SP, Bazil

[51] Meditation Research Program, Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

[52] Max Planck Institute of Psychiatry, Munich, Germany

[53] Department of Psychiatry and Neurosciences, Graduate School of Biomedical and Health Science, Hiroshima University, Hiroshima, Japan Department of Neuropsychiatry, Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan

[54] Institute of Mental Health, Singapore

[55] Yong Loo Lin School of Medicine, National University of Singapore, Singapore

[56] Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

[57] Section for Experimental Psychopathology and Neuroimaging, Department of Psychiatry, University of Heidelberg, Heidelberg, Germany

[58] SAMRC Unit on Risk & Resilience in Mental Disorders, Dept of Psychiatry & Neuroscience Institute, University of Cape Town. South Afrika

[59] Germany Center for Mind, Brain and Behavior - CMBB, Marburg, Germany

[60] Brain and Mental Health, QIMR Berghofer Medical Research Institute, Brisbane, Australia School of Biomedical Sciences, Faculty of Medicine, University of Queensland, Brisbane, Australia

[61] Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Marina del Rey, CA,USA

[62] Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands Leiden Institute for Brain and Cognition, Leiden University Medical Center, The Netherlands

[63] Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, Noord Holland 1081 HV, The Netherlands

[64] Insititute for Community Medicine, University Medicine Greifswald, Germany

[65] Department of Psychiatry and Psychotherapy, Jena University Hospital, Jena, Germany

[66] Germany German Center for Mental Health (DZPG), partner site Halle-Jena-Magdeburg, Germany

[67] Germany Center for Intervention and Research on adaptive and maladaptive brain Circuits underlying mental health (C-I-R-C), Halle-Jena-Magdeburg, Germany

[68] Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany

[69] Department of Psychiatry and Behavioral Sciences Division of Child and Adolescent Psychiatry University of California at San Francisco (UCSF), USA

[70] National Institute of Mental Health, USA

[71] University of Southern California, USA

[72] Dept. Psychiatry, Amsterdam UMC, Amsterdam, the Netherlands

Keywords: Normative modelling; depression; cortical thickness; ENIGMA; neuroimaging; individualised predictions

Key points

Question: Can z-scores derived from normative modelling shed light on the heterogeneous group-level findings of cortical thickness abnormalities in major depression and what characterises individuals at the extreme ends of cortical thickness abnormalities?

Finding: We confirmed a large overlap in z-score distributions between depressed individuals and healthy controls and a heterogeneous spatial distribution of extreme z-deviations across brain regions across individual patients. Lower z-scores for cortical thickness were related to more severe clinical characteristics.

Meaning: Our findings confirm the heterogeneity in individual variation in the location and extent of CT abnormalities across patients with MDD and stress the importance of individualised predictions when examining cortical thickness abnormalities.

Abstract

Importance: Major depressive disorder (MDD) is highly heterogeneous, with marked individual differences in clinical presentation and neurobiology, which may obscure identification of structural brain abnormalities in MDD. To explore this, we used normative modeling to index regional patterns of variability in cortical thickness (CT) across individual patients.

Objective: To use normative modeling in a large dataset from the ENIGMA MDD consortium to obtain individualised CT deviations from the norm (relative to age, sex and site) and examine the relationship between these deviations and clinical characteristics.

Design, setting, and participants: A normative model adjusting for age, sex and site effects was trained on 35 CT measures from FreeSurfer parcellation of 3,181 healthy controls (HC) from 34 sites (40 scanners). Individualised z-score deviations from this norm for each CT measure were calculated for a test set of 2,119 HC and 3,645 individuals with MDD. For each individual, each CT z-score was classified as being within the normal range (95% of individuals) or within the extreme range (2.5% of individuals with the thinnest or thickest cortices).

Main outcome measures: Z-score deviations of CT measures of MDD individuals as estimated from a normative model based on HC.

Results: Z-score distributions of CT measures were largely overlapping between MDD and HC (minimum 92%, range 92-98%), with overall thinner cortices in MDD. 34.5% of MDD individuals, and 30% of HC individuals, showed an extreme deviation in at least one region, and these deviations were widely distributed across the brain. There was high heterogeneity in the spatial location of CT deviations across individuals with MDD: a maximum of 12% of individuals with MDD showed an extreme deviation in the same location. Extreme negative CT deviations were associated with having an earlier onset of depression and more severe depressive symptoms in the MDD group, and with higher BMI across MDD and HC groups.

3

Extreme positive deviations were associated with being remitted, of not taking antidepressants and less severe symptoms.

Conclusions and relevance: Our study illustrates a large heterogeneity in the spatial location of CT abnormalities across patients with MDD and confirms a substantial overlap of CT measures with HC. We also demonstrate that individualised extreme deviations can identify protective factors and individuals with a more severe clinical picture.

1. Introduction

Major depressive disorder (MDD) is a highly prevalent mental illness, impacting more than 300 million people worldwide[1–3]. Neuroimaging research in the last decades has aimed to identify the neural basis of MDD. Cortical thickness has shown to be genetically and phenotypically independent from other imaging phenotypes and to be affected by complex and distinct pruning and myelination processes related to typical and atypical learning and development trajectories in clinical and subclinical populations[4–8]. While earlier studies found medium to large effect sizes for cortical thickness (CT) alterations in MDD (Cohen's d 0.48-0.60)[9,10], more recent studies on CT suggest that these effect sizes have been overestimated, potentially due to small sample sizes and publication bias (e.g., maximum Cohen's d=0.13, in larger samples[11–13]). Rather, having a diagnosis of MDD seems to only subtly affect regional CT, evidently in frontal (including orbitofrontal cortex (OFC) areas, the anterior cingulate (ACC), the dorsomedial prefrontal cortex (PFC)) and posterior cingulate and temporal areas[12,14,15]. These small observed effect sizes imply that many individuals with MDD show CT values that overlap with those of healthy individuals[11,16], and thus cannot be directly used for classification purposes.[17]

The association between these small effect sizes and the large inter-individual clinical heterogeneity within MDD[18–20] remains unresolved. To a certain degree, stratifications of MDD by clinical features (such as age of onset or use of antidepressants) point to subgroups with stronger morphological alterations[12,21,22]. Still, there remains a large portion of MDD patients with CT patterns practically not distinguishable from HC. In addition, while most studies find thinner cortices[23–25] to be associated with depression, also thicker cortices[24,26], or no difference[27] have been reported, especially in periods of rapid brain maturation such as adolescence[28]. Due to the general change of CT with age[29,30], these CT findings may be the result of a complex interaction between the effects of maturation, ageing and the MDD diagnosis.

The group-average comparison approach that is typical of neuroimaging studies is meant to look at the group as a whole and does not identify individual differences within a group, as it treats individualised variability as a nuisance, allowing only for inferences at the group level. In addition, a complex interaction between brain maturation or ageing and diagnosis may be obscured by limiting the analysis to group comparisons, particularly when age-interaction effects are not modelled. Normative modelling[31,32] (NM) is an alternative

approach to identifying morphological brain alterations, with a focus on positioning each individual's brain measure on a normative scale. The framework has been successfully applied to clinical neuroimaging data including CT measures, uncovering the anatomical heterogeneity in various mental disorders[28,29,33–39]. NM establishes the range of (statistically) normal variation in the data based on the covariates in the model (e.g., age and sex) and subsequently parses each individual's raw score onto an individualised z-score deviation from that norm. As one practical use case, NM allows to detect individuals with extreme deviations (e.g., +/- 2 standard deviations) from the norm, which might be beneficial in cases of large heterogeneity, subtypes[40], or where anatomical differences are subtle[12].

Beyond this, multi-site neuroimaging data are sensitive to effects induced by site differences[41–43]. Especially in the case of small effect sizes, site effects can thus easily overshadow effects of interest. Extending previous work[44], we have developed linear and non-linear versions of normative models that can deal with these site differences[43]. Our approach thus combines the individualization aspect of NM with the ability to correct for site effects, making normative modelling applicable for multi-site neuroimaging data[43,44].

In this study, we aim to elucidate the heterogeneity in CT alterations across individuals with MDD by examining the variation in CT z-scores for different brain regions in a large, pooled depression data set from the ENIGMA MDD consortium using NM[45,46]. First, we aim to re-estimate group average differences in CT between MDD and HC based on z-scores from NM that reflect an age, sex and site normed metric, hypothesizing a similar anatomical CT deficit pattern compared to our previous ENIGMA MDD findings of CT alterations in MDD[12]. Further, we aim to investigate characteristics of individuals with extreme deviations in either direction (positive and negative deviations), quantifying their frequency and anatomical distribution. For this we explored the association of extreme CT deviations with clinical and lifestyle features, including symptom severity, number of depressive episodes, use of antidepressant medication, age of onset, childhood trauma and body mass index (BMI).

2. Methods

2.1. Sample

The ENIGMA MDD working group combines neuroimaging and clinical data across 4,597 MDD patients and 5,926 healthy controls (HC) from 53 international cohorts. Data from 34

cohorts were included in this study. All sites obtained approval from local institutional review boards and ethics committees. All participants provided informed consent.

2.2. Demographic and clinical characteristics

Demographic information include age, sex, BMI and MDD status (HC, or MDD; for demographics, clinical characteristics, in- and exclusion criteria, diagnostic instruments used and scan protocol details per cohort, see **Tables S1.1-3**). Clinical information included episode status (HC, first episode depression, recurrent episode depression), remission status (HC, acutely depressed, remitted), antidepressant use at the time of scan (HC, antidepressant-free, antidepressant user), age of onset of depression (MDD with =< 21 years age of first onset, MDD with > 21 years age of onset[12,21]), Beck Depression Inventory (BDI)[47] total score (only for MDD), Hamilton Depression Rating Scale[48] (HDRS, only for MDD) total score and Childhood Trauma Questionnaire (CTQ) total score[49] (see **Tables S1.4-5**).

2.3. Image processing and analysis

3D structural T1-weighted brain MRI scans were acquired at each site (image acquisition parameters are detailed in **Table S1.3**). Analysis and quality control of the images were performed locally using harmonised protocols provided by the ENIGMA consortium (http://enigma.ini.usc.edu/protocols/imaging-protocols). Cortical parcellations were acquired using FreeSurfer (versions 5.1 and 5.3)[50] based on the Desikan-Killiany atlas[51] (34 regions in the left and 34 in the right hemisphere, plus one average across all regions). Measures were visually inspected and statistically analysed for outliers using standardised ENIGMA protocols http://enigma.ini.usc.edu/protocols/imaging-protocols). As our prior studies did not show lateralisation effects of CT alterations in MDD[52], CT values were averaged for each region across the left and right hemispheres, resulting in 34 bilateral regions and one whole-brain average CT measure. The final sample included in this study amounted to 5,300 HCs and 3,645 individuals with MDD from 34 sites including 40 different MR scanning platforms.

2.4. Training and test set

HC data was divided into a training set (60% of HCs, N=3,181) and test set (40% of HCs, N= 2,119) using a within-site split. The normative model was not trained on all HCs to ensure that

potential differences between HC and MDD in deviation z-scores were not due to worse fit of the model in the MDD group compared to HC because the model was trained on all HCs. The training and test sets had comparable distributions in terms of age and sex (**Figure 1,** for a more detailed analysis see **supplemental text S2.2a and table S2.2a,b**). All individuals with MDD (N=3,645) were part of the MDD test set. The MDD test and the HC test sets were imputed, scaled, and centred across sites, based on the HC training set using k=10 nearest neighbour imputation. An overview of the data processing pipeline may be found in **Figure 1** and **supplemental S2.2.**

2.5.   Normative Modelling

We applied our previously developed NM framework predicting CT measures based on age, sex and site[43]. A hierarchical Bayesian approach in *Stan,* described elsewhere[43], was applied, including the covariates, age, sex and site. A more in-depth description of the NM process may be found in **supplemental text S2.4**. All code can be found in: https://github.com/likeajumprope/Normative-ENIGMA-MDD. An interactive version of all figures can be found here: https://likeajumprope.github.io/Normative-ENIGMA-MDD/

2.6.   Statistical analysis

All statistical analyses reported below were performed on deviations in CT from the norm (z-scores) derived from the MDD and HC test sets. A hierarchical Bayesian regression model with age, sex and site as predictors was used to create a model mapping the normative variation in each CT measure onto age and sex, while accommodating for multi-site effects[43]. We further used a B-splines model with 5 knots to model non-linear associations between age and CT, which allowed us to calculate z-scores for each participant *i* and region *j* using the following:

$$z_{ij} = \frac{f(x_{ij}) - y_{ij}}{\sqrt{\sigma^2 + \sigma_*^2}}$$

Where $f(x_{ij})$ is the prediction of the model, y_{ij} is the true value, σ^2 is a random gaussian noise component and σ_*^2 is the predictive variance from the model (for more details see [43,53]).

2.6.1. Average group differences in z-scores per ROI

T-tests were performed to examine MDD vs HC group differences in z-scores. Region-wise overlap scores, Cohen's *d* metrics and distributional overlap percentages between HCs and individuals with MDD[11] were also calculated. Due to residual associations with z-scores, age and sex were included as covariates in all statistical models described below. P-values were corrected using false-discovery-rate (FDR) correction for 35 statistical models (34 regions + 1 global brain measure).

2.6.2. Extreme deviations per ROI and global measure

Individuals were divided into subgroups for each bilateral region, based on their z-score deviation: (1) falling within the *norm* (*normal*; between and 2.5th and 97.5th percentile, or z-scores between -1.96 or 1.96), (2) *extreme positive deviators* from the norm (*supra-normal*; above 97.5thth percentile with z-scores >= 1.96), and (3) *extreme negative deviators* (*infra-normal*; below 2.5th percentile, with z-scores =< -1.96). This resulted in a 0 (normal), -1 (infra-normal) or 1 (supra-normal) score for each cortical region for each individual. Percentages of individuals with MDD with at least one infra- or supra-normal z-score were compared to the percentages of HCs with infra- and supra-normal z-scores using a chi-squared test.

2.6.3. Spatial location indifferent summary deviation scores

To account for the heterogeneity in the spatial location of the infra- and supra-normal z-scores across different individuals, we derived the following spatial location indifferent summary measures that target the analysis of extreme deviations[54,55]:

- *Average z-score*: the average z-score across all 34 cortical regions.

- *Load score*: individual total number of regions with infra- or supra-normal z-scores across all brain regions. Load scores were created by separately summing the number of infra and supra normal z-score deviations per individual, resulting in one positive and one negative load scores with a range from 0 – 35 each[54,55].

- *Extremity score*: z-score with the largest negative or positive value across all regions per individual resulting in one positive and one negative extremity score per individual[54,55].

3. Results

3.1. Model fit

Model fits of the normative models to the training were satisfactory in terms of Explained Variance, Root Mean Squared Error, Correlation and Mean Standardized Log Loss for the test sets. A distribution of extreme deviations close to the expected values of 5% for both test sets indicated a good generalisability of the training model **(Table S1.9)**. A summary can be found in **supplemental text S2.5** and in **tables S1.10-13**.

3.2. Group average differences in z-scores

Significantly lower z-scores in individuals with MDD compared to HC were observed in 25 out of 35 brain regions, with greatest differences in the fusiform gyrus (Cohen's d=-0.17), inferior (d=-0.13) and middle temporal gyrus (d=-0.13), the bank of the superior temporal sulcus (d=-0.12) and the insula (d=-0.12, **Figure 2d,e**). These results aligned with those of a previous publication[12] (ρ=0.70, see **supplemental text S2.6**). No average regional CT increases in MDD compared with HC were detected. Across all regions, z-score distributions were shifted towards more negative in MDD (**Figure 2f**). The percentage of z-score distribution overlap between HC and MDD ranged between 92% and 98%. The fusiform gyrus, the inferior frontal gyrus pars opercularis and average CT showed the lowest overlap, indicating the largest difference in z-score distributions between HC and MDD (92.3%, p<0.001; 93.3%, p<0.001; 93.4% p<0.001; respectively; **Figure 2a-c, Table S1.7)**. For a full region-wise overview of Cohen's d and distributional overlap scores, see **Table S1.7**.

3.3. Heterogeneity in spatial location of extreme negative and positive deviations

In MDD, 30.6% of individuals showed at least one cortical region with a supra-normal z-score (extreme positive deviation), versus 32.1% in HC (χ^2=11.755, p<0.001), while 34.5% of individuals with MDD had at least one cortical region with an infra-normal z-score (extreme negative deviation) compared to 30.0% of HC (χ^2=12.305, p<0.001). An illustration of the number of regions with extreme deviations is given in **Figure 3.**

No more than 12.1% (pericalcarine gyrus, cuneus) of individuals with MDD showed a positive extreme (supra-normal) z-score within the same region (**Figure 3b**) and no more than

10

11.9% (fusiform gyrus) of individuals showed an infra-normal z-score within the same region (**Figure 3a**). This illustrates the heterogeneity that is evident in the spatial location of extreme negative and positive deviations across individuals with MDD. A list of the percentages of extreme deviations for each region can be found in **Table S1.8**. Example profiles of individuals with the highest negative load and extremity score and the most negative average z-score can be found in **S2.10**.

3.4. Deviation scores and correlation with clinical variables

An overview of all associations between all summary deviation scores and clinical covariates is given in **Table 1** and in **Tables S1.14-S1.18 and S2.9**.

3.5. *Average z-scores*

Lower average z-scores were associated with a higher likelihood to be in the MDD group (0.83 relative risk (RR), p<0.001) and being acutely depressed (RR=0.78; p<0.001), compared to being in the HC group. A one unit decrease, equivalent to one standard deviation (SD), in average z-score was associated with a 0.99-point increase in BDI scores (p=0.04), while no association with HDRS scores was observed. A one SD decrease in average z-scores was associated with a 0.99-point BMI (kg/m^2) increase across MDD and HC (p<0.001, see **S1.14**).

3.6. *Load scores*

There were no differences between MDD and HC for positive or negative load scores (**Table 1 & S1.15-16**). For negative load, having more regions with extreme negative deviations was associated with a higher likelihood of not taking antidepressants at time of scanning (vs. HC: RR=1.03; p=0.048) and of having an early onset of MDD (vs HC: RR=1.04, p=0.05, vs. late onset MDD: RR=0.96, p=0.05). A one-unit decrease in positive load scores was associated with a 0.24-point decrease in HDRS scores (p=0.05). Across groups a one unit decrease in positive load scores led to a 0.19 km/m^2 increase in BMI (p=0.01), while a one unit increase in negative load scores was associated with a 0.16 km/m^2 increase in BMI (p = 0.048).

2.7. *Extremity scores*

Positive and negative extremity were defined as the most positive or negative individual z-score across all regions. Greater negative extremity scores were associated with a higher

11

likelihood of being in the MDD group (vs. HC: RR=1.14, p=0.001), of having an early onset of MDD (vs. HC: RR=1.21, p<0.001, vs. late onset: RR=0.89 p=0.02), being either acutely depressed or remitted (acute vs. HC: RR=1.22 p<0.001, remitted vs. HC: RR=1.27 p<0.001), using or not using antidepressants (AD use vs. HC: RR=1.14 p<0.001, AD free vs. HC: RR=1.13 p=0.01) and having higher BDI (0.823-point increases, p=0.04) and BMI scores (0.49-point increase, p=0.04, across groups). Greater positive extremity scores showed a link to a lower likelihood of being in the antidepressant user group (vs. HC: RR=0.87, p=0.009, vs. antidepressant-free group: RR=0.88, p=0.03), of being in the late onset MDD group (vs. HC: RR=0.89, p=0.020), and of being in the acutely depressed group (vs. remitted: RR=0.84, p=0.038). Positive extremity scores were negatively associated with disease severity (-0.56-point HDRS decrease per unit increase) and with BMI (-0.109 unit decrease in BMI, across groups, see **Table 1 & S1.17-18**).

4. Discussion

In this large study using NM, we used our previously developed algorithm[43,44] to characterise the heterogeneity of CT in depression. We found lower z-scores in individuals with MDD compared to HC, mostly in temporal, frontal and parietal areas, and in the insular cortex, with small effect sizes (Cohen's *d* ranging 0.05-0.2) and with a large overlap in z-scores between MDD patients and HC (distributional overlap ranging 92-98%). The magnitude of these effects and the anatomical distribution of the deviations are in line with our previous results from group average comparison studies[12,14,15].

4.1. Heterogeneity in cortical thickness deviations in depression

The main aim of our NM approach was to examine interindividual differences in CT alterations between individuals with MDD. Assigning either a supra-normal, infra-normal or normal z-score per individual per cortical region revealed that ~70% of individuals with MDD showed no extreme deviation in any brain region, thus having CT values that fall fully within the norm. The remaining ~30% of individuals with MDD, however, exhibited wide-spread heterogeneity in degree and spatial distribution of these extreme z-scores. No single cortical region showed an extreme deviation in more than 12.1% of individuals with MDD. This lack of spatial overlap in CT alterations demonstrates the large variation of CT abnormalities in MDD and is in line with spatial percentage overlap in CT z-scores observed in previous NM studies in other disorders such as schizophrenia, ADHD and bipolar disorder[33-35,56]. The variation in extreme

12

deviations in different cortical regions across individuals with MDD may be associated with distinct behavioural and symptom phenotypes that have been reported within MDD[57,58]. Importantly, the high degree of regional heterogeneity within MDD suggests that group average comparisons are not representative of a specific profile of CT alterations apparent in every individual patient (see also case study examples in **S2.8**).

Even though we observed considerable heterogeneity in the spatial location of CT alterations at the regional level they may still form part of the same higher order functional circuit. This is supported by findings showing that regional extreme deviations from NM can be coupled into common higher-order functional circuits and networks[59]. Regional overlap between individuals in extreme z-scores derived from NM increased from <7% up to 21% when integrated into higher order functional networks, in particular in circuits including the prefrontal cortex (PFC) and other frontal areas[59]. This network character of CT alterations could be the driving underlying factor for clinical phenotypes in MDD, re-integrating the heterogeneity at the level of regional CT alterations.

4.2. Large overlap in z-score distributions of cortical thickness measures between individuals with depression and healthy controls

Importantly, we found a high similarity between individuals with MDD and HCs in CT measures. The maximum group difference between distributions in z-scores of regional cortical thicknesses was 8%. In other words, 92% of the distributions in CT z-scores overlapped between MDD and HC. Similar findings have been reported previously[11], with distributional overlap scores of structural brain alterations between healthy individuals and individuals with MDD of between 87% and 95%.

Around 30% of HCs also showed extreme deviations in CT, which was similar to the percentage of individuals with MDD with extreme deviations and reflects the high distributional overlap between MDD and HC. It should be noted, however, that this percentage rises with a larger number of regions analyzed. Given this high overlap between MDD and HC, CT alterations may be driven by other factors commonly associated with MDD but present across individuals with and without depression, such as genetic[60,61], sociodemographic[62], clinical (e.g. childhood trauma[63]), nutritional[64], activity-related[65,66], and lifestyle factors (e.g., smoking, drinking[67,68]). This hypothesis is supported by our observed negative association between BMI and z-scores of CT alterations across HC and MDD individuals, without any interaction between BMI and diagnosis. A similar association between obesity and CT

13

alterations has been found previously in healthy adults[69–74]. Several studies suggest that increased neuroinflammation may be a mediating factor in the association between obesity and cortical thinning[75–77] (although also thickening has been reported[76]) and might contribute to the pathophysiology of cortical thickness alterations in depression[70].

4.3.   Associations with clinical characteristics within MDD

While one implication of this study is that extreme negative deviations to some extent differentiate between healthy controls and depressed individuals, we also found support that extreme deviations can be used to distinguish between subgroups within depression.

Within depression, having higher positive extremity scores decreased the likelihood of being acutely depressed compared to being remitted. This suggests that slight alterations in CT may normalise following remission, which is in line with previous studies showing attenuated reductions in prefrontal and anterior cingulate cortices and the hippocampus in remitters at a 3-year follow up compared to non-remitters[78]. Alternatively, more extreme positive deviations in CT may protect against chronic depression (or develop in a compensatory manner), as increased orbitofrontal and inferior temporal volumes[79] and thicker frontal[80] and anterior cingulate[79] cortices have been shown to predict future remission.

We further found that, within MDD, more extreme positive CT deviations (greater positive extremity scores) were associated with a decreased likelihood of taking antidepressants. While antidepressants have been associated with neuroprotective effects and neurogenesis[81] (see also[12,21]), large-scale cross-sectional studies have consistently shown most wide-spread effects with largest effect sizes of structural brain alterations in those with MDD taking antidepressants. This may in part be driven by overall illness severity, as individuals who are more severely depressed or have had a more chronic or recurrent course of depression are more likely to use antidepressants. This is indirectly supported by our finding of negative associations of positive extremity and load scores with overall severity of depressive symptoms (HDRS scores). Nonetheless, longitudinal studies are required to clarify these associations.

Further, more extreme negative deviations were associated with an earlier age of MDD onset (<21 years of age) compared to adult age of onset in MDD. This could either point to more extreme negative deviations in CT constituting a risk factor for the development of depression, or a consequence of longer illness duration (co-correlated with earlier age of onset), which have both been reported previously[82].

Unlike previous studies, we did not find any association between summary scores of deviations and childhood trauma[63,83]. This discrepancy may stem from the different approach of our study compared to previous studies[63], as we used summary scores of extreme deviations across cortical regions instead of examining scores per cortical region[83]. In addition, the age-corrected z-score measures used in this study do not allow testing for interactions between CT and age, which have been reported previously[63], and the inclusion of the total CTQ scores did not allow for any interaction with the subscales of the measurement.

4.4. Limitations

Despite being the largest NM study in MDD to date, our study is limited by the cross-sectional nature of the current dataset, not allowing us to determine a causal direction for any of the observed associations. The use of longitudinal data to establish normative trajectories could lead to an increased understanding of the causal processes underlying the association between z-score deviations and clinical variables.

Further, the current study is limited by the clinical covariates currently available in the ENIGMA MDD data set. Detailed information on illness history, genetic and environmental risk factors[61,84] and duration, type and dosage of antidepressant treatment[15], were not available for many sites. The inclusion of those parameters into the normative model could be the target of future studies.

5. Conclusions and future directions

While most individuals with MDD do not show extreme deviations in CT, the spatial location and distribution of those who do show extreme alterations is highly heterogeneous across individuals. The similarity of the percentage of individuals with extreme deviations in MDD and HC groups points to an important role of non-depression specific mechanisms affecting CT in both MDD and HC (e.g., BMI). Extreme negative deviations predicted MDD group membership, while extreme positive deviations were linked to more favourable clinical characteristics within MDD, though these differences were subtle. This suggests that regional CT may not provide diagnostic biomarkers for MDD. Likely we need more temporally and spatially finely grained imaging techniques to detect structural brain deviations that can better differentiate between MDD and HC or use data-driven techniques that demarcate biotypes.

Other neuroimaging modalities, such as functional brain measures, may have the potential to yield larger effect sizes (although see[85]). This limitation in diagnostic sensitivity, however, does not necessarily limit the use of neuroimaging measures as markers of treatment response[85] or general disease prognosis[86].

6.   Author contributions

J.M.M.B: Conceptualization, Writing - Original Draft, Methodology, Software, Formal analysis, Validation, Visualisation, Data curation; L.S.v: Writing - Review & Editing; E.P: Writing - Review & Editing, Project administration, Supervision, Data curation; C.D: Writing - Review & Editing; L.K.M.H: Writing - Review & Editing; S.E.E.C.B: Data curation, Project administration; J.B: Resources, Data curation; F.B: Writing - Review & Editing; K.B: Resources, Data curation, Writing - Review & Editing, Funding acquisition; L.M.B: Writing - Review & Editing;K.B: Data curation, Writing - Review & Editing; R.B: Methodology, Writing - Review & Editing; B.C: Data curation, Writing - Review & Editing; K.R.C: Investigation, Resources, Writing - Review & Editing, Funding acquisition; U.D: Investigation, Resources, Data curation, Writing - Review & Editing, Funding acquisition; D.D: Formal analysis, Resources, Data curation, Writing - Review & Editing, Supervision; K.D: Resources, Data curation; J.W.E: Formal analysis, Data curation C.H.Y.F: Investigation, Resources, Writing - Review & Editing; P.F: Investigation, Data curation, Writing - Review & Editing; B.R.G: Writing - Review & Editing; I.H.G: Writing – Review & Editing, J.G: Investigation, Data curation, Writing - Review & Editing; A.G: Investigation; R.G: Resources, Writing - Review & Editing; H.J.G: Resources, Data curation, Writing - Review & Editing, Project administration; N.A.G: Data curation, Writing - Review & Editing, Project administration; D.G: Data curation, Writing - Review & Editing; O.G: Formal analysis, Investigation, Writing - Review & Editing, Supervision; T.H: Data curation, Writing - Review & Editing; G.B.H: Formal analysis; J.H: Investigation, Resources, Data curation, Writing - Review & Editing; B.J. H: Investigation; S.N.H: Resources, Data curation; M.H: Writing - Review & Editing, Project administration; I.B.H: Methodology, Formal analysis, Data curation, Writing - Review & Editing, Visualization, Funding acquisition; T.C.H: Formal analysis, Data curation, Writing - Review & Editing, Funding acquisition; N.J: Methodology, Resources, Writing - Review & Editing, Project administration; A.J.J: Writing - Review & Editing; A.J: Data curation, Writing - Review & Editing; T.K: Resources, Data curation; T.K: Data curation, Writing - Review & Editing T.T.J.K: Resources, Writing - Review & Editing,

16

7. Acknowledgements

18

19

8. Conflicts of interest

K.B has received research funding as principal or coordinating investigator from the German Ministry of Education and Research (BMBF). HJG has received travel grants and speakers honoraria from Fresenius Medical Care, Neuraxpharm and Janssen Cilag as well as research funding from Fresenius Medical Care. IBH is the Co-Director, Health and Policy at the Brain and Mind Centre (BMC) University of Sydney. The BMC operates an early-intervention youth services at Camperdown under contract to headspace. He is the Chief Scientific Advisor to, and a 3.2% equity shareholder in, InnoWell Pty Ltd which aims to transform mental health services through the use of innovative technologies. J.C.S states that within the last twelve (12) months of the date set forth below, neither h nor, to the best of my knowledge, any member of his family has any interest in or has taken any action which would violate the Conflict of Interest Policy, except such interest or action as he has fully disclosed below: ALKERMES, ALLERGAN, ASOFARMA, ATAI, BOEHRINGER Ingelheim, COMPASS, JOHNSON & JOHNSO, LIVANOVA, PFIZER, PULVINAR NEURO LLC, RELMADA, SANOFI, SUNOVIAN. G.B.S-Z is a full-time U.S government employee. He is listed as a coinventor on a patent for the use of ketamine in major depression and suicidal ideation. Dr. Zarate is listed as a coinventor on a patent for the use of (2R,6R)-hydroxynorketamine, (S)-dehydronorketamine and other stereoisomeric dehydro and hydroxylated metabolites of (R,S)-ketamine metabolites in the treatment of depression and neuropathic pain. G.B.S-Z is listed as co-inventor on a patent application for the use of (2R,6R)-hydroxynorketamine and (2S,6S)-hydroxynorketamine in the treatment of depression, anxiety, anhedonia, suicidal ideation and post-traumatic stress disorders. G.B.S-Z has assigned his patent rights to the U.S. government but will share a percentage of any royalties that may be received by the government. The views

expressed are his own and do not necessarily represent the views of the National Institutes of

Health, the Department of Health and Human Services, or the United States Government.

9. Figures
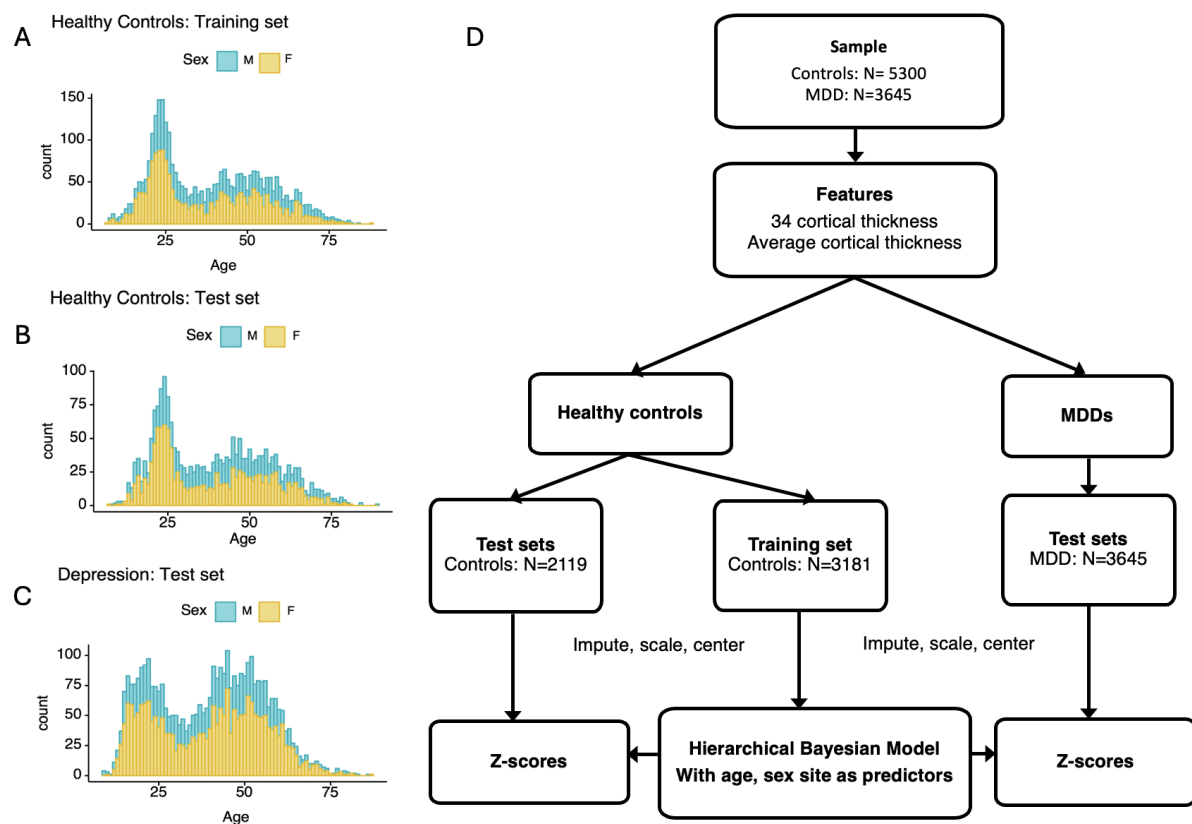
Figure 1. Distribution of age and sex for the three partitions of the data used in this study. A) Training set, healthy controls. B) Test set, healthy controls. M= male, F=female C) Test set, MDD. D) Illustrated Pre-Processing pipeline. The sample consisted of 34 regional features and one whole brain measure from 5300 healthy controls and 3645 individuals with MDD from the ENIGMA MDD consortium.
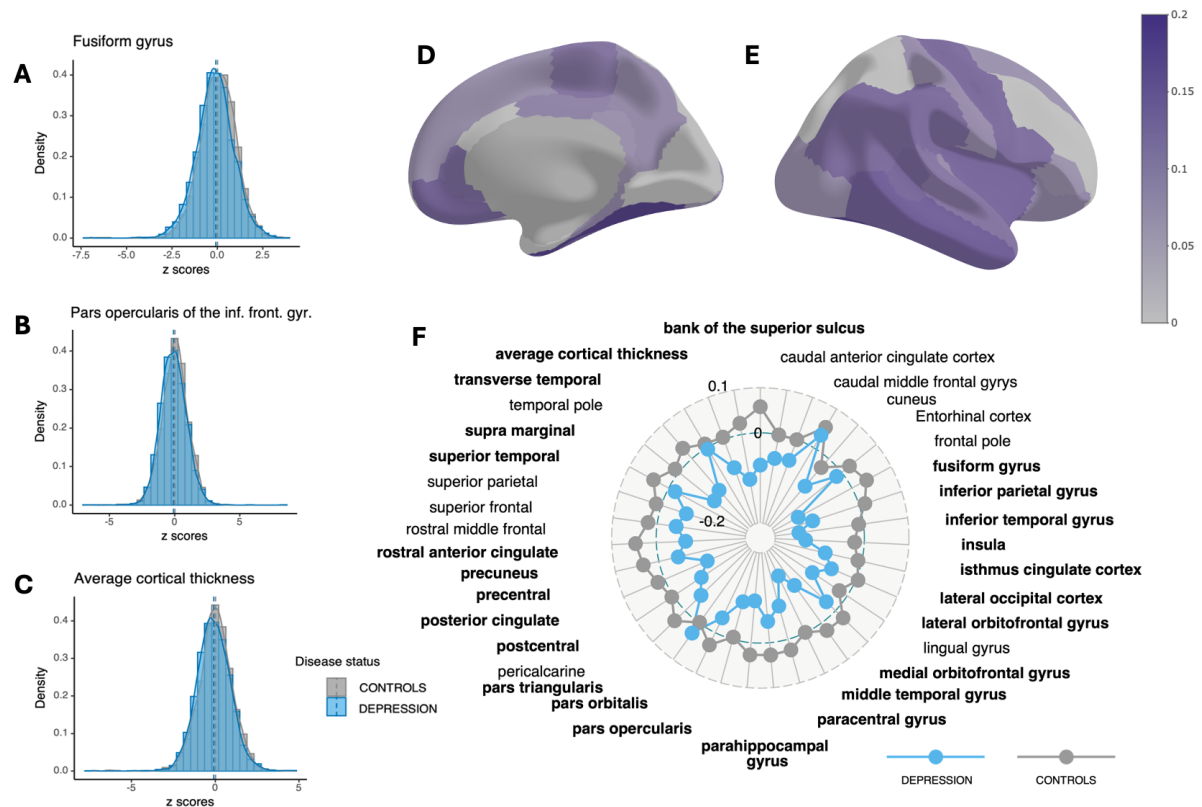
Figure 2. A, B, C: Distributional overlap between individuals with MDD and HCs (HC) for the A. fusiform gyrus, B: the pars opercularis of the inferior frontal gyrus and C: average CT, showing the lowest distributional overlap between MDD and HC groups. **D, E:** Cohen's d effect sizes of differences in z-scores between MDD and HCs for each significant cortical region (FDR corrected, only significant regions are displayed). Red: Individuals with MDD have lower z-scores (thinner cortices) on average than HCs. An overview of Cohen's d effect sizes for z-score differences in different cortical regions is provided in Table S1.7 in the supplementary material. D) medial view. E) lateral view. **F:** Spider plot showing the average z-scores for individuals with MDD and HCs for all cortical regions. All regions that show significant differences are plotted in bold.
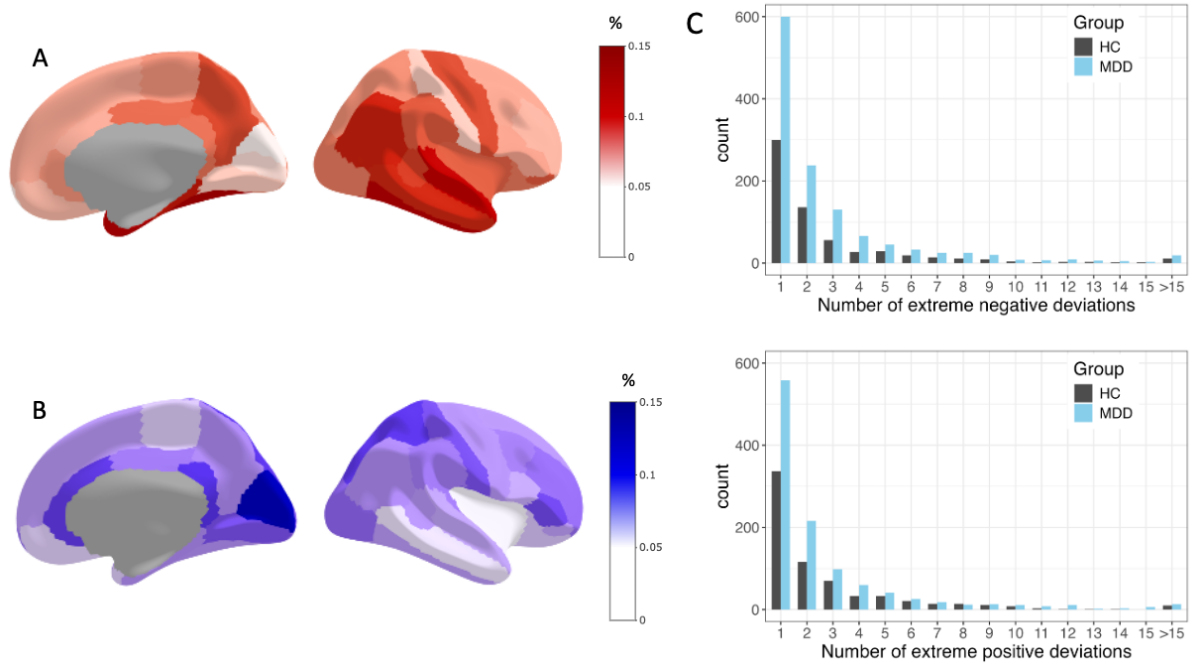
Figure 3. A. B: Heterogeneity in cortical thickness deviations in depression**.** Percentage of individuals with MDD with an infra-normal z-scores (extreme negative deviations; z < -1.96) and supra-normal z-scores (extreme positive deviations; z >1.96) per cortical region. **(A)** infra-normal z-scores; **(B)** supra-normal z-scores. **C:** Distribution of positive load score (number of extreme negative deviation; z > 1.96) **(C, top)** and negative load score (number of extreme positive deviations; z < -1.96) **(C, bottom)** in individuals with MDD and healthy controls.

Independent variable / Dependent var.		Average z ↑	Positive Load ↑	(absolute) Negative Load ↑	Positive extremity ↑	(absolute) Negative extremity ↑
Group	MDD vs *HC*	↓ (0.83)	*n.s.*	*n.s.*	*n.s.*	↑ (1.14)
Episode status	FE vs *HC*	↓ (0.78)	*n.s.*	*n.s.*	*n.s.*	↑ (1.20)
	RE vs *HC*	↓ (0.84)	*n.s.*	*n.s.*	*n.s.*	↑ (1.15)
	FE vs *RE*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	-
Remission status	ACUTE vs *HC*	↓ (0.78)	*n.s.*	*n.s.*	*n.s.*	↑ (1.22)
	REM vs *HC*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	↑ (1.27)
	ACUTE vs *REM*	*n.s.*	*n.s.*	*n.s.*	↓ (0.85)	-
Antidepressant use	AD use vs *HC*	↓ (0.88)	*n.s.*	*n.s.*	↓ (0.87)	↑ (1.14)
	AD free vs *HC*	↓ (0.90)	*n.s.*	↑ (1.03)	*n.s.*	↑ (1.13)
	AD use vs *AD free*	*n.s.*	*n.s.*	*n.s.*	↓ (0.88)	*n.s.*
Age of onset of MDD	EAO vs *HC*	↓ (0.78)	*n.s.*	↑ (1.04)	*n.s.*	↑ (1.21)
	LAO vs *HC*	↓ (0.84)	*n.s.*	-	(0.88)	*n.s.*
	LOA vs *EOA*	*n.s.*	*n.s.*	↓ (0.96)	*n.s.*	↓ (0.89)
Severity	BDI	↓ ($-0.99pt$)	*n.s.*	*n.s.*	*n.s.*	↑ ($0.82pt$)
	HDRS	*n.s.*	↓ ($-0.23pt$)	*n.s.*	↓ ($-0.55pt$)	*n.s.*
BMI*	BMI	↓ ($-0.99pt$)	↓ ($-0.19pt$)	↑ (0.16pt)	↓ ($-0.66pt$)	↑ ($0.48pt$)
Childhood Trauma**	CTQ	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*

Table 1. Odds ratios and effect direction from predicting clinical covariates with brain summary scores with logistic and linear regression. Reference category in *italics*. MDD: Major depressive disorder. HC: healthy controls. FE: First episode depression. RE recurrent episode depression. ACUTE: acutely depressed. REM: remitted depressed. AD use: antidepressant use at the time of the scan. AD free: antidepressant free at the time of the scan. EAO: early age of onset (=< 21 years). LAO: late age of onset (> 21 years). BDI: Beck depression inventory. HDRS Hamilton depression rating scale. BMI: Body Mass Index. * main effects reported, interaction effect with MDD status not significant. **Neither main nor interaction effects significant. ↑ increase. ↓ decrease. *pt*: point. FDR adjusted. n.s. means no significant association was found. Full tables can be found in supplementary material S1.14-S1.18.

10. References

1. Ferrari, A. J. *et al.* Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med.* **10**, e1001547 (2013).

2. Smith, K. Mental health: A world of depression. *Nature Publishing Group UK* http://dx.doi.org/10.1038/515180a (2014) doi:10.1038/515180a.

3. Liu, Q. *et al.* Changes in the global burden of depression from 1990 to 2017: Findings from the Global Burden of Disease study. *J. Psychiatr. Res.* **126**, 134–140 (2020).

4. Nunes, A. S. *et al.* Atypical age-related changes in cortical thickness in autism spectrum disorder. *Sci. Rep.* **10**, 11067 (2020).

5. McLaughlin, K. A. *et al.* Widespread reductions in cortical thickness following severe early-life deprivation: a neurodevelopmental pathway to attention-deficit/hyperactivity disorder. *Biol. Psychiatry* **76**, 629–638 (2014).

6. Lotter, L. D. *et al.* Regional patterns of human cortex development correlate with underlying neurobiology. *Nat. Commun.* **15**, 7987 (2024).

7. Sowell, E. R. *et al.* Longitudinal mapping of cortical thickness and brain growth in normal children. *Journal of Neuroscience* **24**, 8223–8231 (2004).

8. Winkler, A. M. *et al.* Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *Neuroimage* **53**, 1135–1146 (2010).

9. Niu, M. *et al.* Common and Specific Abnormalities in Cortical Thickness in Patients with Major Depressive and Bipolar Disorders. *EBioMedicine* **16**, 162–171 (2017).

10. Lener, M. S. *et al.* Cortical abnormalities and association with symptom dimensions across the depressive spectrum. *J. Affect. Disord.* **190**, 529–536 (2016).

11. Winter, N. R. *et al.* Quantifying Deviations of Brain Structure and Function in Major Depressive Disorder Across Neuroimaging Modalities. *JAMA Psychiatry* **79**, 879–888 (2022).

12. Schmaal, L. *et al.* Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol. Psychiatry* **22**, 900–909 (2017).

13. Kochunov, P. *et al.* Comparison of regional brain deficit patterns in common psychiatric and neurological disorders as revealed by big data. *Neuroimage Clin* **29**, 102574 (2021).

14. Suh, J. S. *et al.* Cortical thickness in major depressive disorder: A systematic review and meta-analysis. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **88**, 287–302 (2019).

15. Li, Q. *et al.* Meta-analysis of cortical thickness abnormalities in medication-free patients with major depressive disorder. *Neuropsychopharmacology* **45**, 703–712 (2020).

16. Fried, E. I. & Kievit, R. A. The volumes of subcortical regions in depressed and healthy individuals are strikingly similar: a reinterpretation of the results by Schmaal et al. *Mol. Psychiatry* **21**, 724–725 (2016).

17. Belov, V. *et al.* Multi-site benchmark classification of major depressive disorder using machine learning on cortical and subcortical measures. *arXiv [q-bio.QM]* (2022).

18. Monroe, S. M. & Anderson, S. F. Depression: The Shroud of Heterogeneity. *Curr. Dir. Psychol. Sci.* **24**, 227–231 (2015-6).

19. Fried, E. I. Moving forward: how depression heterogeneity hinders progress in treatment and research. *Expert Rev. Neurother.* **17**, 423–425 (2017).

20. Goldberg, D. The heterogeneity of "major depression". *World Psychiatry* **10**, 226–228 (2011).

21. Schmaal, L. *et al.* Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. *Mol. Psychiatry* **21**, 1–7 (2015).

22. Ho, T. C. *et al.* Subcortical shape alterations in major depressive disorder: Findings from the ENIGMA major depressive disorder working group. *Hum. Brain Mapp.* **43**, 341–351 (2022).

23. Boes, A. D., McCormick, L. M., Coryell, W. H. & Nopoulos, P. Rostral Anterior Cingulate Cortex Volume Correlates with Depressed Mood in Normal Healthy Children. *Biol. Psychiatry* **63**, 391–397 (2008-2).

24. Fallucca, E. *et al.* Distinguishing Between Major Depressive Disorder and Obsessive-Compulsive Disorder in Children by Measuring Regional Cortical Thickness. *Arch. Gen. Psychiatry* **68**, 527 (2011).

25. Shad, M. U., Muddasani, S. & Rao, U. Gray matter differences between healthy and depressed adolescents: a voxel-based morphometry study. *J. Child Adolesc. Psychopharmacol.* **22**, 190–

197 (2012-6).

26. Reynolds, S. *et al.* Cortical thickness in youth with major depressive disorder. *BMC Psychiatry* **14**, 83 (2014-3).

27. Whittle, S. *et al.* Structural brain development and depression onset during adolescence: a prospective longitudinal study. *Am. J. Psychiatry* **171**, 564–571 (2014).

28. Bethlehem, R. A. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).

29. Wierenga, L. M., Langen, M., Oranje, B. & Durston, S. Unique developmental trajectories of cortical thickness and surface area. *Neuroimage* **87**, 120–126 (2014).

30. Frangou, S. *et al.* Cortical thickness across the lifespan: Data from 17,075 healthy individuals aged 3–90 years. *Hum. Brain Mapp.* **31**, hbm.25364 (2021-2).

31. Marquand, A. F., Wolfers, T., Mennes, M., Buitelaar, J. & Beckmann, C. F. Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **1**, 433–447 (2016).

32. Marquand, A. F. *et al.* Conceptualizing mental disorders as deviations from normative functioning. *Mol. Psychiatry* **24**, 1415–1424 (2019).

33. Lv, J. *et al.* Individual deviations from normative models of brain structure in a large cross-sectional schizophrenia cohort. *Mol. Psychiatry* **26**, 3512–3523 (2021).

34. Wolfers, T. *et al.* Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. *JAMA Psychiatry* **75**, 1146–1155 (2018).

35. Wolfers, T. *et al.* Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models. *Psychol. Med.* **50**, 314–323 (2020).

36. Zabihi, M. *et al.* Fractionating autism based on neuroanatomical normative modeling. *Transl. Psychiatry* **10**, 384 (2020).

37. Marquand, A. F., Rezek, I., Buitelaar, J. & Beckmann, C. F. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biol. Psychiatry* **80**, 552–561 (2016).

38. Verdi, S. *et al.* The heterogeneous brain: Mapping individualised patterns of atrophy in Alzheimer's disease using spatial normative models. *Alzheimers. Dement.* **17**, (2021).

39. Storsve, A. B. *et al.* Differential longitudinal changes in cortical thickness, surface area and volume across the adult life span: Regions of accelerating and decelerating change. *Journal of Neuroscience* **34**, 8488–8498 (2014).

40. Jiang, Y. *et al.* Two neurostructural subtypes: results of machine learning on brain images from 4,291 individuals with schizophrenia. *medRxiv* (2023) doi:10.1101/2023.10.11.23296862.

41. Glocker, B., Robinson, R., Castro, D. C., Dou, Q. & Konukoglu, E. Machine Learning with Multi-Site Imaging Data: An Empirical Study on the Impact of Scanner Effects. *arXiv [eess.IV]* 1–5 (2019).

42. Gronenschild, E. H. B. M. *et al.* The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical Volume and Cortical Thickness Measurements. *PLoS One* **7**, e38234 (2012).

43. Bayer, J. M. M. *et al.* Accommodating site variation in neuroimaging data using normative and hierarchical Bayesian models. *Neuroimage* **264**, 119699 (2022).

44. Kia, S. M. *et al.* Hierarchical Bayesian Regression for Multi-site Normative Modeling of Neuroimaging Data. in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020* 699–709 (Springer International Publishing, 2020). doi:10.1007/978-3-030-59728-3_68.

45. Schmaal, L. *et al.* ENIGMA MDD: seven years of global neuroimaging studies of major depression through worldwide data sharing. *Transl. Psychiatry* **10**, 172 (2020).

46. Thompson, P. M. *et al.* ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl. Psychiatry* **10**, 100 (2020).

47. Beck, A. T., Guth, D., Steer, R. A. & Ball, R. Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. *Behav. Res. Ther.* **35**, 785–791 (1997).

48. Hamilton, M. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* **23**, 56–62 (1960).

49. Bernstein, D. P. J. Childhood Trauma Questionnaire. *Assessment of family violence: A handbook for researchers and practitioners.* (1998) doi:10.1037/t02080-000.

50. Fischl, B. *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in

the human brain. *Neuron* **33**, 341–355 (2002).

51. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).

52. de Kovel, C. G. F. *et al.* No Alterations of Brain Structural Asymmetry in Major Depressive Disorder: An ENIGMA Consortium Analysis. *Am. J. Psychiatry* **176**, 1039–1049 (2019).

53. Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes for Machine Learning*. (MIT Press, 2005).

54. Elad, D. *et al.* Improving the predictive potential of diffusion MRI in schizophrenia using normative models—Towards subject-level classification. *Hum. Brain Mapp.* **42**, 4658–4670 (2021).

55. Bouix, S. *et al.* Increased Gray Matter Diffusion Anisotropy in Patients with Persistent Post-Concussive Symptoms following Mild Traumatic Brain Injury. *PLoS One* **8**, e66205 (2013).

56. Wolfers, T. *et al.* Replicating extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder. *Hum. Brain Mapp.* **42**, 2546–2555 (2021).

57. Fried, E. I. & Nesse, R. M. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *J. Affect. Disord.* **172**, 96–102 (2015-2).

58. Toenders, Y. J. *et al.* Neurovegetative symptom subtypes in young people with major depressive disorder and their structural brain correlates. *Transl. Psychiatry* **10**, 108 (2020).

59. Segal, A. *et al.* Regional, circuit and network heterogeneity of brain abnormalities in psychiatric disorders. *Nat. Neurosci.* **26**, 1613–1629 (2023).

60. Kendler, K. S., Fiske, A., Gardner, C. O. & Gatz, M. Delineation of two genetic pathways to major depression. *Biol. Psychiatry* **65**, 808–811 (2009).

61. Kwong, A. S. F. *et al.* Genetic and Environmental Risk Factors Associated With Trajectories of Depression Symptoms From Adolescence to Young Adulthood. *JAMA Netw Open* **2**, e196587 (2019).

62. Akhtar-Danesh, N. & Landeen, J. Relation between depression and sociodemographic factors. *Int. J. Ment. Health Syst.* **1**, 4 (2007).

63. Tozzi, L. *et al.* Interactive impact of childhood maltreatment, depression, and age on cortical

brain structure: mega-analytic findings from a large multi-site cohort. *Psychol. Med.* **50**, 1020–1031 (2020).

64. T S Sathyanarayana, R., M R, A., B N, R. & K S Jagannatha, R. Understanding nutrition, depression and mental illnesses. *Indian J. Psychiatry* **50**, 77–82 (2008).

65. Morgan, W. P. Selected physiological and psychomotor correlates of depression in psychiatric patients. *Res. Q.* **39**, 1037–1043 (1968).

66. Martinsen, E. W., Strand, J., Paulsson, G. & Kaggestad, J. Physical fitness level in patients with anxiety and depressive disorders. *Int. J. Sports Med.* **10**, 58–61 (1989).

67. Quittschalle, J. *et al.* Association of alcohol and tobacco consumption with depression severity in the oldest old. Results from the age different old age cohort platform. *Int. J. Environ. Res. Public Health* **18**, 7959 (2021).

68. Hämäläinen, J. *et al.* Cigarette smoking, alcohol intoxication and major depressive episode in a representative population sample. *J. Epidemiol. Community Health* **55**, 573–576 (2001).

69. Laurent, J. S. *et al.* Associations Among Body Mass Index, Cortical Thickness, and Executive Function in Children. *JAMA Pediatr.* **174**, 170–177 (2020).

70. Cho, J., Seo, S., Kim, W.-R., Kim, C. & Noh, Y. Association Between Visceral Fat and Brain Cortical Thickness in the Elderly: A Neuroimaging Study. *Front. Aging Neurosci.* **13**, 694629 (2021).

71. Kurth, F. *et al.* Relationships between gray matter, body mass index, and waist circumference in healthy adults. *Hum. Brain Mapp.* **34**, 1737–1746 (2013).

72. Pannacciulli, N. *et al.* Brain abnormalities in human obesity: a voxel-based morphometric study. *Neuroimage* **31**, 1419–1425 (2006).

73. Shaw, M. E., Sachdev, P. S., Abhayaratna, W., Anstey, K. J. & Cherbuin, N. Body mass index is associated with cortical thinning with different patterns in mid- and late-life. *Int. J. Obes.* **42**, 455–461 (2018).

74. McWhinney, S. R. *et al.* Association between body mass index and subcortical brain volumes in bipolar disorders-ENIGMA study in 2735 individuals. *Mol. Psychiatry* **26**, 6806–6819 (2021).

75. Prats-Soteras, X. *et al.* Inflammatory agents partially explain associations between cortical

thickness, surface area, and body mass in adolescents and young adulthood. *Int. J. Obes. (Lond)* **44**, 1487–1496 (2020).

76. Gómez-Apo, E., Mondragón-Maya, A., Ferrari-Díaz, M. & Silva-Pereyra, J. Structural brain changes associated with overweight and obesity. *J. Obes.* **2021**, 6613385 (2021).

77. Hagenberg, J. *et al.* Dissecting depression symptoms: Multi-omics clustering uncovers immune-related subgroups and cell-type specific dysregulation. *Brain Behav. Immun.* **123**, 353–369 (2025).

78. Frodl, T. S. *et al.* Depression-related variation in brain morphology over 3 years: effects of stress? *Arch. Gen. Psychiatry* **65**, 1156–1165 (2008).

79. Phillips, J. L., Batten, L. A., Aldosary, F., Tremblay, P. & Blier, P. Brain-volume increase with sustained remission in patients with treatment-resistant unipolar depression. *J. Clin. Psychiatry* **73**, 625–631 (2012).

80. Motter, J. N. *et al.* Cortical thickness predicts remission of depression with antidepressants in patients with late-life depression and cognitive impairment. *J. Affect. Disord.* **295**, 438–445 (2021).

81. Pilar-Cuéllar, F. *et al.* Neural plasticity and proliferation in the generation of antidepressant effects: hippocampal implication. *Neural Plast.* **2013**, 537265 (2013).

82. Truong, W. *et al.* Changes in cortical thickness across the lifespan in major depressive disorder. *Psychiatry Res.* **214**, 204–211 (2013).

83. Ahn, S. J. *et al.* What is the impact of child abuse on gray matter abnormalities in individuals with major depressive disorder: a case control study. *BMC Psychiatry* **16**, 397 (2016).

84. Li, J. *et al.* Cortical structural differences in major depressive disorder correlate with cell type-specific transcriptional signatures. *Nat. Commun.* **12**, (2021).

85. Cash, R. F. H., Cocchi, L., Lv, J., Fitzgerald, P. B. & Zalesky, A. Functional Magnetic Resonance Imaging-Guided Personalization of Transcranial Magnetic Stimulation Treatment for Depression. *JAMA Psychiatry* **78**, 337–339 (2021).

86. Schmaal, L. *et al.* Predicting the Naturalistic Course of Major Depressive Disorder Using Clinical and Multimodal Neuroimaging Information: A Multivariate Pattern Recognition Study.

Biol. Psychiatry **78**, 278–286 (2015).