Title: Heterogeneity in Depression: evidence for distinct clinical and neurobiological profiles

Authors: *Kayla Hannon*[1], Ty Easley[1], Wei Zhang[1], Daphne Lew[2,3], Vera Thornton[4], Aristeidis Sotiras[1,3], Yvette I. Sheline[5], Andre Marquand[6], Deanna M. Barch[1,4,7], Janine D. Bijsterbosch[1]

Affiliations:

1. Department of Radiology, Washington University School of Medicine, Saint Louis, Missouri 63110, USA
2. Division of Biostatistics, Washington University School of Medicine
3. Institute for Informatics, Washington University School of Medicine
4. Department of Psychiatry, Washington University School of Medicine
5. Perelman School of Medicine, University of Pennsylvania
6. Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre
7. Department of Psychological & Brain Sciences, Washington University

Corresponding authors:

- Kayla Hannon: khannon@wustl.edu
- Janine Bijsterbosch: janine.bijsterbosch@wustl.edu

Abstract

Addressing heterogeneity in depression is critically important to overcome replicability challenges and gain insights into neural etiology. We developed a novel hierarchical framework to systematically disentangle clinical and neurobiological sources of heterogeneity utilizing population data from the UK Biobank. Firstly, we defined patient subgroups who uniquely shared isolated clinical characteristics of depression (e.g., symptoms of anhedonia, depressed mood, and somatic disturbance; severity indices of lifetime chronicity and acute impairment; and late onset). Our results revealed distinct neurobiological features robustly associated with each clinically isolated subgroup, providing symptom-level insights into the neural etiology of depression, and supporting a one-to-one mapping between clinical and neurobiological sources of heterogeneity. Secondly, we investigated residual neurobiological heterogeneity within each subgroup using data-driven clustering. Our findings revealed stable neurobiological clusters that differed in cognitive ability within two clinical subgroups (chronicity and acute impairment), providing evidence that multiple neurobiological mechanisms may give rise to the same clinical presentation (many-to-one mapping).

Introduction

Depression is one of the most common mental health disorders, as illustrated by a 2017 meta-analysis across eighty-three studies that found overall depression prevalence to be 27%[1]. Despite its prevalence, depression often goes underdiagnosed (especially in middle and older age[2]), and treatment efficacy is poor such that only 30% of patients respond to the first line of treatment[3]. Gaining insights into non-invasive neuroimaging correlates of depression has the potential to improve diagnosis and treatment.

Prior work has suggested multiple neuroimaging correlates of depression, including structural, diffusion, and resting state abnormalities in the frontal cortex, deep nuclear gray matter, and temporal and parietal regions[4–6]. However, these findings have suffered from poor replicability[5,7] and/or small magnitude of effect[8] . Several prominent studies, both broadly in neuroimaging[7,9] and specifically in depression[5,10], have suggested these replicability issues could largely be due to variability among patients in their clinical symptomatology and/or neurobiology, known as "patient heterogeneity".

Prior work to investigate heterogeneity in depression includes initiatives such as RDoC and HiTOP, which have attempted to focus on transdiagnostic clinical commonalities such as RDoC's Loss subconstruct[11] and HiTOP's Internalizing disorders[12]. Data-driven approaches have performed clustering on depression symptom scores to identify clinical subtypes[13–16]. Alternatively, other studies have focused on neurobiological sources of heterogeneity by performing data-driven clustering analyses on neuroimaging biomarkers[17], albeit with challenges linked to overfitting and replicability[18]. To our knowledge, no prior work has attempted to parse heterogeneity by investigating the complex interactions between clinical and neurobiological heterogeneity. In this study, we leverage a large population dataset to jointly disentangle clinical and neurobiological sources of depression heterogeneity.

Approaches that do not account for both clinical and neurobiological sources of heterogeneity may be missing important interactions in depression. Heterogeneity is often implicitly or explicitly conceptualized as the presence of discrete subtypes within the overall patient population. Each subtype is thought to be

characterized by a specific set of clinical features which arise from a specific set of neurobiological features, which we refer to as a one-to-one brain-symptom mapping (Fig. 1A). In this conceptualization, the subtypes could be derived by investigating either the neurobiological or clinical features in isolation. Alternatively, it is possible that a specific set of clinical features may arise from multiple different neurobiological patterns in different patients, which we refer to as many-to-one brain-symptom mapping (Fig. 1B; previously described as equifinality[19]). We first investigate the one-to-one model of depression heterogeneity by identifying neurobiological profiles for multiple clinically homogenous groups, and then use data-driven approaches to investigate residual neurobiological heterogeneity even after accounting for clinical variance, thereby explicitly testing potential many-to-one mapping.

The above conceptualization of heterogeneity in depression also assumes that subtypes represent discrete clusters of individuals (Fig. 1C) rather than a continuous gradient across individuals (Fig. 1D). However, prior evidence consistently finds patient distributions are more continuous in nature (for example see Fig. 1f in Drysdale et al[18] and Fig. 3B in Wen et al[20]). It is important to note that even if heterogeneity follows a relatively continuous distribution, a cut-off (i.e., threshold that defines sub-groups within the continuous variation) may offer important clinical value if it is stable, replicable, and clinically informative. Therefore, identifying stable and clinically informative cut-offs (as opposed to discrete, separated clusters) may present a more realistic and relevant delineation approach for depression. The use of such cut-offs is not uncommon in clinical practice, for example, depression severity is a continuous dimension that is commonly divided into subclinical, mild, moderate, and severe thresholds to guide treatment.

In this study, we parse clinical and neurobiological heterogeneity in depression accounting for the possibility of one-to-one and many-to-one brain-symptom relationships, while investigating the potential value of identifying cut-offs within continuous dimensions. Our findings indicate that both clinical and neurobiological sources of heterogeneity need to be considered simultaneously to gain robust insights into depression etiology.

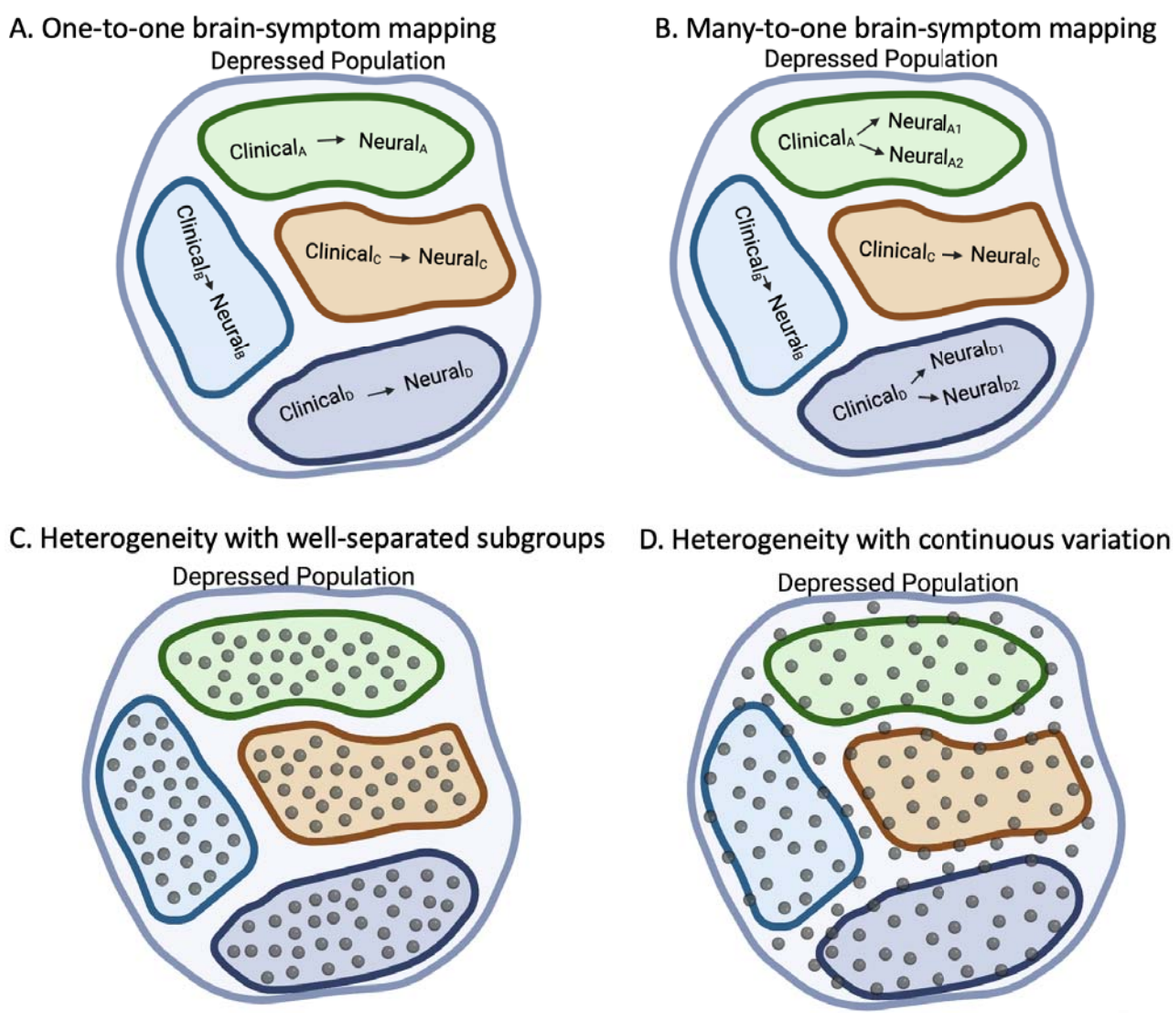

Figure 1. Theoretical perspectives on the relationship between clinical and neurobiological heterogeneity in the depression population. (A). *In the one-to-one theory, each clinical feature is expected to have a one-to-one association with a neurobiological explanation. Note that one-to-one mapping may be multivariate in nature, such that a set of symptoms may be linked to a complex neurobiological pattern.* ***(B).*** *In the many-to-one theory, multiple distinct neurobiological mechanisms can give rise to the same clinical presentation. Note that the one-to-one and many-to-one theories are not mutually exclusive and many-to-one mapping may be observed in some clinically homogeneous groups, but not others. Heterogeneity may either take the form of well-separated clusters* ***(C)*** *or may vary along more continuous dimensions* ***(D).*** *Note that the definition of cut-offs is*

potentially still feasible and desirable even among continuous variation as long as the cut-offs are stable and clinically informative.

Results

Clinical heterogeneity (Fig. 2A)

If heterogeneity led to decreased effect sizes in previous work as is posited, then clinically homogeneous groups should be characterized by stronger and more robust neuroimaging correlates compared to a heterogeneous comparison group. To test this hypothesis, we isolated six clinically homogenous groups and a comparison heterogeneous group (n=4216). The clinically homogeneous groups are three symptom groups (anhedonia n=313, depressed mood n=279, somatic disturbance n=336), two severity groups (chronic n=343, acute impairment n=529), and one age of onset group (late onset n=636) (more information in Tables 1B and 2 in Methods). Notably, the clinically homogeneous groups were explicitly chosen to isolate clinical features along three axes of clinical variation (symptomatology, severity, age of onset), which enables the investigation of differential neurobiological profiles. These clinically homogenous groups are therefore not intended to comprehensively map all individual variation in depression. Indeed, the majority of patients present a multifaceted (rather than isolated) clinical profile and are grouped in the 'heterogeneous group'. The imaging features (Table 1A) investigated included multiple MRI characteristics: gray matter volume (GMV), cortical thickness (CT), DTI fractional anisotropy (FA), and functional connectivity (FC). Normative modeling was used to estimate (separately for each imaging feature) the degree of deviation from a healthy control group (N=8565) for depression participants in all groups. An ANOVA was used to test whether normative deviations of imaging features differed significantly between the groups. For more information on these methods, see Fig. 2A or the Methods section.

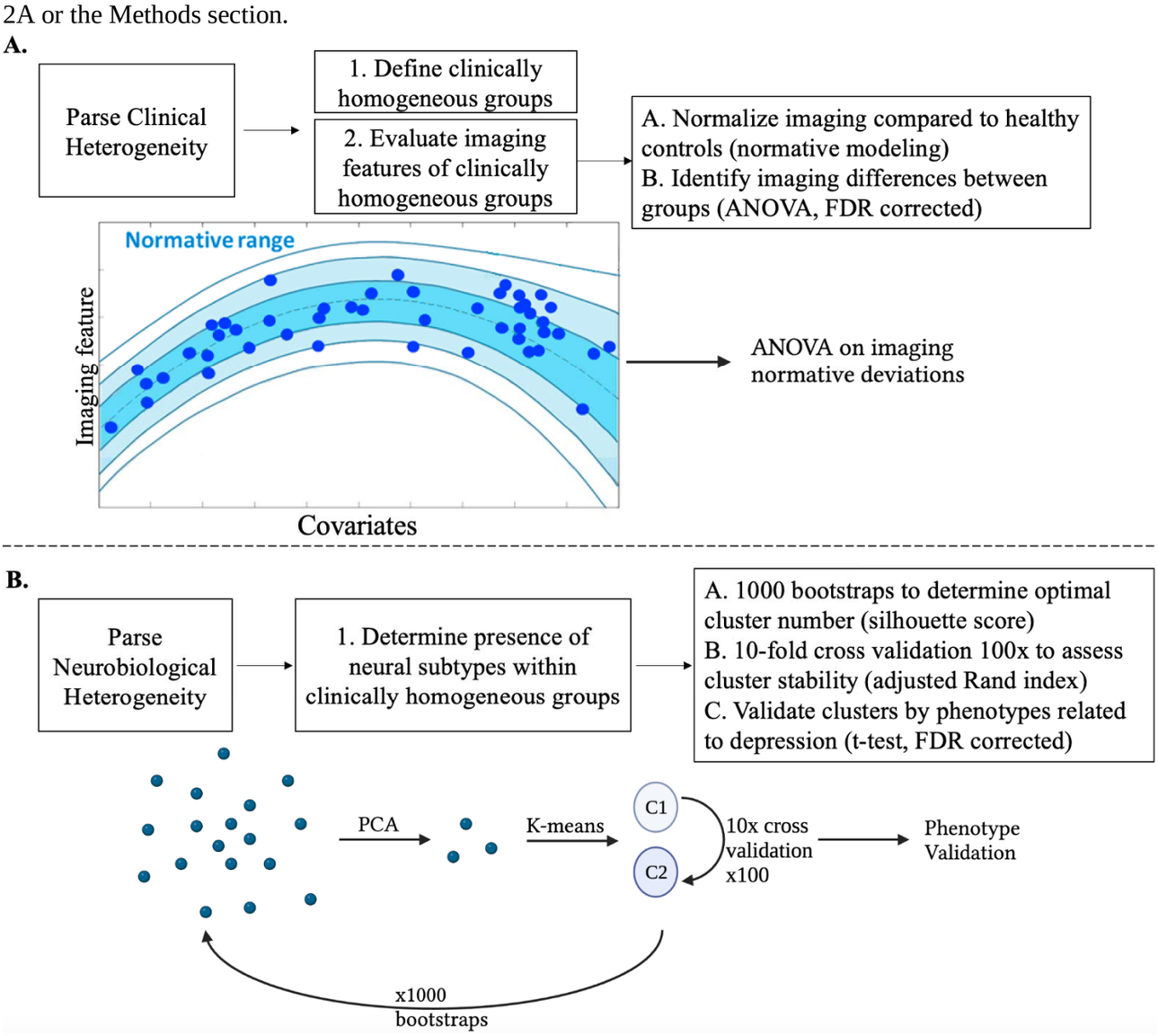

***Figure 2. Analytic framework of study**. **(A)** explains the methodology of our first research question, to parse clinical heterogeneity, with a visualization of the normative model and analysis below. To parse clinical heterogeneity, we isolated clinically homogeneous groups of individuals who are high in one clinical feature but low in all others and a comparison heterogeneous depressed group (with any combination of the clinical features). We then applied normative modeling to their imaging features to normalize compared to healthy controls and employed an ANOVA with false discovery correction to determine which imaging features differed significantly between clinically homogeneous groups. **(B)** explains the methodology of our second research question, to parse residual neurobiological heterogeneity within the clinically homogeneous groups, with a*

visualization of the clustering and validation below. To find clusters within the clinically homogeneous groups based on imaging features, we dimensionally reduced with PCA and performed k-means clustering on the imaging features within each clinically homogeneous group. To determine the optimal cluster number, we performed that pipeline on 1000 bootstraps (randomly sampled the imaging data with replacement) and selected the cluster assignment that had the maximum Silhouette score the most times. We then assessed the stability of the clusters by performing 10-fold cross fold validation 100 times and then calculating the Adjusted Rand Index between the folds. See Methods for further explanation.

16 out of 88 imaging features showed significant group differences after correction for multiple comparisons ($F(6,6645)>2.78$, $p_{FDR}<.05$; Fig. 3). Each of these 16 features were significant more than 80 times over 100 bootstraps of the ANOVA (Supplemental Figure 1), indicating they are highly replicable. Tukey's post-hoc analysis revealed that effects were driven predominantly by differences between the late onset, acute impairment, and heterogenous groups (see full results in Supplemental Table 1). The heterogeneous group was significantly different from at least one other group in 10 out of the 16 features. The acute impairment group differed from at least one other group with more negative normative deviations in many gray matter volumes (insula, amygdala, cingulate gyrus, and frontal orbital GMV) and cortical thicknesses (fusiform, inferiorparietal, and superiorparietal CT). The late onset group differed from at least one other group in all 5 significant CT features (Fig. 3E), and in frontal orbital GMV with more positive normative deviations than all the other groups. Depressed mood differed significantly from at least one other group in the superior longitudinal fasciculus FA and cingulate gyrus GMV. Somatic disturbance differed significantly from at least one other group in cingulate gyrus and frontal orbital GMV as well as inferiorparietal CT. These findings reveal that our clinically homogeneous groups were characterized by distinct neurobiological profiles. The results for the heterogeneous group are of specific interest in light of the one-to-one mapping model as an explanation for low effect sizes and lack of replicability. Notably, the deviations for the heterogeneous group were significantly smaller deviations (closer to zero in magnitude of deviation from controls) for many of the features. These findings support the suggestion that heterogeneity contributed to reduced effect sizes in previous studies.

To decipher if the increased effect sizes for the more clinically homogenous groups reflect larger mean deviations or decreased variance in deviations across individuals within a group, we calculated the intraclass correlation (ICC) for each clinically homogeneous group and the heterogenous group. Resulting ICC estimates were: $ICC_{Anhedonia}=0.88$, $ICC_{DepressedMood}=0.88$, $ICC_{SomaticDisturbance}=0.86$, $ICC_{LateOnset}=0.89$, $ICC_{AcuteImpairment}=0.89$, $ICC_{Heterogenous}=0.88$. As these are quite similar, the driving force of the increased effect sizes of the clinically homogeneous groups is likely due to larger means, not due to less variance.

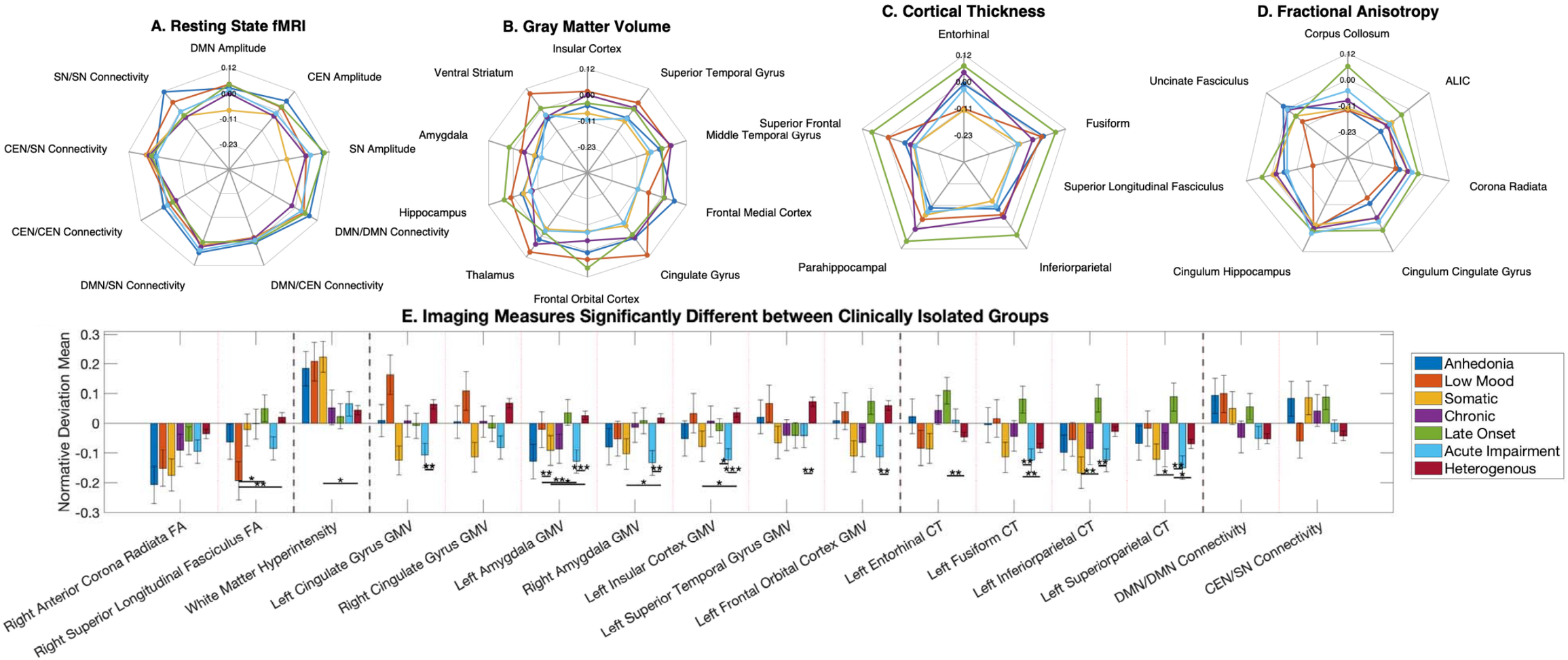

Figure 3. Distinct neurobiological profiles of clinically homogeneous groups*. *(A-D). *Summary multimodal imaging profiles of the clinically homogeneous groups, as represented by their mean normative deviations (relative to a healthy control population). Results were averaged for left and right imaging features where relevant.* ***(E).*** *Results from an ANOVA to test which imaging measures differed significantly between the 7 clinical groups (the clinically homogeneous groups and the heterogeneous group). Bars show the group mean and error bars indicate standard error of mean. Results were significant after false discovery rate multiple comparisons correction across the 88 imaging measures. DMN = default mode network; CEN = central executive network; SN = salience network; ALIC = Anterior Limb of Internal Capsule; FA = fractional anisotropy; GMV = gray matter volume; CT = cortical thickness. * signifies significance at p<0.05, ** signifies p<0.01, and *** signifies p<0.001.*

Neurobiological heterogeneity (Fig. 2B)

We next investigated the many-to-one brain-symptom model for residual neurobiological heterogeneity within clinically homogeneous groups. To address this question, we investigated the presence of stable participant clusters, using k-means to cluster PCA-embedded normative deviations of imaging features (separately within each clinically homogeneous group). We determined the optimal number of clusters, using most frequently highest Silhouette score over one thousand bootstraps, to be two clusters for each clinically homogeneous group (see Supplemental Figure 2). See Methods or Fig. 2B for more information on the analysis pipeline. The presence of stable participant clusters suggests the presence of some residual neurobiological variation even within groups that are highly clinically homogeneous.

On visual inspection, the clusters are relatively continuous in nature (Fig. 4A; and notably consistent with previous work[18,20]) and it is therefore important to determine the stability of the cluster assignments[18]. We used the Adjusted Rand Index (ARI) to assess cluster stability across 100 repeated 10-fold cross validation iterations. A representative example of the stability of cluster assignments between cross validation folds for the acute impairment group is visualized in Fig. 4A and B. ARIs were above 0.6 for all clinically homogeneous groups except for the depressed mood group, and all ARIs exceeded a null calculated from permuted data (Fig. 4C). The null was synthesized by randomly permuting the participant labels independently within each imaging feature (i.e., breaking the correlation between imaging features) for all clinically homogenous groups. The imaging profiles of the clusters are in Supplemental Figure 3.

We furthermore tested the clinical utility of our data-driven clusters using depression relevant phenotypes of cognitive score, trait neuroticism, and socioeconomic deprivation. Cognitive score was calculated based on the 2-item summary measure proposed in Lyall et al 2016[21], trait neuroticism was calculated based on a 12-item sum score[22], and socioeconomic deprivation was quantified using the Townsend Deprivation Index[23]. All UK Biobank (UKB) variable IDs are available in Supplemental Table 2. Two-sided paired t-tests with false discovery rate (FDR) multiple comparison correction were used to determine the statistical difference between the clusters for each clinically homogenous group in these phenotype measures. No significant cluster differences were observed for Neuroticism or the Townsend Deprivation Index post multiple comparisons correction (see Supplemental Table 3). However, the clusters in the chronicity and acute impairment groups significantly differed in cognitive score after correction for multiple comparisons (chronicity $T(341)=2.51$, $p_{FDR}=0.032$; acute impairment $T(527)=3.58$, $p_{FDR}<0.001$; Fig. 4D; full results in Supplemental Table 3). In post hoc analysis, we determined specifically that the verbal-numerical reasoning measure (also called Fluid Intelligence) included in the summary measure was significant for both chronicity ($T(341)=0.105$, $p_{FDR}=0.033$) and acute impairment groups ($T(527)=0.020$, $p_{FDR}=0.0023$). These findings are consistent with prior work[20], and with hypothesized relationships between depression and cognition[24].

For comparison, k-mean clustering was also performed on the combined participant sample including all six clinically homogeneous and the heterogeneous comparison group (combined n=6,652). The resulting clusters replicated the data-driven clusters derived within clinically homogeneous groups well, but did not capture the clinically homogeneous groups (see Supplemental Table 4). These results support the stability of the data-driven clusters in Fig. 4 and highlight the importance of separately parsing clinical and neurobiological sources of heterogeneity.

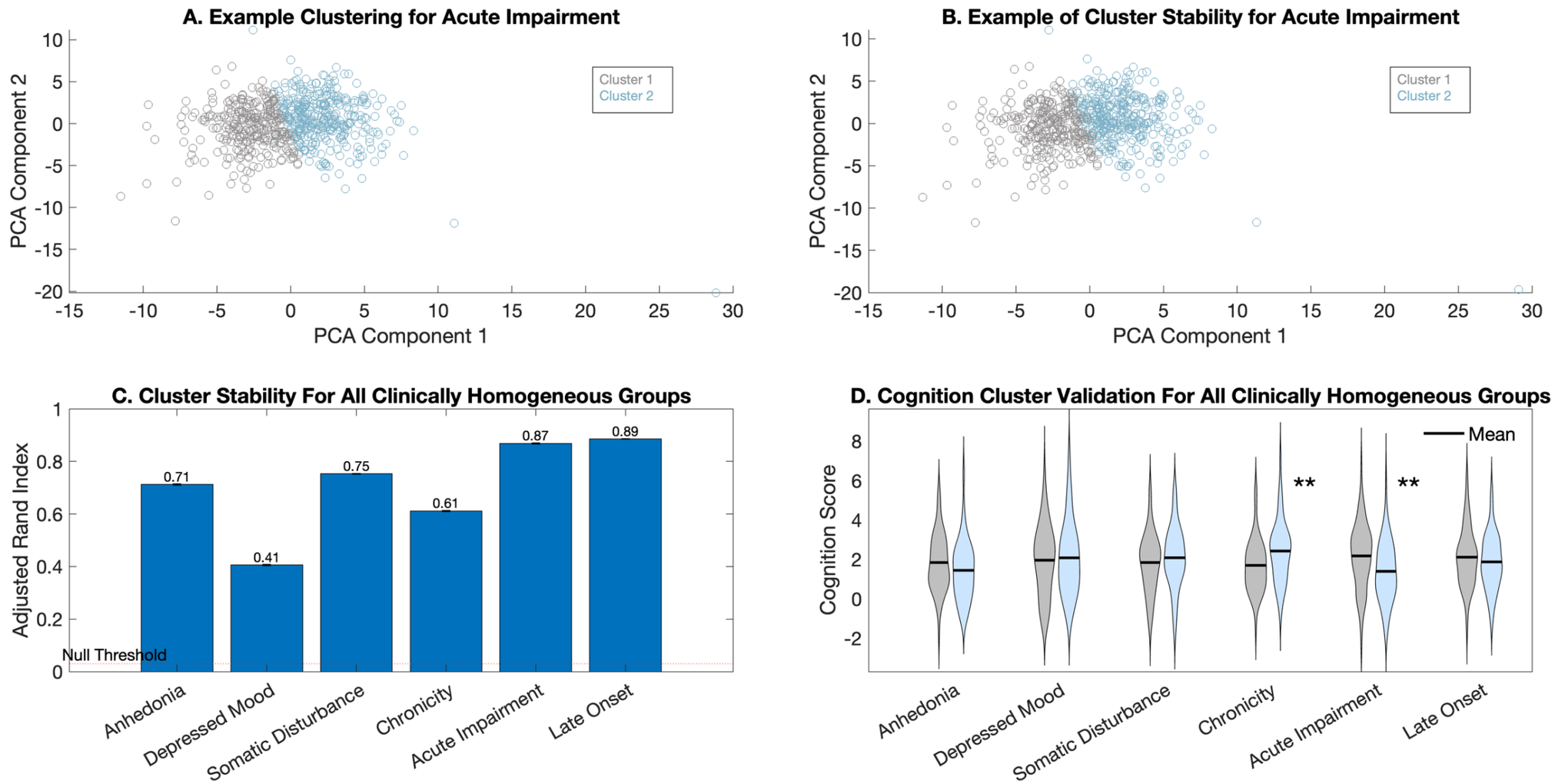

Figure 4. Stable and reproducible data-driven clusters within clinically homogeneous groups. (A). *Visualized are the first two PCA components (selected for visualization purposes) and the cluster assignment for one representative cross-validation fold in the acute impairment group.* ***(B).*** *Results from a second representative cross-validation fold to visually indicate cluster stability. Each participant is shown with the same color as in A, but the PCA scores are recomputed using data from the new cross-validation fold.* ***(C).*** *The Adjusted Rand Indices, which indicates similarity in clustering results across cross-validation folds adjusted for random chance, for the clusters of each clinically homogeneous group. The null threshold was obtained from permuted data, and error bars indicate the standard error of mean for each group.* ***(D).*** *The clusters were further validated by investigating differences in cognitive ability between clusters. The clusters within the chronicity and acute impairment groups differed significantly in cognition (** denotes <0.01 post false discovery rate correction across the 6 comparisons).*

Discussion

In this study we aimed to jointly disentangle clinical and neurobiological sources of depression heterogeneity, investigating the presence of one-to-one and many-to-one brain-symptom mapping. Firstly, we showed that each clinically isolated feature was associated with distinct neuroimaging correlates, providing symptom-level insights into the neurobiological etiology of depression. Secondly, our findings revealed often larger magnitude of neuroimaging effects within clinically homogeneous groups compared to a 'typical' heterogeneous comparison group, pointing to the impact of heterogeneity in prior studies as an obfuscating factor. Thirdly, we revealed evidence for neurobiological clusters that differed in cognitive ability within two clinically homogeneous groups (chronicity and acute impairment). To our knowledge, this is the first evidence indicative of the presence of many-to-one mapping, suggesting that multiple distinct neurobiological mechanisms may give rise to the same clinical presentation.

Our novel approach allowed us to derive unique insights into differential imaging profiles of depression by dissociating symptoms and other clinical features of severity and age of onset, that are typically highly colinear in conventional depression study cohorts. Several of the imaging features showed normative deviations in opposite directions across different clinically homogeneous groups (Fig. 3B), which supports the previously posited idea that research findings differ based on what clinical feature is overrepresented by chance. For example, the depressed mood group showed increased connectivity between the ventral and POS2 DMN and decreased connectivity between the left CEN and the anterior SN, while the anhedonia group showed increased connectivity between the left CEN and the anterior SN. As rumination is a key characteristic in depressed mood[25], these results are consistent with prior evidence suggesting that rumination is driven by increased Default Mode Network connectivity, which fails to be suppressed by the CEN and SN when their connectivity is decreased[26,27]. Anhedonia is a construct commonly juxtaposed to rumination[26] and showed increased (rather than decreased) connectivity between the left CEN and the anterior SN. As such, these findings may suggest that previous inconsistencies in the literature with evidence for both increased and decreased connectivity associated with depression[28] may be due to overrepresentation of different clinical features in different study samples.

Another example of an imaging feature showing normative deviations in the opposite direction than the literature would suggest was the observed positive cortical thickness findings of the late onset group (Fig. 3B). Although late onset depression is typically associated with cortical thinning[29,30], research has shown that remittance in late onset depression was associated with relatively larger cortical thickness[31]. As our late onset group deliberately contains individuals who are currently not high in depression symptoms to achieve the desired dissociation of clinical features, our findings are indeed consistent with prior literature pointing to larger cortical thickness as a protective mechanism linked to remission[31]. Similarly, one might speculate that our finding of increased gray matter volumes of the cingulate gyrus in the depressed mood group may be indicative of another protective mechanism. As the cingulate gyrus is important in emotional regulation[32], individuals in our depressed mood group may be protected from more extensive symptoms of depression due to their larger cingulate gyri. Although this hypothesis requires further testing, such a protective effect could explain the increase in cingulate volume in our depressed mood group, contrary to prior findings of decreased cingulate volume in conventional depression study cohorts[33].

More broadly, our results showed larger magnitudes of neuroimaging effects in many of the clinically homogenous groups as compared to the comparison heterogeneous group, which supports the idea posited in the literature that heterogeneity is leading to smaller effect sizes. In the age of population neuroimaging datasets, this finding points to the value of carefully parsing clinical heterogeneity to investigate one-to-one brain-symptom modeling and identify robust and replicable neuroimaging correlates.

Moving beyond one-to-one brain-symptom mapping, our results revealed the presence of residual neurobiological heterogeneity, supporting the many-to-one brain-symptom mapping model. We found highly

stable neuroimaging-driven participant clusters that differed on cognitive ability in two of our clinically homogenous groups: chronicity and acute impairment, which indicates the presence of many-to-one mapping in these groups. We performed extensive stability and null testing to rigorously verify our clusters. To our knowledge this was the first time that clinical and neurobiological sources of heterogeneity were distinguished and rigorously tested.

In the context of addressing heterogeneity in mental health, the literature has often distinguished between clustering approaches that aim to define distinct subgroups of patients[34] and dimensional approaches that aim to identify one or more continuous axes of patient variation[20,35]. Indeed, recent work sought to directly compare cluster-based and dimensional representations of heterogeneity in a psychosis patient sample[36]. Importantly, our methodology and results bridge across these characterizations of heterogeneity because we perform k-means clustering on a dimensional PCA embedding of imaging features. Although a direct quantitative comparison of our clusters against a dimensional solution was not feasible due to the invalidity of Akaike Information Criterion (AIC) and other measures of model fit in the context of k-means clustering (as k-means does not estimate the fit of a likelihood function), we note that the resulting clusters were largely driven by the first PCA (Fig. 4A). These findings suggests that dimensional and cluster-based representations of heterogeneity offer complementary rather than mutually exclusive approaches, that can be productively combined to explain patient variability while also informing clinical practice. Specifically, clustering can offer meaningful cut-offs – even within an underlying dimensional distribution – that may be more pragmatically useful to inform clinical decision-making than dimensions (as long as the cut-offs are stable and clinically relevant). Another important distinction is that we are not investigating nor claiming the presence of discrete (i.e., highly separable) clusters. As noted in the introduction, a prominent example of clinically informative cut-offs along a dimensional distribution are the categories (or 'clusters') of subclinical, mild, moderate, and severe depression along the continuum of depression severity. Notably, cut-offs within a relatively continuous patient distribution can make validation based on external clinical phenotypes more challenging because average differences between clusters are relatively small. Nevertheless, we validated two of our cluster solutions (within chronicity and acute impairment groups) based on cognitive measures, which may indicate differential etiological origin and comorbid disease progression[37].

The current study, while aided by rich data of the UK Biobank, still had key limitations for future work to build on. We decided to tightly control the clinically homogeneous groups to isolate individual clinical features, which was important to gain insights into sources of heterogeneity and to identify distinct neuroimaging patterns. However, the specific inclusions/exclusion criteria of our clinically homogenous groups focus on relatively atypical patients because the majority of patients with depression have a more mixed clinical presentation. Future directions for this work include investigating the neuroimaging correlates of homogeneous groups with a more mixed clinical presentation to test whether the one-to-one brain-symptom mapping model extends to more complex and multivariate clinical profiles. Importantly, as we started with relatively atypical patients in our clinically homogeneous groups, the goal of this study was not to suggest that the data driven clusters are final subtypes of depression, but rather point to the importance of accounting for both clinical and neurobiological heterogeneity when investigating depression. Furthermore, this study focused on interactions between clinical and neurobiological heterogeneity, but future work may want to extend our approach into other sources of heterogeneity. For example, recent work[38] has highlighted the importance of genetic variation in depression. Future efforts are also needed to further validate the neurobiological clusters identified in this work. The number of participants with complete data on clinically relevant phenotypes for validation was relatively small, which limited the number of phenotypes we could use for validation. Although our findings provide preliminary support for the presence of clinically relevant differences between neurobiologically driven clusters, future work could investigate more validation of phenotypes. For example, future validation of neurobiological

clusters may investigate differences in treatment response based on these clinical and neurobiological groups, which could be particularly valuable as treatment prediction is a major clinical challenge.

In summary, our results provide insights into the symptom-level neurobiological etiology of depression, establishing and investigating one-to-one and many-to-one models of heterogeneity while addressing major methodological challenges of reproducibility. Our results suggest the importance for future studies to account for clinical and neurobiological heterogeneity when trying to understand depression etiology and neurobiological mechanisms.

Methods

Dataset: This study utilizes the UK Biobank, a cohort rich in clinical data including depression questionnaires as well as neuroimaging features known as imaging derived phenotypes (IDPs). Participants in our study are identified based on their self-response on the Recent Depressive Symptoms Questionnaire (RDS, Table 1B). RDS was chosen as it was the only depression measure taken on day of scan. It has been previously validated against other common depression questionnaires in Dutt & Hannon et al 2022[8]. The imaging features investigated in this study were chosen as they have been previously associated with depression. These imaging features are imaging derived phenotypes (IDPs) gray matter volume, cortical thickness, white matter hyperintensity, fractional anisotropy, and functional networks in regions described in Table 1B. All imaging data was acquired on a 3T scanner. Gray matter volume and cortical thickness were obtained from T1-weighted images. White matter hyperintensity is from T2 fluid-attenuated inversion recovery (FLAIR) acquisition. Fractional anisotropy and functional networks are from diffusion MRI and functional MRI acquisition, respectively. Data processing for all neurobiological features has been described in previous work[39,40]. The UK Biobank IDs for these imaging measures along with all other variables are in Supplemental Table 2.

A.

Gray Matter Volume	*Cortical Thickness*
Medial & Orbital Frontal Cortex	Superior Frontal Cortex
Anterior & Posterior Cingulate Gyrus	Superior & Inferior Parietal Cortex
Anterior & Posterior Superior Temporal Gyrus	Fusiform Cortex
Anterior, Posterior, & Temporo-occipital Middle Temporal Gyrus	Entorhinal Cortex
Insular Cortex	Parahippocampal Cortex
Hippocampus	*Fractional Anisotropy*
Amygdala	Cingulum Bundle at Cingulate Gyrus
Thalamus	Cingulum Bundle at Hippocampus
Ventral Striatum	Splenium, Genu, & Body of Corpus Callosum
Functional Networks	Uncinate Fasciculus
Within network amplitudes & partial between-network correlations:	Superior Longitudinal Fasciculus
Dorsal, Ventral, & POS2 Default Mode Network	Anterior, Posterior, & Superior Corona Radiata
Right & Left Central Anterior & Posterior Executive Network	Anterior Limb of Internal Capsule
Salience Network	*White Matter Hyperintensity*
	Total volume

B.

RDS Questions	RDS Score System
RDS1. Depressed mood in last 2 weeks	(1) Not at all
RDS2. Unenthusiasm/disinterest mood in last 2 weeks	(2) Several days
RDS3. Tenseness/restlessness mood in last 2 weeks	(3) More than half the days
RDS4. Tiredness/lethargy mood in last 2 weeks	(4) Nearly every day

Table 1. Imaging Features and Depression Questionnaire used in study. (A). *The neurobiological imaging features included in this study as they are related to depression, POS2 = parietal-occipital sulcus area 2.* ***(B).*** *Recent Depressive Symptoms (RDS) measure used to quantify depression symptoms in this study, sum range 4-16.*

Clinically homogeneous groups: Due to indications in the literature that they may exhibit distinct neuroimaging profiles, six clinical features were investigated: symptoms of anhedonia, depressed mood, and somatic disturbance; severity indices such as lifetime chronicity and acute impairment; and clinical subgroups such as late onset depression. Anhedonia, somatic disturbance, and depressed mood have consistently been found as symptom-based subtypes[13–16]. Anhedonia (inability to feel pleasure) has been linked to abnormality in Salience (Sal) and Central Executive Networks (CEN)[41] as well as decreased gray matter volume (GMV) of the amygdala[42,43] and anterior cingulate gyrus (ACC)[44]. Somatic disturbance (increased or decreased psychomotor activity) is believed to be linked to abnormal activity in the Salience Network[45]. Depressed mood (defined as low mood) is associated with reduced volume in frontal regions such as medial frontal cortex and ACC[33]. Severity is another important feature of depression, which takes two possible forms. Trait depression (i.e.,

baseline predisposition) leads to numerous episodes of depression throughout the lifetime which we refer to as chronic severity or chronicity. On the other hand, acute severity refers to state depression (being in a current state of severe depression). High chronicity has been associated with a reduction in hippocampal volume[10,46]. Acute severity, on the other hand, may be associated with more state-dependent changes such as abnormal functional connectivity, particularly in Default Mode Network (DMN).[47] The final clinical feature relates to age of first depression episode onset. When depression episodes begin in late life (60 years or older) it is referred to as late onset. Late onset patients are more likely to have reduced fractional anisotropy (FA)[48] and white matter hyperintensities (WMH)[49] even after accounting for age. We decided to include this subgroup as the UK Biobank cohort skews older. Notably, all clinical features of symptoms, severity, and age of onset are highly collinear in typical depression cohorts. To our knowledge, our study is the first to select participants in order to dissociate these clinical features of depression and allow robust investigation into distinct neuroimaging correlates. The prior neuroimaging findings summarized above were used to guide the selection of our imaging features.

A

	Sample Size	# of Episodes	RDS Score	Age Onset (yrs)
Late Onset	636	<=10	Sum <=5	>=60
Chronicity	343	>=10	Sum <=5	<60
Acute Impairment	529	<=10	Sum >=10	<60
Depressed Mood	279	<=10	RDS1 > 2, RDS2-3 <=2	<60
Anhedonia	313	<=10	RDS2 >2, RDS1,3 <=2	<60
Somatic Disturbance	336	<=10	RDS3 >2, RDS1-2 <=2	<60
Heterogenous	4216	>=1	Sum >=5	Any
Healthy Control	8565	0	Sum = 4	Never

B

	Age (mean±SD)	Sex (% male)	RDS Sum	RDS Anhedonia	RDS Mood	RDS Restless	RDS Lethargy	Age Onset	Number of episodes
Anhedonia	62.6±7.6	66.8	8.3	3.4	1.4	1.4	2.2	12.8	1
Depressed Mood	61.4±7.5	37.3	8.7	1.6	3.2	1.5	2.3	15.3	2
Somatic Disturbance	61.0±8.0	46.1	8.6	1.4	1.4	3.3	2.4	14.2	2
Chronicity	63.6±7.5	45.8	4.5	1.0	1.0	1.1	1.3	21.6	33
Acute Impairment	60.2±7.2	22.6	12.0	2.9	3.0	2.7	3.4	13.8	2
Late Onset	69.8±4.4	47.6	4.4	1.0	1.0	1.0	1.3	63.9	1
Heterogenous	61.6±7.5	35.6	6.2	1.3	1.4	1.4	2.0	34.7	6
Healthy Control	64.7±7.2	58.4	4	1	1	1	1	N/A	N/A

Table 2. Clinically Homogeneous Group Information. (A). *Inclusion criteria for clinically homogeneous groups.* ***(B).*** *Demographics and Clinical Characteristics. Means if not otherwise stated. RDS = Recent Depressive Symptoms measure (sum range 4-16), RDS# = RDS question number (see Table 1), range 1-4*

Clinically homogenous groups are comprised of individuals who share their clinical features but do not have any of the other clinical features investigated. For example, the individuals in the anhedonia group scored high in

anhedonia but low in depressed mood and somatic disturbance, were not late onset (first episode of depression at 60 years or older) nor had a high number of reported depressive episodes in their lifetime. The late onset group is comprised of individuals who experienced their first episode of depression at 60 years or older but are not currently experiencing high symptoms nor had more than 10 episodes in their lifetime. Isolation of these rare individuals is uniquely enabled in our work by leveraging a large population dataset. Further inclusion details are shown in Table 2A. Clinical characterization (like RDS score, age of onset, and number of lifetime episodes) and demographic information like age and sex of each group are in Table 2B. Symptoms and severity were measured based on response to the RDS questionnaire. Age of onset was determined by report of first episode of depression (UKB variable ID 20433, using instance 0 data). Chronicity was defined by reported number of episodes (UKB variable ID 4620 and 5386, using instance 2 data). **Imaging features**: We investigated 88 imaging features of the clinically homogenous groups based on features that literature has shown to be relevant to depression (Table 1A). These imaging features are Imaging Derived Phenotypes from the UK Biobank whose exact IDs are in Supplemental Table 2.

Normative modeling: To be highly sensitive to biologically meaningful variation, we employed normative modeling on the imaging features to acquire "normative deviations" that represent how each person differs from the normal population for each imaging measure, accounting for covariate effects. Normative modeling has successfully been applied to parse heterogeneity in other mental disorders[50–52].The normative model applied is SinhArcsinh-warped Bayesian linear regression[53] with covariates age (UKB variable ID 21003, instance 2), sex (UKB variable ID 31, instance 2), head motion (UKB variable ID 25741, instance 2), and image acquisition site (UKB variable ID 54, instance 2). This algorithm was selected as previous research found it represented neuroimaging features most accurately of all algorithms assessed[53]. More precisely, the normative model is a b-spline basis expansion over the covariates (age, sex, head motion, acquisition site) with equally spaced knot points employing a SinhArcsinh warp to model non-linearity in the imaging feature (the response variable in this model). The prior weight precision, noise precision, and warp shape parameters were controlled by model hyper parameters which were estimated using an Empirical Bayes approach. The normative deviation is a Z-statistic of the model's residuals, measuring how much the individual's imaging feature deviated from the model's expected value based on their covariates. These z-scores were calculated by subtracting the true imaging value from the estimated prediction (after warping), divided by the standard deviation accounting for all estimated variance components, as described in previous publications[54]. Normative modeling accuracy statistics are in Supplemental Table 5.

Statistical comparison of normative deviations across clinically homogenous groups: To determine the imaging features that are significantly different between clinically homogeneous groups, we applied two-sided one-way ANOVA with false discovery rate (FDR) multiple comparison correction over 88 imaging feature comparisons. To further test whether the clinically homogeneous groups differed in terms of within-group variability, we measured homogeneity of the imaging features of the clinically homogeneous groups using c-k intraclass correlation (ICC) on the imaging features of each clinically homogenous group. C-k ICC measures the degree of consistency for averages of k participants made of the fixed levels of c imaging features[55], also sometimes referred to as Cronbach's alpha. We also assessed the replicability of the significant imaging features results by repeating the ANOVA over 100 bootstraps (randomly sampling the participants with replacement), accounting for FDR correction within each bootstrap (see Supplemental Figure 1). We also performed a two-sided t-test against 0 for each imaging feature for each clinically homogeneous group, see Supplemental Figure 4 for results.

		N	Age (mean±SD)	Sex (% male)	Depression score (4-16)
Anhedonia	Cluster 1	196	62.2±7.4	44.9	8.3
	Cluster 2	117	63.3±7.9	47.9	8.4
Low Mood	Cluster 1	168	60.6±7.6	35.7	8.8
	Cluster 2	111	62.4±7.2	39.6	8.6
Somatic Disturbance	Cluster 1	212	61.4±7.6	47.6	8.4
	Cluster 2	124	60.2±8.5	43.5	8.8
Chronic	Cluster 1	176	63.7±7.3	41.5	4.5
	Cluster 2	167	63.4±7.8	50.3	4.4
Late Onset	Cluster 1	305	69.5±4.5	50.5	4.4
	Cluster 2	224	70.2±4.2	43.8	4.3
Acute Impairment	Cluster 1	344	60.0±6.9	29.1	12.1
	Cluster 2	292	60.4±6.9	34.2	11.9

Table 3. Demographic information for the clusters within each clinically homogeneous group.

Data-driven clustering within clinically homogenous groups: We performed clustering to test for the possibility of many-to-one mapping within clinically homogeneous groups (Fig. 4B). Feature reduction was performed using principal component analysis (PCA) applied to 88 imaging features keeping the minimum number of principal components that explain at least 75% of the variance as features in subsequent k-means clustering. This procedure and the below steps were performed independently within each clinically homogeneous group.

Identifying the optimal number of clusters: We measured the silhouette score for *k* from 2 through 10 to determine how many clusters is best. We bootstrapped this data 1000 times (randomly sampling the data with replacement) and determined optimal cluster number as the *k* that has the highest silhouette score the most times over the 1000 bootstraps. We performed the PCA, feature selection, and k-means clustering pipeline for each bootstrap. The feature selection (selecting the components that explained 75% of the variance) was highly consistent: the first 32 components explained 75% of the variance for every bootstrap.

Cluster stability testing: Once we determined the optimal cluster number for each clinically homogeneous groups, we assessed the stability of the clusters. To do this, we performed 10-fold cross validation, which was repeated over 100 iterations. In each fold, 90% of the participants were used to repeat the PCA, selection of PCA components (32 components), and k-means clustering (k=2) at the found dimensionality. We calculated the adjusted Rand index (ARI) between each fold, which provides a measure of the stability of the cluster assignments. For each fold, we trained the PCA and k-means on 90% of data, then applied the PCA coefficients and k-means centroids to the remaining 10% culminating in cluster assignments for all 100% of data for that fold. In that iteration, the ARIs between cluster assignments for all 10 folds were computed. The mean ARI for that entire iteration was then calculated. This process was repeated for 100 iterations (with new folds every time) and the mean ARIs for each iteration were averaged (and STD calculated) one final time to get the final mean ARI and STD shown in Fig. 4C.

Null model: We synthesized a null dataset by randomly permuting the participant labels 1000 times independently within each imaging feature (i.e., breaking the correlation between imaging features) within each

clinically homogenous group. The resulting null data were processed through the pipeline described above (including PCA, component selection, and cross-validation performed separately within each clinically homogeneous group), which resulted in a set of ARI thresholds based on null data. The 95% confidence interval for the ARIs of the null data of all clinically homogeneous groups was (0.030,0.031).

Cluster validation: We validated the clusters by investigating whether any clusters significantly differed on the following depression-relevant phenotypes: cognitive ability (quantified using the PCA weights calculated in Lyall et al 2016[21] that utilized variables 20016 and 20023), Neuroticism (see Supplemental Table 2 for variable IDs), and Townsend socioeconomic deprivation index (UKB variable 189). We conducted paired t-tests between the clusters within each clinically homogenous group, using false discovery rate correction for the 6 comparisons. Cognitive ability was selected as it is highly clinically pertinent to depression, particularly to the older adults overrepresented in the UK Biobank[56]. Neuroticism was selected as it is a strong correlate of depression, considered to be indicative of "trait-like" depression[57]. Finally socioeconomic deprivation was chosen as it has been shown to be an important modulator of depression[58].

Acknowledgements
Janine Bijsterbosch is supported by the NIH (R34 NS118618; R01 MH128286) and the McDonnell Center for Systems Neuroscience.

Data Availability
This research was performed under UK Biobank application number 47267. UK Biobank data (Miller et al., 2016; Sudlow et al., 2015) is available following an access application process. For more information: https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access.

Code Availability
All analysis code for this article is available at: https://github.com/PersonomicsLab/Heterogeneity.

References
1. Wang, J. *et al.* Prevalence of depression and depressive symptoms among outpatients: a systematic review and meta-analysis. *BMJ Open* **7**, e017173 (2017).
2. VanItallie, T. B. Subsyndromal depression in the elderly: underdiagnosed and undertreated. *Metabolism* **54**, 39–44 (2005).
3. Khan, A. & Brown, W. A. Antidepressants versus placebo in major depression: an overview. *World Psychiatry* **14**, 294–300 (2015).
4. Zeng, L.-L. *et al.* Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. *Brain* **135**, 1498–1507 (2012).
5. Kennis, M. *et al.* Prospective biomarkers of major depressive disorder: a systematic review and meta-analysis. *Mol Psychiatry* **25**, 321–338 (2020).
6. Saberi, A., Mohammadi, E., Zarei, M., Eickhoff, S. B. & Tahmasian, M. Structural and functional neuroimaging of late-life depression: a coordinate-based meta-analysis. *Brain Imaging and Behavior* **16**, 518–531 (2022).
7. Greene, A. S. *et al.* Brain–phenotype models fail for individuals who defy sample stereotypes. *Nature* **609**, 109–118 (2022).
8. Dutt, R. K. *et al.* Mental health in the UK Biobank: A roadmap to self□report measures and neuroimaging correlates. *Human Brain Mapping* **43**, 816–832 (2022).
9. Marek, S. *et al.* Reproducible brain-wide association studies require thousands of individuals. *Nature* (2022) doi:10.1038/s41586-022-04492-9.
10. Sheline, Y. I., Liston, C. & McEwen, B. S. Parsing the Hippocampus in Depression: Chronic Stress, Hippocampal Volume, and Major Depressive Disorder. *Biological Psychiatry* **85**, 436–438 (2019).
11. Woody, M. L. & Gibb, B. E. Integrating NIMH Research Domain Criteria (RDoC) into depression research. *Current Opinion in Psychology* **4**, 6–12 (2015).
12. Kotov, R. *et al.* The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J. Abnorm. Psychol.* **126**, 454–477 (2017).
13. van Loo, H. M., de Jonge, P., Romeijn, J.-W., Kessler, R. C. & Schoevers, R. A. Data-driven subtypes of major depressive disorder: a systematic review. *BMC Med* **10**, 156 (2012).
14. Hybels, C. F., Blazer, D. G., Landerman, L. R. & Steffens, D. C. Heterogeneity in symptom profiles among older adults diagnosed with major depression. *Int. Psychogeriatr.* **23**, 906–922 (2011).

15. Bracht, T., Linden, D. & Keedwell, P. A review of white matter microstructure alterations of pathways of the reward circuit in depression. *Journal of Affective Disorders* **187**, 45–53 (2015).
16. Chen, L.-S., Eaton, W. W., Gallo, J. J. & Nestadt, G. Understanding the heterogeneity of depression through the triad of symptoms, course and risk factors: a longitudinal, population-based study. *Journal of Affective Disorders* **59**, 1–11 (2000).
17. Drysdale, A. T. *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* **23**, 28–38 (2017).
18. Dinga, R. *et al.* Evaluating the evidence for biotypes of depression: Methodological replication and extension of Drysdale et al. (2017). *NeuroImage: Clinical* **22**, 101796 (2019).
19. Cicchetti, D. & Rogosch, F. A. Equifinality and multifinality in developmental psychopathology. *Dev Psychopathol* **8**, 597–600 (1996).
20. Wen, J. *et al.* Characterizing Heterogeneity in Neuroimaging, Cognition, Clinical Symptoms, and Genetics Among Patients With Late-Life Depression. *JAMA Psychiatry* **79**, 464 (2022).
21. Lyall, D. M. *et al.* Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. *PLoS ONE* **11**, e0154222 (2016).
22. Bauermeister, S. & Gallacher, J. *A psychometric evaluation of the 12-item EPQ-R neuroticism scale in 502,591 UK Biobank participants using item response theory (IRT)*. https://www.researchsquare.com/article/rs-13803/v1 (2020) doi:10.21203/rs.2.23234/v1.
23. Foster, H. M. E. *et al.* The effect of socioeconomic deprivation on the association between an extended measurement of unhealthy lifestyle factors and health outcomes: a prospective analysis of the UK Biobank cohort. *The Lancet Public Health* **3**, e576–e585 (2018).
24. Perini, G. *et al.* Cognitive impairment in depression: recent advances and novel treatments. *NDT* **Volume 15**, 1249–1258 (2019).
25. Brinker, J. K. & Dozois, D. J. A. Ruminative thought style and depressed mood. *J. Clin. Psychol.* **65**, 1–19 (2009).
26. Williams, L. M. Precision psychiatry: a neural circuit taxonomy for depression and anxiety. *The Lancet Psychiatry* **3**, 472–480 (2016).
27. Goulden, N. *et al.* The salience network is responsible for switching between the default mode network and the central executive network: Replication from DCM. *NeuroImage* **99**, 180–190 (2014).
28. Wang, L., Hermens, D. F., Hickie, I. B. & Lagopoulos, J. A systematic review of resting-state functional-MRI studies in major depression. *Journal of Affective Disorders* **142**, 6–12 (2012).
29. Mackin, R. S. *et al.* Patterns of Reduced Cortical Thickness in Late-Life Depression and Relationship to Psychotherapeutic Response. *The American Journal of Geriatric Psychiatry* **21**, 794–802 (2013).
30. Lim, H. K. *et al.* Regional Cortical Thickness and Subcortical Volume Changes Are Associated with Cognitive Impairments in the Drug-Naive Patients with Late-Onset Depression. *Neuropsychopharmacol* **37**, 838–849 (2012).
31. Motter, J. N. *et al.* Cortical thickness predicts remission of depression with antidepressants in patients with late-life depression and cognitive impairment. *Journal of Affective Disorders* **295**, 438–445 (2021).
32. Posner, M. I., Rothbart, M. K., Sheese, B. E. & Tang, Y. The anterior cingulate gyrus and the mechanism of self-regulation. *Cognitive, Affective, & Behavioral Neuroscience* **7**, 391–395 (2007).
33. Killgore, W. D. S. & Yurgelun-Todd, D. A. Ventromedial prefrontal activity correlates with depressed mood in adolescent children: *NeuroReport* **17**, 167–171 (2006).
34. Kaczkurkin, A. N. *et al.* Neurostructural Heterogeneity in Youths With Internalizing Symptoms. *Biological Psychiatry* **87**, 473–482 (2020).

35. Kaczkurkin, A. N. *et al.* Evidence for Dissociable Linkage of Dimensions of Psychopathology to Brain Structure in Youths. *AJP* **176**, 1000–1009 (2019).
36. Smucny, J. *et al.* Latent Profiles of Cognitive Control, Episodic Memory, and Visual Perception Across Psychiatric Disorders Reveal a Dimensional Structure. *Schizophrenia Bulletin* **46**, 154–162 (2020).
37. Richard, E. *et al.* Late-Life Depression, Mild Cognitive Impairment, and Dementia. *JAMA Neurol* **70**, 383 (2013).
38. The Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* **18**, 199–209 (2015).
39. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* **19**, 1523–1536 (2016).
40. Alfaro-Almagro, F. *et al.* Confound modelling in UK Biobank brain imaging. *NeuroImage* **224**, 117002 (2021).
41. Harvey, P.-O., Pruessner, J., Czechowska, Y. & Lepage, M. Individual differences in trait anhedonia: a structural and functional magnetic resonance imaging study in non-clinical subjects. *Mol Psychiatry* **12**, 767–775 (2007).
42. Gibson, K. *et al.* Relations Between Subcortical Volumes, Anhedonia Severity, and Reward-Oriented Behaviors in an Anhedonic Transdiagnostic Sample. *Biological Psychiatry* **89**, S277 (2021).
43. Gruenbaum, S. E. *et al.* Inhibition of glutamine synthetase in the central nucleus of the amygdala induces anhedonic behavior and recurrent seizures in a rat model of mesial temporal lobe epilepsy. *Epilepsy & Behavior* **51**, 96–103 (2015).
44. Liu, X., Li, L., Li, M., Ren, Z. & Ma, P. Characterizing the subtype of anhedonia in major depressive disorder: A symptom-specific multimodal MRI study. *Psychiatry Research: Neuroimaging* **308**, 111239 (2021).
45. Zhang, T. *et al.* Abnormal connectivity of anterior-insular subdivisions and relationship with somatic symptom in depressive patients. *Brain Imaging and Behavior* **15**, 1760–1768 (2021).
46. Sheline, Y. I. 3D MRI studies of neuroanatomic changes in unipolar major depression: the role of stress and medical comorbidity. *Biological Psychiatry* **48**, 791–800 (2000).
47. Tozzi, L. *et al.* Reduced functional connectivity of default mode network subsystems in depression: Meta-analytic evidence and relationship with trait rumination. *NeuroImage: Clinical* **30**, 102570 (2021).
48. Cheng, Y. *et al.* Delineation of Early and Later Adult Onset Depression by Diffusion Tensor Imaging. *PLoS ONE* **9**, e112307 (2014).
49. Salo, K. I., Scharfen, J., Wilden, I. D., Schubotz, R. I. & Holling, H. Confining the Concept of Vascular Depression to Late-Onset Depression: A Meta-Analysis of MRI-Defined Hyperintensity Burden in Major Depressive Disorder and Bipolar Disorder. *Front. Psychol.* **10**, 1241 (2019).
50. Andrews, D. S., Marquand, A., Ecker, C. & McAlonan, G. Using Pattern Classification to Identify Brain Imaging Markers in Autism Spectrum Disorder. in *Biomarkers in Psychiatry* (eds. Pratt, J. & Hall, J.) vol. 40 413–436 (Springer International Publishing, 2018).
51. Verdi, S. *et al.* The heterogeneous brain: Mapping individualised patterns of atrophy in Alzheimer's disease using spatial normative models. *Alzheimer's & Dementia* **17**, (2021).
52. Wolfers, T. *et al.* Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. *JAMA Psychiatry* **75**, 1146 (2018).
53. Fraza, C. J., Dinga, R., Beckmann, C. F. & Marquand, A. F. Warped Bayesian linear regression for normative modelling of big data. *NeuroImage* **245**, 118715 (2021).

54. Rutherford, S. *et al.* The normative modeling framework for computational psychiatry. *Nat Protoc* **17**, 1711–1734 (2022).
55. McGraw, K. O. & Wong, S. P. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* **1**, 30–46 (1996).
56. Butters, M. A. *et al.* Pathways linking late-life depression to persistent cognitive impairment and dementia. *Dialogues in Clinical Neuroscience* **10**, 345–357 (2008).
57. Vittengl, J. R. Who pays the price for high neuroticism? Moderators of longitudinal risks for depression and anxiety. *Psychol. Med.* **47**, 1794–1805 (2017).
58. Assari, S. Social Determinants of Depression: The Intersections of Race, Gender, and Socioeconomic Status. *Brain Sciences* **7**, 156 (2017).