Elucidating depression heterogeneity using clinical, neuroimaging and genetic data

Luigi Angelo Maglanoc

Dissertation for the degree of Philosophiae Doctor (Ph.D)

Department of Psychology

Faculty of Social Sciences

University of Oslo

2019
Acknowledgments

I have worked in some capacity or another with cognitive neuroscience since 2011. Finishing this PhD is bittersweet, because I will not continue working with neuroscience and/or in academia (at least for now). The (paraphrased) quote by Daniël Lakens is quite apt: “If not me, then someone else.” Luckily, UiO is full of passionate, hard-working and talented researchers, so the one who is missing out is likely me. And the future is bright, as I have seen first-hand the new wave of undergraduate and graduate students.

Although the PhD dissertation itself is a sole endeavour, a long list of people and groups have been directly or indirectly involved over the years that lead to its fruition. My PhD journey was contingent on the funding from UiO and the psychology department. I have always been aware of how fortunate we are here in Norway as PhD candidates compared to other countries, especially in terms of receiving a respectable salary. The three papers themselves were possible thanks to the funding from the Research Council of Norway and the Regional Ethical Committee of South East Norway for funding the research, and my co-authors for making them come to fruition. For paper II, I thank the UK Biobank for letting me access and use their data. What an amazing resource that has and will have a large impact in the life sciences. Thanks to my co-supervisor Nils Inge and the rest of the Clinical Neuroscience Research group: Rune, Eva and Brage, with whom I started my PhD with. The multimodal imaging group, Ole, Christine and the rest of NORMENT have accommodated me over my extremely slow (self-inflicted) transition. It was truly a pleasure to see all the fantastic work that was being done there and will continue for years to come. Thanks to the administration (including IT) at the psychology department and USIT. It goes without saying that the all the research participants deserve my utmost gratitude. I have met almost all of you in person, and this thesis, and neuroscience/psychology research would not be possible without you.

My family has always been supportive and believed in me even when I did not, and for that, I will always be grateful. I thank Dani for being very involved in the data collection for papers I and III, especially for the MRI-scanning sessions (which involved many Fridays and weekends). I thank my fellow office mates at the psychology department: Trine, Cecilie, Unni and Erlend (to name a few), and Kristine at NORMENT for random laughs that made
the days enjoyable. Linn gets special thanks for her positivity that made the last few months fun despite the workload and stress.

Tobias has been particularly important in the culmination of the three research papers. I have learned a lot from working with him and am extremely grateful for all the help and feedback he has given to me (and many times over weekends). NORMENT and UiO are truly lucky to have such a competent and dedicated researcher.

The most instrumental person for this PhD, and in fact, my (short) academic career in neuroscience since my master’s thesis is Lars. He is the main reason that I have gotten to work with and learn advanced methods, that were among the key reasons I got my new job. What a fantastic leader that manages to find time to supervise X number of PhD candidates, Y number of postdocs, be involved with Z number of projects, and still have a huge role in doing the writing and analyses in studies (and being a parent). The countless revisions he has made to my three PhD papers has definitely improved their quality. To this day, I still have no idea how he manages all of this.

For life in general, the most important person is without a doubt my fiancée. Thank you for being you, being my better half, supporting me and putting up with me all these years, especially in my tough times. Without you I would not have been able to complete this PhD thesis, let alone much else.

And a final special thanks to my cat for being fluffy.
# Table of Contents

1 GENERAL SUMMARY ........................................................................................................ 1

2 LIST OF PAPERS ............................................................................................................ 3

3 LIST OF ABBREVIATIONS ........................................................................................... 4

4 INTRODUCTION............................................................................................................ 7
  4.1 Depression diagnosis and clinical characteristics .............................................. 7
    4.1.1 Clinical depression heterogeneity ................................................................. 9
  4.2 Depression and the brain ................................................................................. 11
    4.2.1 Morphometric MRI studies ........................................................................... 11
    4.2.2 DTI studies ..................................................................................................... 12
    4.2.3 Resting-state fMRI studies .......................................................................... 13
    4.2.4 Depression heterogeneity is reflected in the brain ...................................... 15
  4.3 Polygenetic architecture of depression ............................................................ 17
  4.4 Depression overlap with anxiety and neuroticism ............................................ 19
    4.4.1 Anxiety ......................................................................................................... 19
    4.4.2 Neuroticism ................................................................................................... 20

5 MAIN RESEARCH OBJECTIVES ............................................................................. 23
  5.1 Paper 1 ............................................................................................................... 23
  5.2 Paper 2 ............................................................................................................... 24
  5.3 Paper 3 ............................................................................................................... 24

6 METHODS ..................................................................................................................... 27
  6.1 Participants .......................................................................................................... 27
  6.2 Clinical and cognitive inventories ..................................................................... 28
  6.3 Subject clustering and centrality analyses ...................................................... 28
  6.4 Genetic analyses ................................................................................................. 29
  6.5 MRI acquisition .................................................................................................. 30
  6.6 Resting-state fMRI ............................................................................................ 31
    6.6.1 Independent component analysis ................................................................. 31
    6.6.2 Static functional connectivity ........................................................................ 32
    6.6.3 Dynamic functional connectivity ................................................................. 33
    6.6.4 Global levels of FC ....................................................................................... 33
6.7 Morphometric MRI ....................................................................................................... 34
6.8 Diffusion based measures of white matter ................................................................... 34
6.9 Multimodal data fusion ................................................................................................. 35
6.10 Statistical analyses ......................................................................................................... 36
  6.10.1 Paper 1 .................................................................................................................. ............................. 36
  6.10.2 Paper II ................................................................................................................. ............................. 37
  6.10.3 Paper III................................................................................................................ ............................. 38
6.11 Ethical considerations .................................................................................................... 38

7 SUMMARY OF PAPERS ............................................................................................. 41
  7.1 Paper I: ........................................................................................................................... 41
  7.2 Paper II: .......................................................................................................................... 42
  7.3 Paper III: ........................................................................................................................ 43

8 GENERAL DISCUSSION ............................................................................................ 45
  8.1 Are current models of depression supported by neuroimaging? .........................45
  8.2 Data-driven depression subgroups ............................................................................47
  8.3 Alternative models to tackle depression heterogeneity ............................................49
    8.3.1 Transdiagnostic approach .......................................................................................... ................ 49
    8.3.2 Normative modelling....................................................................................................... .................. 50
  8.4 Advances in genetics .................................................................................................... 53

9 METHODOLOGICAL CONSIDERATIONS .................................................................. 55
  9.1 Participants, clinical and cognitive measures ............................................................ 55
  9.2 Dimensionality and brain parcellation......................................................................... 57
  9.3 Estimation of dFC .......................................................................................................... 58
  9.4 Polygenic risk scores ...................................................................................................... 59

10 CONCLUDING REMARKS ....................................................................................... 61

11 REFERENCES ............................................................................................................... 63

PAPERS I-III .................................................................................................................. 85
1 General Summary

Depression is a debilitating disorder with a high prevalence compared to most other mental disorders. It has become increasingly apparent that depression is clinically heterogenous, most notably in terms of the breadth and range of possible symptom configurations and profiles. Although there has been a plethora of neuroimaging studies over the years, findings have been inconsistent, leading to a lack of robust and specific imaging-based biomarkers for depression across neuroimaging modalities, which may be partly explained by its clinical heterogeneity. The complex nature of depression is also reflected in its genetic architecture, with more and more studies confirming its highly polygenic nature. To complicate things further, depression is comorbid with other mental disorders, in particular with anxiety, and showing strong links and phenomenological overlap with neuroticism. The main aim of this thesis is to elucidate depression heterogeneity through symptoms, advanced neuroimaging and genetics, which hopefully will further our understanding of the complex phenomenology and etiology of depression.

Our approach in paper one was to identify subgroups of depression based on symptoms of anxiety and depression using a data-driven clustering approach, and to test to which extent the symptom-based subtyping was supported by functional imaging. With the understanding that depression can be conceptualized along several continua in terms of the specific symptom and its severity, we did not distinguish between individuals with or without a history of clinically diagnosed depression in the cluster analysis. This approach yielded five subgroups of depression characterized by specific symptom profiles as assessed with centrality measures from network-based graph theory, with a presence of cases and controls across all subgroups. These subgroups were supported by differences in brain functional connectivity patterns from a subsample with resting-state functional magnetic resonance imaging (fMRI), which in particular implicated a fronto-temporal brain network. In contrast, we found no significant associations between brain functional connectivity patterns and dimensional or categorical measures of depression.

A goal in clinical neuroscience is not only providing a better understanding of the phenomena at hand, but also being able to say something about the individual in question, rather than just average group differences. To this end, we used a machine learning approach in paper II to map various conceptualizations of resting-state fMRI-based brain functional connectivity to cognitive and mental health traits related to depression in a large population-
based cohort (UK Biobank). We also predicted age and classified sex to use as a benchmark comparison for our predictions. Further, we wanted to assess the mapping between the genetic underpinning of these and related traits with brain functional connectivity using polygenic risk scores based on previously published genome-wise association studies (GWAS). Our results showed high prediction accuracy for age and sex, as well as robust prediction of fluid intelligence and years of educational attainment. In contrast, we observed low prediction accuracy for symptom loads of depression and anxiety and trait level neuroticism, as well as all of the polygenic risk scores. This study further demonstrates the heterogeneity of depression and related mental health traits, but also the need to refine polygenic risk scores.

One explanation for the inconsistent findings in the neuroimaging literature of depression is the relative absence of studies integrating different imaging modalities together. Multimodal fusion approaches take into account the covariation across modalities, which may lead to unseen patterns that have the potential for increased clinical sensitivity. To this end, we used linked independent component analysis (LICA) to combine data from morphometric, diffusion weighted and resting-state fMRI in individuals with or without a history of depression. Our findings revealed strong associations with age and sex with brain components related to global properties of cortical macrostructure, diffusion-based properties of white matter integrity, and default mode network amplitude. In contrast, we found no association of these brain components with case-control status (depression vs healthy controls), nor symptom loads for depression and anxiety across groups, or any interaction effects with age or sex. The machine learning analyses were generally in line with the univariate analyses, showing low model performance for classifying cases from controls and predicting symptom loads for depression and anxiety, but high model performance for predicting age.

The findings of the three papers and thesis confirms the clinical heterogeneity of depression, and that advanced neuroimaging by itself is not enough to elucidate the neurobiological underpinnings. One of the keys to solving the depression riddle are more precise methods of stratification at the individual level. This could ultimately pave a path towards the development of personalized treatment and prevention of depression.
2 List of Papers


3 List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD</td>
<td>Attention deficit/hyperactivity disorder</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the receiver operating curve</td>
</tr>
<tr>
<td>BAI</td>
<td>Beck Anxiety Inventory</td>
</tr>
<tr>
<td>BDI-II</td>
<td>Beck Depression Inventory</td>
</tr>
<tr>
<td>BOLD</td>
<td>Blood oxygen level dependent</td>
</tr>
<tr>
<td>CES-D</td>
<td>Center for Epidemiologic Studies Depression Scale</td>
</tr>
<tr>
<td>CNV</td>
<td>Copy number variant</td>
</tr>
<tr>
<td>dFC</td>
<td>Dynamic functional connectivity</td>
</tr>
<tr>
<td>DMN</td>
<td>Default mode network</td>
</tr>
<tr>
<td>DSM</td>
<td>Diagnostic and Statistical Manual for Mental Disorders</td>
</tr>
<tr>
<td>DTI</td>
<td>Diffusion Tensor Imaging</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography</td>
</tr>
<tr>
<td>EPQ-S</td>
<td>Short Form Eysenck personality Questionnaire</td>
</tr>
<tr>
<td>FA</td>
<td>Fractional anisotropy</td>
</tr>
<tr>
<td>FC</td>
<td>Functional connectivity</td>
</tr>
<tr>
<td>fMRI</td>
<td>Functional magnetic resonance imaging</td>
</tr>
<tr>
<td>GAD-7</td>
<td>Generalized Anxiety Disorder 7-item scale</td>
</tr>
<tr>
<td>GMD</td>
<td>Grey matter density maps</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association study</td>
</tr>
<tr>
<td>HRSD</td>
<td>Hamilton Rating Scale for Depression</td>
</tr>
<tr>
<td>ICA</td>
<td>Independent component analysis</td>
</tr>
<tr>
<td>IDS</td>
<td>Inventory for Depressive Symptomatology</td>
</tr>
<tr>
<td>QIDS</td>
<td>Quick Inventory for Depressive Symptomatology</td>
</tr>
<tr>
<td>LICA</td>
<td>Linked independent component analysis</td>
</tr>
<tr>
<td>MADRS</td>
<td>Montgomery-Åsberg Depression Rating Scale</td>
</tr>
<tr>
<td>MDD</td>
<td>Major depressive disorder</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>MD</td>
<td>Mean diffusivity</td>
</tr>
<tr>
<td>NESDA</td>
<td>Netherlands Study of Depression and Anxiety</td>
</tr>
<tr>
<td>OCD</td>
<td>Obsessive-compulsive disorder</td>
</tr>
<tr>
<td>OFC</td>
<td>Orbital frontal cortex</td>
</tr>
</tbody>
</table>
PCA – Principal component analysis
PHQ – Patient Health Questionnaire
RD – Radial diffusivity
RMSE – Root mean squared error
SDS – Zung Self-Rating Depression Scale
sFC – Static functional connectivity
4 Introduction

Depression is the most common among the mood disorders and among the world’s most prevailing mental health disorders affecting one in 25 people across the globe (World Health Organization, 2017). Although there has been a wealth of research over the years, much remains to be understood, in part due to depression heterogeneity. As such, it has become a matter of public health to solve the so-called depression riddle, with the overarching goal of this thesis being to elucidate depression heterogeneity. The first introductory section will discuss the main concepts of this thesis relating to depression heterogeneity: clinical aspects, neurobiology, genetics, and overlap with anxiety and neuroticism. This leads onto the main research objectives of this thesis, followed by a general discussion of the three papers.

4.1 Depression diagnosis and clinical characteristics

The specific criteria for a major depressive episode based on the Diagnostic and Statistical Manual for Mental Disorders (DSM) IV (American Psychiatric Association, 1994) are having depressed mood or loss of interest almost every day for at least 2 weeks. Further, 5 of 7 of the following symptoms have to be present: change in appetite, change in sleep, change in activity, fatigue or loss of energy, guilt/worthlessness, lowered concentration, and suicidal thoughts. However, what additional symptoms there are and even what exactly constitutes depression symptoms is still debatable. For instance, the Beck Depression Inventory (BDI-II) has additional symptoms of past failure, feelings of punishment, feelings of dislike, self-criticalness, crying, agitation, indecision, and a loss of sexual interest. In this particular case, there is more focus on somatic and affective rather than cognitive symptoms (see Figure 1). However, the type of symptoms examined is dependent upon the clinical inventory used. For instance, the Patient Health Questionnaire (PHQ) only focuses on the symptoms that are in the DSM. Fried (2017) found that 7 common depressive questionnaires, BDI-II, Hamilton Rating Scale for Depression (HRSD), Center for Epidemiologic Studies Depression Scale (CES-D), Inventory for Depressive Symptomatology (IDS), Quick Inventory of Depressive Symptomatology (QIDS), Montgomery-Åsberg Depression Rating Scale (MADRS) and Zung Self-Rating Depression Scale (SDS) showed a low mean overlap (Jaccard index = 0.36). This can be considered problematic when attempting to relate findings from studies that use different depression clinical inventories. Symptoms that are only captured by some
questionnaires include mood reactivity, grief, interpersonal sensitivity, leaden paralysis (IDS), lonely, talked less, people disliked me (CES-D), inner tension, inability to feel (MADRS), hypochondriasis and loss of insight (HRSD) (Fried, 2017). These instruments also differ on whether they explicitly disaggregate symptoms (i.e. increase or decrease), in particular for neurovegetative symptoms such as appetite, weight, sleep (insomnia/hypersomnia) and psychomotor disturbances (e.g. van Loo, de Jonge, Romeijn, Kessler, & Schoevers, 2012). They also vary on the detail of certain symptoms, with the HRSD (Hamilton, 1960) for instance separating questions for initial, middle, and delayed insomnia.

Figure 1. Presence or absence of 52 symptoms of depression across 7 depression clinical inventories. A colored or hollow circle indicates that the symptom is directly or indirectly assessed respectively, in a given questionnaire. Figure reprinted from “The 52 symptoms of major depression: Lack of content overlap among seven common depression scales” by E. I. Fried, 2017, Journal of Affective Disorders. 208. p. 193. Reprinted with permission.
4.1.1 Clinical depression heterogeneity

The relatively low overlap amongst depression inventories may reflect covert heterogeneity in depression: many symptom profiles can lead to an MDD diagnosis. This means that any two individuals with MDD may not have any single symptom in common. Somewhat surprisingly, Fried and Nesse (2015) found that of the 3073 participants in the “Sequenced Treatment Alternatives to Relieve Depression” (STAR*D) clinical trial (Rush et al., 2004), there were 1030 unique symptom profiles based on a subset of 12 items from the Quick Inventory of Depressive Symptoms (QIDS-16). Strikingly, the most frequent profiles explained less than 2% of the variation, in which one of these only involved insomnia and no other symptoms. These authors speculate this partly explains why the first treatment stage of the STAR*D clinical trial only lead to one in four patient reaching remission. In a follow-up study using network analysis, Fried, Epskamp, Nesse, Tuerlinckx and Borsboom (2016) found that energy loss was the most central item among 28 symptoms. Additionally, several non-DSM symptoms were more important than symptoms based on the DSM (see Figure 2), including sympathetic arousal and panic, which provides a shift in perspective of what is conventionally thought of as important symptoms of depression. Taking this into account, it may not be surprising that 227 different symptom profiles lead to a diagnosis of a major depressive episode (Zimmerman, Ellison, Young, Chelminski, & Dalrymple, 2015).

There is evidence that certain types of depression treatments alleviate specific depression related symptoms. Bekhuis and colleagues (2018) compared the efficacy of a 6-month Short Psychodynamic Supportive Psychotherapy with combined therapy (psychotherapy with pharmacotherapy) on specific depressive symptoms as measured by the Symptom Checklist-90. They found that combined therapy decreased symptoms of feeling entrapped, emotional lability, worry, hopelessness, obsessive thoughts, blue mood and low energy more than psychotherapy alone. Network analyses revealed that changes in feeling entrapped and emotional lability were associated with change in obsessive thoughts, blue mood, worry, low energy and hopelessness, thus revealing nuances not discovered by conventional analyses. A recent preliminary study (Mullarkey, Stein, Pearson, & Beevers, 2019) directly investigated whether and to what extent an 8-week internet intervention (Deprexis) impacted specific depression symptoms as measured by the self-report version of the QIDS using network analysis. They found that Deprexis improved sadness and indecision, which were associated with change in self-dislike, fatigue, anhedonia, suicidality, and slowness in the intervention group compared to the waitlist group. Early, middle, late and
hyper insomnia, appetite and weight gain were not affected by the intervention group compared to the waitlist group. In an randomized control trial investigating the efficacy of attentional bias modification, node strength centrality analyses indicated that an increase in interest (from baseline to post session) was the symptom that changed the most in the network of 17 HRSD symptoms (Kraft et al., 2019).

Figure 2. A centrality plot based on node strength of the 28 depression symptoms of the clinician rated IDS. Briefly, node strength is the degree to which a given symptom is directly connected to the other symptoms based on the sum of the edges (i.e. correlation of the symptoms). Image adapted with permission from “What are ‘good’ depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis” by E. I. Fried, S. Epskamp, R. M. Nesse, F. Tuerlinx, & D. Borsboom, 2016, Journal of Affective Disorders. 189. p. 317.
Not only is depression common across psychiatric disorders, but there is evidence that it exacerbates the risk of developing other disorders. In a Danish population-based cohort of over 5 million individuals followed-up from 1900 to 2015, Plana-Ripoll and colleagues (2019) found that there was an almost 80 times higher risk of receiving a diagnosis of neurotic disorders (which includes anxiety) after receiving a diagnosis of a mood disorder. Depressive symptoms are also highly prevalent in non-psychiatric disorders. For instance, a systematic review and meta-analysis found that the mean prevalence in multiple sclerosis was 30.5% for a depression diagnosis, and 35% for clinically significant depressive symptoms. Even higher rates are seen in neurodegenerative disorders, such as Parkinson’s disease, with rates of 40-50% (Reijnders, Ehrt, Weber, Aarsland, & Leentjens, 2008). To complicate things further, individuals from the general healthy population can also experience symptoms related to depression from time to time in varying degrees.

4.2 Depression and the brain

4.2.1 Morphometric MRI studies

There is a myriad of morphometric magnetic resonance imaging (MRI) studies of depression, in particular, related to the volumetry of certain subcortical structures. One of the most implicated subcortical structures in depression is the hippocampus, with roughly 800 studies registered in Pubmed (as of March 2019), commonly linked with cognitive decline (e.g. Videbech & Ravndal, 2004). See for instance the meta-analysis by Gerrlings and Gerritsen (2017) which implicates that hippocampal volume atrophy is more prominent in late-onset than early-onset depression.

The amygdala is another subcortical brain region generally associated with depression, both in terms of structural atrophy and aberrant function (see below), and is an integral part of the extrahypothalamic stress-response system (e.g. Ressler, 2010). A recent study found an association between recent life stress (6 months) and amygdala atrophy but not with temporally distant life stress, suggesting rapid and reversible change (Sublette et al., 2016). The cingulate is another region often linked with depression, because of its implied role in the modulation of emotional behaviour (e.g. Drevets, Savitz, & Trimble, 2008). An early meta-analysis of 10 studies found evidence of right, but in particular, left volume reduction in the subgenual cingulate (Hajek, Kozeny, Kopecek, Alda, & Höschl, 2008).
Similar to the cingulate, regions of the frontal lobe, including the orbitofrontal cortex (OFC), are implied to play a role in emotional regulation and thus are often associated with depression. A recent study found structural asymmetry of the dorsolateral prefrontal cortex in patients with depression compared to healthy controls, which was also related to symptom severity (Liu et al., 2016).

The ENIGMA consortium conducted a meta-analysis to investigate cortical abnormalities (Schmaal et al., 2017) and identified thinner cortical grey matter in the OFC, anterior and posterior cingulate, insula and temporal lobes in adults with MDD compared to healthy adults. Further, they found adolescents with MDD showed reduced surface area in medial OFC, superior frontal gyrus, primary and higher-order visual, somatosensory and motor areas compared to healthy adolescents (see Figure 3). In a preliminary large-scale meta-analysis by the ENIGMA consortium, Ho and colleagues (2019) used shape analysis to probe the subcortical morphometry in 1,781 patients with depression compared to 2,953 healthy controls. Using a region of interest approach, they found that MDD patients had smaller surface area in regions including the nucleus accumbens, subiculum of the hippocampus, and basolateral amygdala compared to healthy controls. Subsequent analyses revealed that smaller surface area of the subiculum of the hippocampus and the basolateral amygdala in the patient group was driven by patients with adolescent-onset MDD, who additionally showed thinner cortex in these two regions. Further, recurrent MDD patients had reduced cortical thickness and surface area in the cornu ammonis 1 of the hippocampus and the basolateral amygdala compared to first-episode MDD patients. Overall, the findings of this study suggest specificity of age at onset and recurrence of depression with respect to subcortical structures.

4.2.2 DTI studies

Diffusion tensor imaging (DTI) provides (indirect) measures of brain white matter coherence and integrity. There is some evidence that white matter integrity in the brain serves as a backbone for functional connectivity, for instance measured using resting-state function MRI (fMRI: see below), and could thus explain functional abnormality in a given disorder such as depression. The results from a meta-analysis (Liao et al., 2013) revealed decreased fractional anisotropy (FA) in the right frontal lobe, right fusiform gyrus, left frontal lobe and right occipital lobe. Additionally, fibre tracking show that the fascicles involved in the aforementioned regions included the right posterior thalamic radiation, right inferior
longitudinal fasciculus, right inferior fronto-occipital fasiculus and interhemispheric fibers running through the corpus callosum. This provides some support that depression may arise as a result of structural and functional disconnections amongst key brain regions, and specifically from this study, related to dysfunctional information transfer between hemispheres. However, caution is warranted when interpreting these results as the sample size was relatively small (n = 231 patients with MDD and n = 261 healthy controls), there was a diverse range of methods used, as well as the heterogeneity of the patient group. A larger and more recent meta-analysis (641 MDD patients, 581 healthy controls) including studies using tract-based spatial statistics (TBSS: Chen et al., 2016) converged on lower FA in the genu of the corpus callosum body and left anterior limb of the internal capsule. Further, they found that severity of MDD was associated with lower FA in the genu of the corpus callosum.

DTI may have longitudinal clinical utility for depression. A longitudinal study reported widespread FA decreases in both high-risk subjects that did not develop depression over a 2-year period and the high-risk subjects that did develop depression (Ganzola et al., 2018), suggesting that these may be trait related. In a study with an 8-week course of antidepressant medication (Korgaonkar, Williams, Song, Usherwood, & Grieve, 2014), a cross-validated regression model revealed 62% accuracy in predicting remission based on increased FA of the cingulum and decreased FA of the stria terminalis tracts, demonstrating that these regions are potential prognostic biomarkers for depression. However, caution is warranted when interpreting the findings from these studies as the sample size were relatively small.

4.2.3 Resting-state fMRI studies

In a neuronal framework, depression is now being understood as the result aberrant interactions in the brain (e.g. Hamilton, Chen, & Gotlib, 2013), similar to the “disconnection syndrome” hypothesis. This can be probed with resting-state fMRI, which essentially measures the spontaneous activity of the brain in the absence of a specific task or stimuli. A review of resting-state fMRI studies in depression reported some consistent findings across studies, which included stronger connectivity between the salience network and the anterior default mode network (DMN), and between the posterior DMN and the central executive network (Mulders, van Eijndhoven, Schene, Beckmann, & Tendolkar, 2015) in patients with depression compared to healthy peers. The DMN has been linked with negative self-
referential processes in depression (Perrin et al., 2012; Sheline et al., 2009), while the central executive network has been involved with goal-directed behavior (Menon & Uddin, 2010). Overall, these findings may be related to the cognitive and emotional dysfunctions seen in MDD (Marchetti, Koster, Sonuga-Barke, & De Raedt, 2012). A meta-analysis of seed-based studies found decreased functional connectivity (FC; i.e. the temporal correlation between brain regions or networks) within the frontoparietal network, and increased FC between frontoparietal regions and parietal regions of the dorsal attention network (Kaiser, Andrews-Hanna, Wager, & Pizzagalli, 2015). One of the more consistent findings is aberrant connectivity within the DMN (e.g. Kaiser, Andrews-Hanna, et al., 2015; Mulders, van Eijndhoven, Schene, Beckmann, & Tendolkar, 2015). This is corroborated by a recent large-scale study by the newly established REST-meta-MDD consortium consisting of 1,300 patients with depression and 1,128 healthy controls (Yan et al., 2019). Briefly, they found that decreased within-network FC in DMN was driven by recurrent depression, and not with first-episode drug-naïve MDD, but also not disease duration. This was also linked with depression symptom severity, but only in patients with recurrent MDD. A recent large-scale resting-state fMRI study (Xia et al., 2019) found some convergence across site specific analysis methods and whether global signal regression was implemented or not. They found that patients with MDD had lower activation (based on regional homogeneity) in the right postcentral gyrus, OFC, and middle and inferior occipital gyri compared to healthy controls. Conversely, patients with MDD had higher activation in part of the inferior frontal gyrus, right supramarginal gyrus, precuneus, and right superior frontal gyrus compared to healthy controls.

An early study found that the cognitive control network, DMN and affective networks showed increased FC to the “dorsal nexus” (a bilateral dorsal medial prefrontal cortex region) in patients with MDD compared to healthy controls (Sheline, Price, Yan, & Mintun, 2010). This was postulated by the authors to represent a mechanism for depression symptoms, although this has not been replicated in subsequent studies. In a seed-based analysis of the amygdala, Cullen and colleagues (2014) found lower FC between the amygdala in adolescents with MDD to the following regions: hippocampus, parahippocampus, and brainstem, which were inversely associated with general depression, dysphoria, and lassitude, and positively associated with well-being. Lastly, the adolescents with MDD showed positive FC between the amygdala and precuneus, whereas the controls exhibited negative FC. Using a seed-based analysis, Rzepa and Mccabe (2018) found that both anhedonia and depression severity in healthy adolescents were associated with decreased FC between the dorsal medial
prefrontal cortex and the precuneus. Additionally, they found that increased FC between the dorsal medial prefrontal cortex and anterior cingulate cortex/paracingulate gyrus was associated with anhedonia, while increased FC between the dorsal medial prefrontal cortex and the frontal pole was associated with depression severity.

A longitudinal study using seed-based analysis found that attenuated left-to-right subgenual cingulate cortex FC at baseline distinguished the resilient from recurring-episode group after 14 month follow-up (Workman et al., 2017). Another potential use of resting-state fMRI is in discriminating between (mood) disorders, rather than just a case-control approach. Yin and colleagues (2018) investigated the connectivity of insular subregions (ventral anterior insular cortex, dorsal anterior insular cortex, and posterior insula) at the whole-brain level in patients with MDD and bipolar disorder. Crucially, the patients with bipolar disorder did not have (hypo)manic episodes prior to and including the scanning session, and diagnosis was determined at follow-up. Briefly, they found similar decreases in insular connectivity in dorsal frontal regions, but the patients with bipolar disorder had additional specific decreased insular connectivity in somatosensory and motor.

### 4.2.4 Depression heterogeneity is reflected in the brain

Despite some overlap in the brain imaging literature of depression, there is also a considerable degree of mixed findings. In one of the largest meta-analyses of depression MRI studies at the time it was published (1728 MDD patients and 7199 controls), the ENIGMA consortium sought to identify subcortical brain volumes that characterize depression (Schmaal et al., 2016). The only significant finding was lower hippocampal volumes in MDD patients compared to controls. However, the effect size was very small (Cohen’s d = -0.14, % difference = -1.24), which is somewhat surprising in light of the sheer number of studies that implicated the hippocampus in depression. Equally as important, hippocampal volume differences are seen in other psychiatric disorders including schizophrenia (e.g. Moberget et al., 2019), where the effect sizes are larger, and even more so for neurodegenerative disorders such as Alzheimer’s disease (e.g. Henneman et al., 2009). There was also a trend towards smaller amygdala, but again the effect size was small (Cohen’s d = 0.11, % difference = -1.23), and only associated with age of onset rather than case-control difference. This null finding is intriguing, because as previously mentioned, the amygdala has classically been considered as one of the more crucial brain regions in depression pathophysiology. The regions that were significantly different between patients with MDD and healthy controls in
the most recent aforementioned ENIGMA study of cortical abnormalities (Schmaal et al., 2017) were also characterized by small effect sizes, ranging from Cohen’s $d = -0.10$ to -0.14 (see Figure 3). These small effect sizes echo the lack of robust and specific neural markers for depression. Further, a study comparing hippocampal subfield volumes in bipolar disorder, MDD and healthy controls did not report any robust volumetric differences between MDD and healthy controls (Cao et al., 2017).

Figure 3. Regions showing cortical thinning in adults with MDD compared to healthy controls. The color indicates the effect size for the regions, which are relatively small. Figure reprinted from “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.” by L. Schmaal et al., 2017, *Molecular Psychiatry*. 22(6). p. 902. Reprinted with permission.

A similar picture of depression heterogeneity is painted in the resting-state fMRI literature. There is considerable inconsistency in terms of whether connectivity of implicated networks such as the salience network, dorsal attention network, and central executive network area increased or decreased. Further, there are inconsistencies on the involvement of additional resting-state networks in depression, such as the so-called “affective network” which includes the amygdala and limbic regions (Sheline et al., 2010). Although there were some consistent findings in the previously mentioned large-scale multi-site study conducted by Xia and colleagues (2019), the between-group differences were partially affected by episode status and age of onset in the patients with MDD. Furthermore, there was low reproducibility of correlations between FC and clinical variables across sites. This last point is an issue inherent in general when linking psychological phenomena to neuroimaging.
illustrated by a recent study that found poor replicability of the association between grey matter volume and 36 psychological measures (Masouleh, Eickhoff, Hoffstaedter, Genon, & Alzheimer’s Disease Neuroimaging Initiative, 2019).

This lack of consistent findings in depression also extends to task-based neuroimaging studies. Müller and colleagues (Müller et al., 2017) conducted a meta-analysis of functional neuroimaging experiments of depression from the past 20 years, which included 34 cognitive and 65 emotional processing tasks, with a total of 1058 MDD patients. Strikingly, in their main analyses, they found no significant differences in MDD patients compared to controls in task-based brain activation for cognitive processing ($p > 0.29$) nor the emotional processing ($p > 0.47$). In secondary analyses splitting the tasks into subcategories, they found no convergence for positive (all $p > 0.15$), negative (all $p > 0.76$) nor memory (all $p > 0.48$) processes in MDD patients compared to controls. Crucially, supplementary analyses accounting for confounds by restricting the analyses to (1) patients not receiving medication, (2) patients without comorbidity and (3) patients with early-onset depression did not change the results. Overall, they conclude that this lack of convergence in aberrant task-based brain activation in patients with MDD for is partly related to differences in experimental paradigms and analyses, but also on the heterogeneity of the depression samples. Taken together, these large-scale studies illustrate that there are no clear neurobiological markers for depression, likely due to the clinical, or additionally, brain heterogeneity in depression.

### 4.3 Polygenetic architecture of depression

Depression is moderately heritable, with estimates from 28% to 44% (Fernandez-Pujals et al., 2015). Over the recent years, the focus has shifted from candidate gene markers of depression to acknowledging its polygenic nature (e.g. McIntosh, Sullivan, & Lewis, 2019), similar to the trend in neuroimaging, a factor shared with many other traits. A recently published genome-wide association study (GWAS) of depression (Wray et al., 2018) with 135,458 and 344,901 controls identified 44 independent risk variants, which has increased to 102 risk variants based on 246,363 cases and 561,190 controls (Howard et al., 2019). Additional analyses in the GWAS by Wray and colleagues (2018) revealed that the risk variants they identified were related to gene expression patterns in prefrontal and anterior cingulate cortex. An important gene variant identified was the splicing regulator RBFOX1, implicated in gene regulatory processes. Both of these GWAS found that the TCF4 was putatively associated with depression, in line with findings of decreased gene expression at the mRNA and protein...
level for patients with depression compared to healthy controls (Mossakowska-Wójcik, Orzechowska, Talarowska, Szemraj, & Galecki, 2018). This is interesting, because TCF4 is involved in the excitability of prefrontal neurons (Rannals et al., 2016), as pointed out by the authors. Relatedly, they also found that the variants identified revealed pathways associated with neurotransmission, and enrichment of the central nervous system.

Polygenic risk scores are one way to encapsulate all the information of a GWAS into one cumulative score. Briefly, polygenic risk scores are the individual level sum of the risk variants identified by a GWA weighted by their effect size. For instance, Levine and colleagues (2014) derived a polygenic score based on the top 11 SNPs of their GWA analysis of 7,000 individuals. This depression polygenic risk score was associated with depressive symptoms based on the CES-D, illustrating the potential utility of polygenic risk scores in predicting future depression severity. To take this a step further, Shen and colleagues (2019) examined the association between depression polygenic score with 210 behavioral and 278 neuroimaging phenotypes in >10,000 individuals in the UK Biobank. They found a positive association between depression polygenic score and several behavioral measures, including sleep problems, stressful life experience, and subjective ratings of physical health. Further, they found an increased effect of depression polygenic risk on those that were exposed to (reported) childhood trauma, stressful life events and those living in more socially deprived areas. Amongst the brain imaging traits, they found a causal effect of liability to depression on lower global white matter microstructure, association-fibre and thalamic-radiation microstructural integrity based on Mendelian Randomization analysis.

However, there are inconsistent findings in polygenic risk score studies (Peyrot et al., 2018) and other genetic studies of depression, which can possibly be partly be attributed to symptom heterogeneity (Kaufman, 2018). A preliminary study sought to investigate the clinical heterogeneity of depression by investigating the genetic underpinnings of depression at the symptom level based on the items from the PHQ-9 questionnaire for 148,752 subjects in the UK Biobank (Thorpe et al., 2019). Their GWA analysis revealed no overlap in genetic loci amongst the 9 symptoms, with genetic correlations (i.e. the proportion of variance shared due to genetic causes) from 0.54 to 0.96. Interestingly, fatigue had a higher SNP heritability than depression, although to what extent this was independent or related to depression is not known in this context. A confirmatory factor analysis based on the genetic correlations amongst these symptoms lead to a 3-factor model representing psychological, neurovegetative, and psychomotor/concentration symptoms. Overall, these genetic studies demonstrate that depression is highly polygenic, and associated with other psychological and
behavioural measures, but also clinically heterogeneous illustrated by genetic specificity at the symptom level.

4.4 Depression overlap with anxiety and neuroticism

4.4.1 Anxiety

Comorbidity between depression and anxiety has been shown in several studies over the years (e.g. Schapira, Roth, Kerr, & Gurney, 1972), with rates of almost 50% based on an epidemiological study and 75% in the NESDA study (Lamers et al., 2011). In fact, the HRSD has three questions probing psychiatric and somatic anxiety and agitation. Some symptoms of anxiety based on the Beck Anxiety Inventory (BAI) include fear of worst happening, terrified or afraid, nervous, fear of choking, fear of losing control, fear of dying and being scared. This comorbidity might explain why the inter-rater reliability for MDD and generalized anxiety disorder reported in DSM-5 field trials for 100 patients (Cohen’s kappa between 0.20-0.39) was in the lower range of the diagnoses assessed (Regier et al., 2013). Studies indicate that depressed patients with a history of anxiety disorder have worse treatment outcomes than those without, and significantly more psychopathology (Brown, Schulberg, Madonia, Shear, & Houck, 1996). Indeed, studies also indicate that those with comorbid anxiety disorders also have poorer overall functioning (e.g. Grunhaus, 1988), higher levels of suicidal ideation, greater pathological worry, and younger age at illness onset (e.g. Zimmerman & Chelminski, 2003). As such, early studies suggested that a key aspect of prognosis is to recognize and identify those with comorbid conditions (e.g. Gorman, 1996).

A handful of studies have attempted to address whether anxiety and depression share neurobiological mechanisms. In one of these studies, Zhao and colleagues (2017) compared brain morphometry in an anxious depression group (n = 32), non-anxious MDD group (n = 22) and controls (n = 43). The MDD group showed thinner cortex in temporal regions and smaller left hippocampus volume compared to healthy controls, while the anxious depression group had thinner cortex in the left superior frontal cortex amongst other regions and increased subcortical volume of the caudate nuclei. Volume of the caudate nuclei correlated with the anxiety/somatization factor score, indicating that this may distinguish anxious from non-anxious depression, although interpretation is limited by the small sample size. Couvy-Duchesne and colleagues (2018) sought to identify cortical endophenotypes for anxiety-depression based on scores from the Somatic and Psychological Health Report. They used
data from 157 monozygotic and 194 dizygotic twins to parcellate cortical thickness and surface area into genetically informed clusters. In an overlapping twin and sibling sample (n = 834), they found a nonlinear negative association between anxiety-depression and smaller surface area of the occipito-temporal region, mainly the right lingual and fusiform gyri. This finding was replicated in a sample from the Human Connectome Project (HCP: n = 890) and was also significantly heritable (h² = 0.55 for the primary sample and h² = 0.63 for the HCP sample).

Anxious depression is often considered a subgroup of depression (Rush, 2007), and existed in some form as a formal part of the DSM-III as noted by Zimmerman and Chelminski (2003). The current DSM (V) now has a formal subgroup of depression, anxious distress that encompasses this. There is some support for this as its own subgroup based on functional neuroimaging studies, however, they are few and sample sizes are small (n = 14 to 25) (Andreescu et al., 2009, 2011; Etkin & Schatzberg, 2011; Waugh, Hamilton, Chen, Joormann, & Gotlib, 2012). There are few structural neuroimaging studies investigating anxious depression, and one of these did not find any differences in grey matter volume compared with patients only with MDD and only with anxiety (van Tol et al., 2010).

4.4.2 Neuroticism

To complicate things further, both depression and anxiety have strong links with neuroticism (e.g. Uliaszek et al., 2010). For instance, one study found that neuroticism acted as a risk factor for depression, and predisposed to comorbidity with anxiety disorders (Xia et al., 2011). Another study found evidence that neuroticism was closely related to an internalizing factor based on structural equation modelling, underlying both anxiety and depression (Griffith et al., 2010). Interestingly, there is substantial phenomenological overlap between symptoms of anxiety and depression with neuroticism traits. This similarity can be observed in the neuroticism subscale of the Short Form Eysenck personality Questionnaire (EPQ-S). For instance, in terms of depression symptoms, the irritability trait is matched by BDI item 17, while the mood trait is matched by BDI item 1 (sadness). An example of the overlap with regards to anxiety is that trait worry is akin to GAD-7 items 2 and 3 (worry related), and the tense trait being similar to GAD-7 item 4 (inability to relax). This phenomenological overlap may not be so surprising considering that some definitions of neuroticism includes the tendency to experience negative affect more broadly and have difficulty regulating emotional experience (Ormel et al., 2013).
In a large GWAS of 12 neuroticism items based on the EPQ-S, Nagel, Watanabe, Stringer, Posthuma and Sluis (2018) found that these 12 traits had high genetic correlations with symptom load for depression, as expected. However, genetic correlations were lower for worry of embarrassment (0.42) and nervous feeling (0.45) but noticeably higher for loneliness (0.83) and miserableness (0.81), possibly suggesting a certain degree of specificity. Interestingly, genetic correlations of these 12 neuroticism items showed the same but weaker pattern with MDD, which points to genetic heterogeneity of MDD. Further, for anxiety disorders, genetic correlations were highest with suffering from nerves (0.78) and sum score neuroticism (0.76). They also identified a “depressed affect” and a “worry” cluster based on hierarchical clustering of the genetic correlations, the former of which the authors point out is identical to what has been found in previous factor analysis of the full EPQ-R neuroticism scale (Mor et al., 2008). This cluster consisted of the items encompassing loneliness, miserableness, mood, and fed-up and exhibited high genetic correlation with MDD (0.66), anxiety disorders (0.67) but especially symptom load for depression (0.86). A follow-up meta-analysis across many neuroticism GWAS (Nagel, Jansen, et al., 2018) demonstrated that 32 of the risk variants identified for neuroticism were also genome-wide significant for the depressed affect cluster, with 7 of these being shared with depression. The genetic correlations of depression and depressed affect with 35 other traits (including subjective well-being, anxiety disorders, educational attainment and more), which further strengthening the relationship between depression and this depressed affect cluster is that. Taken together, these studies show that depression, anxiety and neuroticism have both clinical, neurobiological and genetic overlap.
5 Main research objectives

The overarching aim of the present thesis was to contribute toward disentangling the heterogeneity of depression, and thus, contribute to an increased understanding of depression as whole. Depression heterogeneity encompasses many aspects, but based on the above reviewed literature, the focus will be on bridging the gap between symptoms, polygenic architecture, brain functional connectivity but also multimodal imaging fusion. The rationale is that depression has both complex neurobiological and polygenic aetiology, supported by overlap with other psychological traits and mental disorders, in particular anxiety but also neuroticism. The approach used is mainly data-driven in nature, which may aid in a less constrained characterization of depression.

5.1 Paper 1

The aim of paper 1 was to address symptom heterogeneity in depression and related subclinical phenotypes. As previously mentioned, many symptom profiles can lead to the same depression diagnosis (e.g. Fried & Nesse, 2015). Further, symptoms of anxiety tend to co-occur with symptoms of depression (e.g. Johansson, Carlbring, Heedman, Paxling, & Andersson, 2013; Femke Lamers et al., 2011). In addition, depression may exist on a continuum including the general healthy population, which experience transient depression related symptoms. In order to address this symptom heterogeneity, we used high dimensional data clustering to extract subgroups based on individual symptoms of depression (BDI-II) and anxiety (BAI) across individuals with or without a history of depression in a multisite sample. We characterized the symptom profiles of the resulting subgroups based on mean differences, but also eigenvector centrality from graph-theory to assess their network properties. A secondary aim was to assess whether these symptom-based subgroups are supported by functional neurobiology. We did this by assessing the functional brain network properties of these subgroups based on static and dynamic connectivity from resting-state fMRI in a subset of the total sample.
5.2 Paper 2

The aim of paper 2 was to assess the predictive ability of fMRI-based functional connectivity towards (i) cognitive and mental health traits including depression and (ii) genetic architecture of related traits based on polygenic scores. As previously mentioned, depression has moderate heritability based on twin and family studies, with estimates from 28% to 44% (Fernandez-Pujals et al., 2015), and large GWAS have revealed a highly polygenic architecture (Howard et al., 2019; Wray et al., 2018). In addition to clinical and genetic overlap with anxiety, depression is linked with neuroticism (e.g. Hansell et al., 2012). To establish neuroimaging-based biomarkers that are both robust and replicable, large samples are of the utmost importance. To date, little is known about the clinical utility of fMRI features to the polygenic risk of depression and related mental health characteristics. Even less is known about the sensitivity of specific frequency bands of the fMRI signal that are linked to a range of neural processes (Buzsáki & Draguhn, 2004) and even depression (e.g. Wang et al., 2016). The specific sensitivity of static and dynamic brain functional connectivity to mental health characteristics and their genetic architecture may aid towards the development of a brain-based nosology, prevention and treatment. To this end, we used resting-state fMRI data from the UK Biobank employing a multivariate machine learning approach to predict (i) cognitive traits and dimensional measures of depression, anxiety and neuroticism and (i) their polygenic architecture.

5.3 Paper 3

The main aim of paper 3 was to assess to what extent the combination of neurobiological domains based on multimodal fusion sensitive to categorical and dimensional aspects of depression. As is evident in the introduction to this thesis, a host of studies have identified brain regions and networks implicated in depression across structural and functional neuroimaging modalities. However, there are both mixed findings in the neuroimaging literature, but also little unity across imaging domains due to them being studied in isolation. Multimodal fusion takes into account the contribution of the selected neurobiological domains, and can potentially identify hidden patterns of co-variation, that may be more sensitive than the sum of their parts (Doan et al., 2017). In this paper, we used linked independent component analysis (LICA) to combine macrostructure (cortical surface area and thickness, and grey matter density), white matter diffusion properties, and DMN connectivity (resting-state fMRI). We tested for group differences, but also for interactions with age and
sex based on differences in prevalence and depression characteristics (Husain et al., 2005; Marcus et al., 2005), and associations with symptom load for depression and anxiety. A machine learning approach was used to assess the clinical utility of the resulting brain components, which was compared with age prediction.
6 Methods

6.1 Participants

The thesis uses a cross-sectional design across all three papers. The clinical (inventory) data in paper I comes from individuals with (n = 605) or without a history (n = 437) of depression, which were drawn from four research projects at the Clinical Neuroscience Research Group, Department of Psychology, University of Oslo (from 18 to 72 years of age). These were (1) *Secondary prevention of depression applying an Attention Bias Modification Procedure* (2) *The effects of serotonin polymorphisms and gender on cognitive control and brain function in major depression and healthy control* (3) *Cognitive control, genetics and emotion regulation* and (4) *Executive functions in binge drinking young adults*. Individuals with a history of depression were screened using the Mini International Neuropsychiatric Interview 6.0 (Sheehan et al., 1998) for three of the projects, while Structural Interview for DSM-IV (First, Spitzer, Gibbon, & Williams, 1997) was used for one of them. To take a transdiagnostic approach but also to increase the representativeness and the generalizability of our study, we did not exclude individuals that also had histories of additional Axis-I related disorders, in particular anxiety disorders, but also to a small extent episodes of (hypo)mania to a small extent. The individuals with a history of depression were mainly recruited from outpatient clinics in the Oslo area, while individuals with no history of depression were recruited using posters, newspaper advertisements and social media. The MRI subsample of individuals with (n = 178) or without (n = 72) a history of depression was drawn from project 1. Paper III consists of 71 healthy controls and 170 individuals with a history of depression with complete data across all included MRI measures (see below) from the MRI-subsample in paper I. The exclusion criteria for the MRI-subsample were a history of neurological disorders and MRI contraindications, but see the papers for full details.

The data in paper II comes from the UK Biobank (https://www.UK.Biobankio bank.ac.uk/), which is an unprecedent open access database consisting of a cohort of over 500,000 individuals from the United Kingdom that started data collection in 2006. The UK Biobank has a rich and vast array of measures genetic, physical and health data. According to Web of Science, there are already over 500 published studies that have at least mentioned the UK Biobank, demonstrating the impact it has had in the life sciences. For instance, the GWAS of depression phenotypes by Howard and colleagues (2018) has been cited 54 times while the GWAS of brain imaging phenotypes by Elliot and colleagues (2018)
has been cited 38 times, both published in 2018. As per March 2019, it is possible to get access to over 25,000 T1-based brain images, which is steadily growing, and close to 20,000 fMRI-based brain images. The sample in paper II was drawn from the October 2018 release of the UK Biobank imaging data (Sudlow et al., 2015), which consisted of 12,213 individuals with resting-state fMRI scans (UK Biobank field 20227) that passed the UK Biobank quality assessment (Miller et al., 2016). The exclusion criteria in this study were mental or neurological disorders based on ICD-10 (n = 210) and non-Caucasians (n = 1659), yielding a final sample of 10,343 individuals. The genetic data that was used to run GWA and compute polygenic risk scores consisted of 402,515 individuals (see below).

6.2 Clinical and cognitive inventories

The BDI-II (Beck, 1996) and BAI (Beck, 1996) were used to probe 21 (current) symptoms of depression and anxiety respectively in papers I and III. Additionally, we calculated the originally proposed somatic-affective (12 items for BDI-II and BAI each) and cognitive factor subscales (9 items for BDI-II and BAI each) to assess the specificity of associations with brain functional connectivity (see below). In paper II we used symptom loads from the PHQ-9 (Kroenke, Spitzer, & Williams, 2001) and GAD-7 (Spitzer, Kroenke, Williams, & Löwe, 2006) which mirror the DSM-IV criteria for depression and generalized anxiety disorder respectively. Both of these questionnaires in paper II were taken from the online follow-up assessment in the UK Biobank. We calculated sum neuroticism based on the neuroticism subscale from EPQ-S (Eysenck, Eysenck, & Barrett, 1985) collected from instance 2 (imaging assessment) using a previous implementation by Nagel and colleagues (Nagel, Watanabe, et al., 2018). The twelve item-level neuroticism scores were used in GWA (see below). In paper II, years of educational attainment was calculated based on the “qualifications” variable (UK Biobank field: 6138) following the procedure by Okbay and colleagues (2016) collected at instance 2 (imaging assessment). Fluid intelligence in paper II (UK Biobank field: 20016) was based on the sum of the number of correct answers on the verbal numerical reasoning test (13 items) taken at instance 2 (imaging assessment).

6.3 Subject clustering and centrality analyses

We used subject-level clustering in paper I to identify whether we could identify putative subgroups of depression based on symptoms of depression (BDI) and anxiety (BAI). The main rationale behind this is depression symptom heterogeneity which has been discussed in
the introduction. We did not distinguish between cases and controls in clustering, with the
notion that depression likely exists on a continuum. Clustering models can be roughly divided
into two broad categories: hard clustering, whereby the data (i.e. subjects) belong to one
specific cluster and soft or fuzzy clustering, in which the data (i.e. subjects) belong to each
cluster to a certain degree. Thus, the advantage of soft clustering is that one can estimate the
likelihood or uncertainty of belonging to a cluster, in addition to actual cluster assignment.
The clustering algorithm we used in paper I, high dimensional data clustering in the R
package *HDclassif* (Bergé, Bouveyron, & Girard, 2012), falls under soft clustering. Another
advantage of this specific clustering algorithm is that it can calculate the optimal number of
clusters based on the Bayesian Information Criterion. For specific details of this method, see
the original study and paper II.

We used graph-theoretical and network-based approaches to characterize the
symptom profiles of the subject-level clusters in paper I, based on the partial correlations of
all the symptoms. As illustrated in the introduction, network-based approaches may reveal
patterns amongst nodes (i.e. symptoms) that may not be apparent when using measures that
calculate them in isolation (e.g. mean or sum scores). The main measure we used was
eigenvector centrality on each of the symptoms calculated using the
*eigenvector_centrality_und.m* function in the Brain Connectivity Toolbox (Rubinov &
Sporns, 2010), which reflects how much influence a node (i.e. symptom) has on the entire
network. Thus, a node (i.e. symptom) with high eigenvector centrality in a given subgroup
has high influence in the network that subgroup in graph-theoretical terms. Additional more
specific measures of centrality: node strength, closeness, and betweenness were estimated
using the R package *qgraph* (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom,
2012). For full details of these methods, see the original studies and paper I

### 6.4 Genetic analyses

In paper II, we used GWAS results for broad depression, probable MDD (Howard et al.,
2018), diagnostic MDD (Wray et al., 2018), item-level and sum neuroticism (Nagel,
Watanabe, et al., 2018), anxiety (Otowa et al., 2016) and schizophrenia (Ripke et al., 2014) to
compute polygenic scores through PRSice v.1.25 (Euesden, Lewis, & O’Reilly, 2015). We
ran our own GWA holding out the MRI-imaging participants for years of educational
attainment and fluid intelligence to avoid bias and overfitting, as the original papers (Lee et
al., 2018; Savage et al., 2018) included UK Biobank data in their discovery dataset. Briefly,
GWA are typically performed on a discovery dataset while the resulting polygenic risk scores are estimated on an independent target dataset to avoid overfitting. For full details of the GWA, see the original papers, and paper II for our own GWA and polygenic score computation.

6.5 MRI acquisition

Neuroimaging data for paper I and paper III were collected at the Oslo University Hospital using a 3T Philips Ingenia MRI scanner (Phillips Healthcare) with a 32-channel head coil. Cushions and headphones were used to mitigate scanner noise and head motion. Anatomical T1-weighted 3D turbo field echo images used for co-registration purposes in paper I and analyses in paper III was collected for 74 controls and 194 patients (TR = 3000ms, TE = 3.61ms, FA = 8°, approx. 3 minutes, 1 mm isotropic voxels). Resting-state fMRI data used in paper I and paper III were collected using a T2* weighted single-shot gradient echo EPI sequence for 72 controls and 178 patients (TR = 2500ms, TE = 30ms, FA = 80°, 3.00 mm isotropic voxels, 45 slices, 200 volumes, approx. 8.5 minutes). We instructed participants to keep their eyes open and to remain awake. Diffusion weighted data used in paper III were collected for 72 controls and 184 patients with a dual spin echo, single-shot EPI sequence (TR = 7200ms, TE = 86.5ms, FOV = 224 x 224 mm², 2.0mm isotropic voxels). Thirty-two unique diffusion weighted volumes were collected at b-value = 1000s/mm² in addition to two gradient echo field maps (blip up/down volumes) to correct for magnet field inhomogeneity (TR = 10462 ms, TE1 = 54ms, TE2 = 54ms, flip angle = 90°, 112 x 112 matrix, 60 transverse slices, 2.0mm isotropic voxels).

The MRI-data in paper II was collected by the UK Biobank. Full details of the MRI acquisition and preprocessing (including T1-data which the UK Biobank used to co-register the resting-state fMRI data) can be seen in Miller and colleagues (Miller et al., 2016). Most of the MRI data used in paper II were collected in Cheadle Manchester, UK on a Siemens Skyra 3.0T scanner (Siemens Medical Solutions, Germany) with a 32-channel head coil. A small number of scans (n = 354) were collected at an identical scanner in Newcastle, UK. Briefly, resting-state fMRI data was collected using a T2* weighted multiband (x8 multislice acceleration) EPI sequence for 12,213 individuals as of the October 2018 release (TR = 0.735ms, TE = 39ms, FA = 52°, 2.4 mm isotropic voxels, 490 volumes, 6 minutes). Subjects were instructed to keep their eyes fixated on a crosshair, and “think of nothing in particular”
6.6 Resting-state fMRI

FMRI is based on the blood oxygenation level dependent (BOLD) signal. Here, the level of de-oxygenation in the blood is being measured with relation to specific cognitive tasks based on experimental paradigms. More specifically, it is the change in relative levels of oxyhemoglobin and deoxyhemoglobin caused by the release of oxygen, which decreases the MRI signal intensity (Roskies, 2008). The interest in resting-state fMRI specifically, has increased substantially within the last decade. One major driving force behind this is the finding that the brain’s energy expenditure only increases by 5% during a cognitive task as compared to “rest” (Raichle & Gusnard, 2002). Another reason is that it is less strenuous for participants, which is crucial in certain patient populations, but more importantly reflects the intrinsic properties of the brain.

6.6.1 Independent component analysis

Independent component analysis (ICA), a multivariate data-driven technique based on blind source separation, was used to extract RSNs in papers I and III, and by the UK Biobank in paper II. An advantage of ICA over atlas-based approaches is that it provides unbiased estimates of spatiotemporal RSNs across all included subjects (Cole, Smith, & Beckmann, 2010). It also has advantages over earlier seed-based approaches, in that it requires no a priori hypothesis, extracting RSNs based on covariance, and then creating spatial maps from the BOLD time series (e.g. Damoiseaux et al., 2006). For paper I and III, group-level ICA using FSL’s MELODIC was performed on a balanced subset of individuals with a history of depression and with no history of depression (n = 72 from each group) to avoid bias (Kaufmann et al., 2015). In paper II, the UK Biobank ran a group-ICA on 4100 datasets based on the top 1200 weighted spatial eigenvectors from MELODIC’s Incremental Group-PCA (Miller et al., 2016). For paper I and III, model order (i.e. the number of components) was fixed at 40, while in paper II we chose the provided model order 25. For paper I and III we assessed the spatial maps and frequency profiles based on previous studies (Kelly et al., 2010). For papers I and II, we regressed out the time series of identified noise components. In paper I, several additional components were discarded due to the spatial maps not matching with any established resting-state networks or being a mixture of signal and noise. This yielded 19 ICs in paper 1 (see Figure 4 for examples of common resting-state brain networks) and 21 ICs in paper 2 for the analyses. Finally, dual regression (Nickerson, Smith, Öngür, &
Beckmann, 2017) was run to produce subject specific spatial maps and corresponding time series which were used in network modelling.

**Figure 4.** Example of typical resting-state networks from the data in paper I (with a Z-score threshold of 3 to 15): DMN (IC05), right (red) and left (blue) frontoparietal networks (IC02 and IC03) and a visual network (IC01).

### 6.6.2 Static functional connectivity

Within the last decade, more studies are using a network-based approach for resting-state fMRI, i.e. investigating the functional topography and connectedness of the brain, which is frequently referred to as the “connectome” (e.g. He & Evans, 2010). Most resting-state fMRI studies and the ones mentioned so far have investigated static functional connectivity (sFC). Essentially this is the average functional connectivity across a whole resting-state scanning period. In papers I and III we used partial correlations between the time-series. In contrast to full correlation, partial correlations model the direct connectivity between nodes (e.g. brain networks or brain regions) after controlling for all other connections, which has been shown to be more sensitive to true connections (Smith et al., 2011). These partial correlations were
L1-regularized, estimating regularization strength (lambda) for each individual subject (Friedman, Hastie, & Tibshirani, 2008; Kaufmann et al., 2015; Ledoit & Wolf, 2003). Regularization penalizes model complexity by imposing sparsity, removing likely spurious edges/connections. In paper II, we computed sFC based on the full fMRI time-series and also within the 0.04 – 0.07 Hz using bandpass filtering.

6.6.3 Dynamic functional connectivity

It is quite possible that information is lost with methods estimating sFC, considering that they do no encapsulate how the brain communicates over time. In contrast, dynamic functional connectivity (dFC; e.g. Hutchison et al., 2013) measures variability in the strength or spatial organization of functional connectivity among brain regions. DFC may have improved diagnostic utility, with one study indicated that dFC is better at discriminating between patients with schizophrenia and bipolar disorder as compared to sFC (Rashid et al., 2016). Recent evidence suggests that dFC captures task-based phenotypes such as processing speed to a larger extent than sFC, while both explain self-reported measures such as life-satisfaction equally well (Liégeois et al., 2019). Despite this, sFC still provides useful information as regions exhibiting aberrant sFC may not be the same regions exhibiting aberrant dFC, which was the case for patients within the Alzheimer’s disease continuum (Córdova-Palomera et al., 2017). There are several methods to conceptualize dFC. In papers I and II we used a phase-based method, originally conceived by Glerean, Salmi, Lahnakoski, Jääskeläinen, Sam and colleagues (2012), and later adaptations (Córdova-Palomera et al., 2017; Demirtas et al., 2016). This specific method involves first bandpass filtering the time series (0.04 – 0.07 Hz in our case), and finally the Hilbert transform and estimation of the Kuramoto order. For specific details of this method, see the aforementioned studies and paper I.

6.6.4 Global levels of FC

The methods mentioned above to calculate sFC and dFC are at the edge-level, i.e. connections between nodes. Proxies for functional integration at the whole-brain can be estimated using for instance graph theoretical measures. To measure functional integration based on sFC, we computed global efficiency using the `efficiency_wei.m` function in the Brain Connectivity Toolbox (Rubinov & Sporns, 2010). Global efficiency is the average inverse shortest path length in a network and is primarily influenced by short paths, whereby the length of a path is the shortest distance between any two given nodes. Based on dFC, we
computed metastability, a measure of dynamic stability, and synchrony, a measure of general coherence (e.g. Váša et al., 2015). Metastability is a potential marker for cognitive and behavioural functioning (Hellyer, Scott, Shanahan, Sharp, & Leech, 2015; Kahana, 2006; Naik, Banerjee, Bapi, Deco, & Roy, 2017; Shanahan, 2010), whereas synchrony is hypothesized to be necessary for information exchange within the brain (Fries, 2005).

6.7 Morphometric MRI

We estimated cortical thickness and surface area in paper III using FreeSurfer (http://surfer.nmr.mgh.harvard.edu) (Fischl et al., 2002) to be used in LICA (see below). For more technical details, see paper III. Briefly, FreeSurfer can characterize anatomy such a cortical thickness and folding patterns, as well as structural boundaries at the subcortical level at each vertex (point of triangles). It also involves surface-based registration that accounts for individual anatomy (Fischl, Sereno, & Dale, 1999). Some of the other advantages of FreeSurfer is high specialization in cortical surface representation from the grey matter segmentation, surface-based group registration capabilities and accuracy of subcortical structure measurements. We used a total of 74 healthy controls and 194 patients with a history of depression in paper III. We also created grey matter density maps (GMD) based on the T1-weighted data using voxel-based morphometry as implemented in the computational anatomy toolbox (CA12: http://www.neuro.uni-jena.de/cat/) within SPM12 (https://www.fil.ion.ucl.ac.uk/spm/). The main advantage of this approach is that it can be applied across the whole-brain level and doesn’t need the a priori definition of specific regions (Henley et al., 2010).

6.8 Diffusion based measures of white matter

DTI has been used as a measure for brain microstructure or white matter integrity based on the diffusion of water molecules. The diffusion of water in (normal) brain tissue follows one direction (anisotropic), constrained mainly by features of cellular membranes such as myelination and packing of axons (e.g. O’Donnell & Westin, 2011). Macroscopic diffusion anisotropy therefore is influenced by the underlying tissue orientation, and tissue heterogeneity (Basser, Mattiello, & LeBihan, 1994). Diffusion tensor consist of three eigenvectors (i.e. diffusion direction) and three associated eigenvalues (i.e. the degree of diffusivity). We computed several DTI metrics: FA, mean diffusivity (MD) and radial diffusivity (RD) using FSL tools including DTIFIT. FA describes the degree of anisotropy in
each voxel, reflecting the integrity of white matter (e.g. Basser & Pierpaoli, 1996). MD is a measure of the total magnitude of diffusion in each voxel independent of direction and may be related to ischemic white matter damage (Bijanki et al., 2015). RD measures the transverse direction of diffusion and appears to be modulated by myelin in white matter (Song et al., 2002).

6.9 Multimodal data fusion

There has been a substantial increase in studies reporting findings from multimodal brain imaging data (e.g. Calhoun & Sui, 2016). However, most studies still analyze these neuroimaging modalities separately. Calhoun and Sui (2016) have delineated several levels of multimodal imaging work. One of these levels is data integration, whereby the data are analyzed separately and the overlaid on each other, which does not allow inference of interaction. The next level is one-sided or asymmetric data fusion, where one neuroimaging measure or modality is used to constrain another, such as DTI to constrain fMRI (e.g. Bowman, Zhang, Derado, & Chen, 2012). The disadvantage is that assumptions are imposed on the data that may not transfer from one modality to the next. The last level, which is implemented in paper III, is symmetric data fusion, takes into account the cross-information sin all neuroimaging modalities simultaneously, which may reveal hidden inter-related patterns. Of particular interest is the potential interaction or degree of fusion between structural measures and functional measures (e.g. resting-state fMRI). As pointed by Calhoun (2018), an obvious perspective is that structure impacts function, although the reverse may also be true in terms of brain plasticity (e.g. Draganski et al., 2004). Notwithstanding, the potential interaction of structural and functional brain characteristics may help disentangle depression heterogeneity. Specifically in paper III, we used FMRIB’s LICA (FLICA, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLICA) to fuse data based on cortical thickness, surface area, GMD, DTI (FA, MD and RD), and DMN functional connectivity (resting-state fMRI). Briefly, this creates spatial maps based on the degree of covariance of imaging measures across subjects (see Figure 5 for some examples), and the extent to which a subject contributes to each LICA component (subject weights) (Groves, Beckmann, Smith, & Woolrich, 2011; Groves et al., 2012). Model order was set to 40 for the main analyses and 80 in supplemental analyses (for comparison). For more details of LICA, see paper III.
Figure 5. Example of components from LICA in paper III characterized by regionally specific features. Warm colours indicate positive weighting (e.g. higher FA) whereas blue indicates negative weighting (e.g. lower cortical thickness). For IC11, we see multimodal involvement of all included DTI-based white matter properties, as well as GMD and cortical surface area, explaining 1.6% of the total variance. For IC30, we see covariation of GMD and cortical thickness, explaining 0.9% of the total variance.

6.10 Statistical analyses

6.10.1 Paper 1

For details of the statistical analyses, see paper 1. Briefly, we ran an ANCOVA to test for associations with global efficiency, metastability and synchrony with subject-level clusters, BDI-II and BAI sum and subscale scores. We used the same model to assess for FC-edge differences but using network-based statistics (NBS) for inference (Zalesky, Fornito, &
Bullmore, 2010) which using a family-wise error rate control. This was specifically designed to probe alterations in connectome alterations based on the clustering structure of the edges (i.e. taking into account the interconnectedness of the edges). The first step in this procedure involves setting a threshold at the edge-level, which is user-determined. Essentially, there is a balance between large differences (conservative thresholds) and sensitivity (liberal thresholds). Permutations based testing (10,000 permutations) was used to estimate a null distribution to determine statistical significance.

**6.10.2 Paper II**

We employed a machine learning approach for all of the analyses. A simple definition of machine learning is any approach in which one predicts a target variable (e.g. depression symptom scores) based on a model built on a training set. We used such an approach to predict 2 cognitive and 3 mental health related traits, and 20 polygenic scores based on related traits, using various combinations of difference FC conceptualizations. The specific algorithm we used was shrinkage linear estimation (Schäfer & Strimmer, 2005) from the R-package ‘care’ (http://strimmerlab.org/software/care). One of the advantages of this method is that it can estimate the (relative) feature importance of all the explanatory variables based on the correlation-adjusted marginal correlation scores (Zuber & Strimmer, 2011). There are many measures of model performance (prediction accuracy), which are all based on the relationship between the actual and predicted scores of the target variable. To avoid any potential limitations of any single measure, we computed the most commonly used ones: root mean squared error (RMSE), mean absolute error (MAE), correlation (either Spearman’s rho or Pearson’s r depending on the scale of the target variable) and $R^2$.

To reduce bias and overfitting, we used two approaches. The first was to split our dataset into a training set (80%) and a left-out test set for final model validation (20%). The second was to perform 10-fold cross-validation with 100 repetitions on the training set (80%). Here, the training set is (randomly) partitioned into 10 equal “sets”, in which 9 of them are used to predict the leftover set. This is repeated until each fold has been predicted by the combination of the remaining 9 sets. The mean result across the ten folds represents model performance (i.e. prediction accuracy) Then, the procedure is repeated 100 times, whereby the sampling of individuals in the 10 folds is random in each iteration. This yields 100 estimates of model performance, whereby the mean of this is used as the point estimate for inference. We used permutations-based testing (10,000 permutations) to determine the
statistical significance of each model, bonferroni correcting for the number of models we tested.

6.10.3 Paper III

Univariate analyses using linear models similar to paper I and machine learning analyses similar to paper II were used in paper III. In addition to testing for main effects of depression case-control status, and symptom loads for depression and anxiety on the subject weights of each LICA component, we also tested for their interaction effects with age and sex. To classify patients from controls in the machine learning approach based on the subject weights of the LICA components, we used shrinkage discriminant analysis (Ahdesmäki & Strimmer, 2010) in the R-packages ‘sda’ (http://www.strimmerlab.org/software/sda/). We computed the typical measures of model performance for classification analyses: accuracy (i.e. percentage of correct predictions), sensitivity (i.e. the proportion of correctly identified cases) and specificity (i.e. the proportion of correctly identified controls). The limitation with accuracy it is biased by unequal sample sizes in groups. As such, we computed area under the receiver operating curve (AUC) as our main measure of model performance using the R-package ‘pROC’ (Robin et al., 2011) which unlike accuracy, takes both sensitivity and specificity into account. In this case, the relative feature importance of each LICA component was based on the correlation-adjusted t-scores (Ahdesmäki & Strimmer, 2010). We predicted symptom loads for depression and anxiety using shrinkage linear estimation (which was implemented in paper II). As in paper II, statistical significance in the machine learning approach was determined using permutations-based testing (1000 iterations), based on AUC for classification and RMSE for prediction of symptoms.

6.11 Ethical considerations

All studies were carried out in accordance with the Helsinki Declaration of 1957 as revised in 2008, and papers I and III were approved by the Regional Ethical Committee of South East Norway (REK Sør-Ost), and the UK Biobank Access Committee Project (No. 27412) for paper II. All participants in papers I and III provided written informed consent prior to involvement, and by the UK Biobank for paper II. Information about the studies was provided to the participants during recruitment, which was followed up with more details including written protocols through e-mail or the post, in particular regarding MRI scanning. Each participant was informed that the data they provided were confidential and would be
anonymized in line with the Norway Personal Data Act (https://lovdata.no/dokument/NL/lov/2018-06-15-38). All participants were informed several times of their right to withdraw at any point from the studies, and additionally have their data collected on them deleted. These considerations were especially crucial as all participation in papers I and III were voluntary with no form of monetary compensation. Particular care was taken towards the patient group in papers I and III, as the clinical screening had the potential of bringing back hurtful memories. Participants were only invited to the MRI-session if they did not suffer from claustrophobia or MRI-contraindications (such as pregnancy or metallic implants), which were screened for again prior to the MRI-scanning. For papers I and II, any suspicion of incidental findings was screened initially by the radiographers followed by the radiologist for final confirmation (of which none required referral).
7 Summary of papers

7.1 Paper I:

Data-driven clustering reveals a link between symptoms and functional brain connectivity in depression

Background: Depression varies in symptom constellations at the individual level, compounded by other psychiatric domains, in particular anxiety. Depression may be better characterized by a combination of a dimensional and symptom-based approach, which may lead to identifying specific and robust biomarkers. To this end, we assess the brain functional connectivity patterns from resting-state fMRI in symptom-based clusters of individuals.

Methods: Symptoms of depression and anxiety based on the BDI-II and BAI respectively in individuals with or without a history of depression (n = 1084) were used in high dimensional data clustering to form subgroups. To ascertain whether these were supported by neurobiology, we compared static and dynamic functional connectivity patterns in these subgroups in a subset with resting-state fMRI data (n = 251).

Results: The clustering yielded five subgroups, characterized by distinct symptom profiles based on total severity but especially graph theoretical measures of centrality. This included higher eigenvector centrality of sadness, feelings of dislike, and loss of interest had higher in subgroups 1 compared to the other subgroups, while loss of pleasure, lack of energy, and tiredness had higher eigenvector centrality in subgroup 3 compared to the other subgroups. Noteworthy was the presence of patients and controls across all subgroups. The fMRI analyses revealed a main effect of brain sFC patterns, in particular involving a fronto-temporal network. In contrast, we found no significant brain local or global sFC or dFC patterns with sum scores of symptoms, case-control status or the symptom-based subgroups.

Conclusion: These subgroups were characterized by specific profiles of depression and anxiety symptoms, supported by both a dimensional model and neurobiological differences in resting-state based measures of static brain functional connectivity patterns.
7.2 Paper II:

Predicting cognitive and mental health traits and their polygenic architecture using large-scale brain connectomics

**Background:** The extent to which the neuronal basis and aetiology of cognitive abilities and mental disorders are shared is unknown. Further, the heterogeneity and phenomenological overlap of these traits is reflected by their genetic overlap and polygenic nature. The sensitivity and specificity of brain functional connectivity to dissecting the biological heterogeneity of these traits at the individual level has yet to be established.

**Methods:** The mappings between the brain connectome based on static and dynamic FC and (i) cognitive and mental health traits and (ii) their genetic underpinnings were established using machine learning in 10,343 healthy individuals from the UK Biobank. The cognitive traits included years of educational attainment and fluid intelligence, while the mental health traits included dimensional measures of depression, anxiety and neuroticism. We predicted polygenic scores for educational attainment, fluid intelligence, depression, anxiety, and different neuroticism traits, in addition to schizophrenia. We predicted age and classified sex to serve as a comparison for our main results.

**Results:** We observed high prediction accuracy for age and sex, and robust prediction accuracy for years of educational attainment and fluid intelligence. Age showed a distributed pattern of increased sFC, whereas years of educational attainment and fluid intelligence were characterized by decreased sFC in frontal and default mode networks. In comparison, we observed low prediction accuracy for the mental health traits and all the polygenic scores.

**Conclusion:** Cognitive abilities are reflected in brain connectomics, in contrast to symptom loads of depression, anxiety, trait level neuroticism, and polygenic scores of these traits. These findings may serve as a benchmark for brain connectomic mapping of mental health traits and their genetic architectures.
7.3 Paper III:

Multimodal fusion of structural and functional brain imaging in depression using linked independent component analysis

**Background:** Depression is characterized by aberrant structure and function in distributed brain regions and networks. However, the clinical heterogeneity but also neurobiological complexity of depression has impeded the endeavour towards robust and specific biomarkers. Multimodal fusion of structural and functional brain imaging and a dimensional approach may help parse the neuronal correlates of depression.

**Methods:** We fused cortical macrostructure (thickness, area, gray matter density) diffusion-based properties of white matter integrity (FA, MD, RD) and resting-state fMRI DMN amplitude using LICA in patients with a history of depression (n = 170) and controls (n = 71). We assessed association between the resulting brain components with case-control status, and symptom loads for depression and anxiety, as well as age and sex using univariate and machine learning analyses.

**Results:** The univariate analyses revealed no significant associations with case-control status, nor symptom loads for depression and anxiety with the resulting brain component, nor any interaction effects with age and sex, as well as low prediction accuracy in the machine learning analyses. In contrast, there were strong age and sex associations with mainly global but also regional specific brain components, with varying degrees of multimodal involvement, and high prediction accuracy for age in the machine learning analyses.

**Conclusion:** This study supports accumulating evidence that the magnitude of effect in brain imaging features between patients with a history of depression and healthy controls are small. More precise methods of stratifying depression at the individual level based on large sample sizes are needed to dissect the clinical and neurobiological heterogeneity of depression.
8 General Discussion

8.1 Are current models of depression supported by neuroimaging?

A crucial finding in the current thesis was the lack of statistically significant associations between depression case-control status and brain measures. More specifically, in paper 1 there was no robust association between resting-state brain FC at the edge-level or global level with depression case-control status. This is noteworthy, because we used NBS, which is a sensitive method of probing distinct aspects of network topology due to its improved statistical power compared to traditional mass-univariate testing (Zalesky et al., 2010). This was also the case in terms of the brain components in paper III, which consisted of both global and regionally specific features, with varying degrees of multimodal involvement of cortical macrostructure, white matter diffusion properties and DMN amplitude. This is also a notable finding, because several studies have provided some evidence that multimodal fusion using LICA is more clinically sensitive compared to analyses at the unimodal level (e.g. Alnæs et al., 2018; Doan et al., 2017; Wu et al., 2019). This is in accordance with an increasing number of studies suggesting nuanced differences in structural brain properties in patients with MDD compared to (Schmaal et al., 2017, 2016; Varoquaux, 2018; Wolfers, Buitelaar, Beckmann, Franke, & Marquand, 2015).

Although not statistically significant, the machine learning analyses in paper III revealed that patients with depression can be weakly classified from controls using the brain components from LICA (AUC = 0.57). This demonstrates both the utility of multimodal fusion approaches but also the added information from using machine learning approaches, which takes into account all features used, compared to conventional univariate analyses. With the low model performance in mind, the most important feature for classifying patients with depression was a component characterized by both high and low GMD in cerebellar regions (IC19). This adds to a growing body of research revealing the relevance of the cerebellum in psychiatric disorders such as schizophrenia (Moberget et al., 2018) and even depression (Depping, Schmitgen, Kubera, & Wolf, 2018).

Taking this further, we found no significant interaction effects between case-control status and age or sex in paper III, although to the best of my knowledge, no other study has provided evidence of age-by-group interactions with brain structure or function. There have been some reports of sex-by-group interactions, implicating for instance hippocampal
volumes (e.g. Frodl et al., 2002), and higher GMD in the cerebellum of male patients (Yang et al., 2017), but these were based on small sample sizes.

Another common finding across all three papers is that we found no robust association between resting-state fMRI FC patterns and symptom loads for depression and anxiety, consistent with the large-scale multi-site study of depression and resting-state DMN FC (Yan et al., 2019). This very much fits into one of the main messages by Fried and Nesse (2015): that sum-scores of symptoms may conceal specific associations, arguing for an individual symptom-based approach. With the low model performance of predicting sum scores of depression symptoms in paper III in mind, the most important feature was IC0, which was characterized by low GMD and cortical thickness in temporal and some frontal regions, consistent with previous meta-analyses of cortical morphometry in depression (e.g. Schmaal et al., 2017; Suh et al., 2019). IC0 was also characterized by diffusion-based properties in interhemispheric connections and frontal-striatal thalamic pathways, similar to previous studies although in the opposite direction (Chen et al., 2016).

Keeping in mind the low model performance for predicting sum scores of anxiety symptoms in paper III, the most important feature was IC6, which encompassed positive GMD in the thalamus, corroborating a recent meta-analysis (Wang, Cheng, Luo, Qiu, & Wang, 2018). This component also included negative weighting of GMD in hippocampal and amygdalar regions, in accordance with some studies (Hölzel et al., 2010; Machado-de-Sousa et al., 2014), and low FA in the splenium of the corpus callosum, and high MD and RD in the thalamus, consistent with some studies (Lu, Yang, Chu, & Wu, 2018). To reiterate, inference about the importance of these brain components must be interpreted with caution, as although they are in line with previous studies, the overall model performance was low, but does show the ability of machine learning approaches to quantifying clinical utility.

We found no significant associations between brain FC patterns at the edge and global level with both BDI-II and BAI cognitive or somatic factor scores in paper I. In contrast, we found an association between edge-level resting-state sFC and the symptom-based subgroups identified in paper I. Although we did not find any significant associations between symptom loads for depression (or anxiety) with the various brain measures, this does not necessarily preclude that depression exists on a continuum. Indeed, the most common finding of studies using data-driven clustering approaches to identify subgroups of depression is total severity differences (van Loo, de Jonge, Romeijn, Kessler, & Schoevers, 2012), which was also a characteristic of the subgroups identified in paper I. An additional support for the continuum
or dimensional aspect of depression is that the subgroups in paper I consisted of individuals with and without a history of depression, which was linked with total severity differences.

8.2 Data-driven depression subgroups

In paper I, both subgroups 4 and 5 where characterized by a lack of symptoms, in particular subgroup 4, in which only changes in sleep was present. Sadness and loss in interest have higher eigenvector centrality in subgroup 1, whereas loss of pleasure had higher eigenvector centrality in subgroup 3, which are the main symptoms of depression as defined by the DSM. This symptom specificity in subgroups may suggest that these subgroups are linked to specific mechanisms and environmental triggers. For instance, Lamers and colleagues (2010) identified several depression subgroups in NESDA, whereby the melancholic class was characterized by more smokers and childhood trauma than the atypical class. In contrast, the atypical class was characterized by higher body mass, presence of more women, and metabolic syndrome. Subgroup 3 in paper I was characterized by higher centrality in lack of energy and tiredness compared to the other subgroups. This may have long-term clinical relevance, as a group with remitted depression over a two-year period (van Borkulo et al., 2015) had lower centrality in fatigue or loss of energy at baseline compared to the group with persistent depression. In another longitudinal depression study from NESDA (Rhbergen et al., 2012), those with dysthymia and double depression ended up in the subgroup with better prognosis based on latent class growth analysis, which goes against DSM-IV categorization. Characteristics of the class with the best prognosis included younger age, more females, less childhood adversity, less somatic illness, lower neuroticism and higher extraversion.

An alternative approach to identifying subgroups of depression using a symptom-based approach is to use a (neuro)biological based approach. The study that has received the most attention using such an approach, and initially showed promise, was conducted by Drysdale and colleagues (2017). Briefly, they identified four subgroups of depression by combining resting-state fMRI FC and depression symptoms using canonical correlation analysis followed by clustering, based on a large multisite sample (n = 1,188). These subgroups were characterized by differences in dysfunctional connectivity in limbic and frontostriatal networks, similar to our finding of the fronto-temporal node (IC16) revealing different sFC across the subgroups in paper I. More specifically, a similar finding in paper I and Drysdale and colleagues (2017) was connectivity involving the thalamus in biotypes 1 and 2 and subgroup 3 respectively. Further in paper I, we found that the probability of
belonging to subgroup 1 and 3 were associated with unique brain FC patterns, with only a 5-edge overlap. Interestingly, the subgroups Drysdale and colleagues (2017) identified, had very distinct symptom profiles, and did not differ in overall symptom severity (with the exception of “biotype” 2) in contrast to most studies attempting to identify subgroups of depression using data driven methods (van Loo et al., 2012), including paper I. Briefly, biotypes 1 and 4 were characterized by increased anxiety similar to subgroup 2 in paper I, biotypes 3 was partly characterized by interest similar to subgroup 1, and biotypes 1 and 2 were characterized by increased anergia and fatigue similar to subgroup 3. Remarkably, these biotypes also had differential response to an experimental transcranial magnetic therapy in a subset of the total sample (n = 154).

To the best of my knowledge, two studies have attempted to identify subgroups of depression based on directed FC from fMRI, using Group Iterative Multiple Model Estimation, based on resting-state fMRI (Price, Gates, Kraynak, Thase, & Siegle, 2017) and task-based positive mood induction fMRI (Price, Lane, et al., 2017). A common finding in both of these studies is identifying two subgroups of depression, in which the one of the subgroups consisted of more patients. Beyond that, they revealed different directed pathways, suggesting that resting-state and task-specific fMRI capture unique information. In the task-fMRI study (Price, Lane, et al., 2017), the subgroup with more patients was characterized by hyperconnectivity, especially in ventral affective regions, while in the resting-state fMRI study (Price, Gates, et al., 2017) found different directed FC profiles of the DMN similar to our findings in paper I. Further, one of the subgroups in the resting-state fMRI study (Price, Gates, et al., 2017) had higher rates of anxiety, similar to our findings in paper I highlighting the significance of anxiety in parsing depression heterogeneity.

Although these studies using data-driven methods are promising and directly address depression heterogeneity, they need to be replicated in subsequent studies. In particular, a major caveat to the study by Drysdale and colleagues (2017) is a recent and thorough study which failed to replicate the findings (Dinga et al., 2019), which were amongst other things, related to preprocessing of the data prior to clustering. Further, Marquand, Wolfers and colleagues (2016) question the current use of clustering and related approaches, which need to be validated using more thorough statistics, as implemented by Dinga and colleagues (2019). This calls for alternative approaches in parsing depression heterogeneity which are discussed below.
8.3 Alternative models to tackle depression heterogeneity

8.3.1 Transdiagnostic approach

One of the elements of depression that this thesis has made evident is the overlap with both anxiety and neuroticism. Crucially, this extends to other psychiatric domains, as illustrated in the recent depression GWAS (Wray et al., 2018) where the genetic correlations indicate that depression was associated with bipolar disorder, schizophrenia, autism, and intellectual disability. This again brings into question the nosology of psychiatric disorders, including depression, which may be elucidated by a transdiagnostic approach. Such an approach is supported by the finding that the subgroups identified in paper I did not differ in history of anxiety disorders or (hypo)mania, cutting across traditional diagnostic boundaries.

A review by Buckholtz and Meyer-Lindenberg (2012) proposed that transdiagnostic patterns of dysconnectivity in the brain (e.g. based on fMRI FC) underlie the phenomenological and clinical overlap of psychiatric disorders. For instance, there is meta-analytic evidence (Sha et al., 2018) that the default mode, frontoparietal and sensorimotor networks are dysfunctional across 11 psychiatric disorder, including depression, based on 182 whole-brain resting-state fMRI studies. They also found that network connector nodes primarily involved in communications with cognitive components were affected, highlighting the potential relevance of the brain FC patterns of cognitive traits identified in paper II for depression and other psychiatric disorders. There were some disorder specific abnormalities, including lower activity in the visual system in MDD. These findings are supported to an extent by a similar large-scale meta-analysis, but this time based on seed-based resting-state fMRI studies and voxel-based morphometry of structural MRI (Sha, Wager, Mechelli, & He, 2019). They again identified that the default mode and frontoparietal networks were dysfunctional across the 8 included disorders, but also the salience network. They also found grey matter reductions across disorders localized in regions that are involved in general cognitive performance, corroborating the previous meta-analysis (Sha et al., 2018), further highlighting the relevance of the cognitive traits from paper II in psychiatric disorders.

Xia and colleagues (2018) used sparse correlation analysis to identify resting-state fMRI FC-based dimensions of psychopathology based on 111 item-level psychiatric symptoms, which included the psychosis-spectrum, attention deficit/hyperactivity disorder (ADHD), and depression in a sample of 663 youths. From this they extracted four dimensions of psychopathology: mood, psychosis, fear and externalizing behavior, which were each...
linked with distinct patterns of connectivity. The fear dimension consisted of social phobia and agoraphobia symptoms. The mood dimension consisted of items related to depressive symptoms such as *feeling sad* and *suicidality*, but also mania (e.g. “irritability”), and obsessive-compulsive disorder (OCD: e.g. “recurrent thoughts of self-harm”). Notably, these two specific symptoms of mania and OCD in the mood dimension fit well into a depression and anxiety perspective brought up in this thesis. Specific for the mood dimension was marked increase in connectivity between the ventral attention and salience networks, whereas the fear dimension was characterized by increased connectivity within the fronto-parietal network, with the symptom-based subgroups in paper I showing altered sFC between similar brain networks. Common across these four dimensions was segregation of the DMN and executive networks, in accordance with the aforementioned meta-analyses (Sha et al., 2019, 2018), differential sFC among the subgroups in paper I and feature importance for predicting the cognitive traits in paper II.

A recent study (Grisanzio et al., 2018) combined data-driven hierarchical clustering and a transdiagnostic approach. Specifically, they clustered on measures of self-reported negative mood, anxiety and stress symptoms from patients with MDD (n = 100), panic disorder (n = 53), posttraumatic stress disorder (n = 47), and healthy controls (n = 220). From this they identified 6 subtypes characterized by tension (n = 81; 91%), anxious arousal (n = 55; 13%), general anxiety (n = 38; 9%), anhedonia (n = 29; 7%), melancholia (n = 37; 9%) and normative mood (n = 180; 43%). All of these subgroups consisted of a mix of the patient groups, cutting across traditional diagnostic boundaries, and had differential profiles of behavioural neurocognitive measures, and electroencephalography (EEG) brain activation for a resting condition and an emotional task-based condition. More specifically, these subgroups differed in EEG activation across the frontal region, especially in the resting condition, in which the anhedonia subgroup had the highest EEG activation. For the emotion paradigm, there were differences in parietooccipital activation across subgroups, most notably in the general anxiety subgroup.

### 8.3.2 Normative modelling

A promising approach that may lead to more insight into depression and other mental disorders is normative modelling (Marquand et al., 2019; Marquand, Rezek, Buitelaar, & Beckmann, 2016; Marquand, Wolfers, et al., 2016), a type of dimensional approach. In contrast to case-control designs, normative modeling takes into account that cases, but also
controls, may be heterogeneous and exist on a continuum (see Figure 6), in line with our approach of combining cases and controls prior to clustering in paper I. Currently, most normative modelling approaches have been done with reference to (normal) ageing, mapping individual differences which helps to understand healthy variation (Marquand et al., 2019), tackling both brain and clinical heterogeneity (Marquand, Rezek, et al., 2016; Marquand, Wolfers, et al., 2016). For instance, this approach has been used to establish normative models of regional cortical surface area, thickness (Potvin, Dieumegarde, Duchesne, & Alzheimer’s Disease Neuroimaging Initiative, 2017) which revealed expected patterns of deviations in patients with Alzheimer’s disease. Further studies have used normative modelling to parse brain heterogeneity in ADHD (Wolfers et al., 2019), autism (Zabihi et al., 2018) and schizophrenia (Wolfers et al., 2018).

Figure 6. (A) The cases-control approach assumes homogeneity within each group although this may not be the case. Alternative models could be that (B) there exist distinct subgroups of cases (C) a continuum model which extends from controls to cases or (D) diffusivity and heterogeneity of the cases beyond clearly defined subgroups. Figure reprinted from “Understanding heterogeneity in clinical cohorts using normative models: Beyond case-control studies.” by A. F. Marquand, I. Rezek, J. Buitelaar & C. F. Beckmann, 2016, *Biological Psychiatry.* 80(7). p. 553. Reprinted with permission.

A common finding in these studies using normative modelling was the large intraindividual differences in the respective patient groups, and only a few identical brain abnormalities present in more than two percent of patients (e.g. Wolfers et al., 2018), demonstrating brain heterogeneity. As alluded to earlier, whether or not the inconsistent
findings in the neuroimaging literature of depression are solely due to clinical heterogeneity, or additionally brain heterogeneity, is not known, and can be assessed with normative modelling. I would speculate that normative modelling will reveal substantial brain heterogeneity in depression similar to the other mental disorders, given our findings in papers I to III and the recent literature. A major strength of normative modelling is its potential use to stratify cohorts, moving away from average group statistics which are based on consistent pattern of neurobiological atypicality across individuals. It also tackles brain heterogeneity by accounting for multiple pathological pathways even though these individuals may have the same symptoms (Marquand et al., 2019). A vital consideration is its bottom-up nature, which should be combined with top-down approaches, and validated with external measures such as symptoms (Marquand et al., 2019).

Normative modelling can be constrained to specific brain regions of interest, demonstrated in a study whereby hippocampal volumes were characterized in 19,793 healthy individuals in the UK Biobank, which has potential clinical utility in particular for Alzheimer’s disease (Nobis et al., 2019). Normative modelling can also be applied with other neuroimaging modalities, although currently, few studies have done so. For instance, Barber and colleagues (2019) characterized developmental patterns of whole-brain connectivity of striatal subdivisions from resting-state fMRI in a large developmental sample (8-22 years of age). They combined this with a transdiagnostic approach by assessing striatal age-related connection associations with symptom scales related to ADHD, psychosis, depression, and general psychopathology. Depression was specifically associated with aberrant connectivity between striatal subdivision and limbic regions. For general psychopathology, and common for all the symptom subscales was connectivity between striatal subdivisons and the dorsal posterior insula, which is involved in pain processing (e.g. Segerdahl, Mezue, Okell, Farrar, & Tracey, 2015). Although normative models usually model healthy ageing, which has been the case for all of the studies mentioned above, they can also be used to chart variation in cognition or any other trait which can then be linked to neurobiology (Marquand et al., 2019; Marquand, Rezek, et al., 2016). An exciting application is using the deviations from normative modelling in clustering (Marquand et al., 2019; Marquand, Rezek, et al., 2016; Marquand, Wolfers, et al., 2016), which may lead to more robust and precise stratification of depression. Overall, transdiagnostic approaches are aiding towards the reconceptualization of mental disorders, with normative modelling showing particular potential.
8.4 Advances in genetics

In line with the transdiagnostic approach, more and more studies are revealing genetic overlap between depression and other disorders. To tackle this more directly, the Cross-Disorder Group of the Psychiatric Genetics Consortium and colleagues (2019) conducted a meta-analysis GWAS of anorexia nervosa, ADHD, autism spectrum disorder, bipolar disorder, MDD, OCD, schizophrenia and Tourette’s syndrome based on 232,946 cases and 494,162 controls. Briefly, they found that based on genetic correlations, MDD was most closely related to autism spectrum disorder (0.45) and ADHD (0.44) which the authors note are two childhood-onset disorders. An exploratory factor analysis of these genetic correlations yielded a factor based on mood and psychotic disorders, and a factor based on autism spectrum disorder, ADHD, Tourette’s syndrome and MDD. The authors note that the presence of MDD in both factors may reflect the heterogeneity of MDD, echoing the message of this thesis. A GWAS of mood instability in the UK Biobank (Ward et al., 2019) identified high genetic correlations with MDD (0.74) and anxiety (0.64), and lower genetic correlations with other psychiatric disorders. In paper II we observed that correlations among polygenic scores of various traits varied for each of the three depression phenotypes (probable, broad and diagnostic). For instance, out of the three depression phenotypes, the polygenic score for sum neuroticism was most strongly correlated with broad depression ($r = 0.38$). Correlations among polygenic risk scores for the depression phenotypes varied for each neuroticism facet, with the strongest correlation being between broad depression and feeling miserable ($r = 0.30$). However, the broad and probable depression polygenic scores overlapped with the GWAS discovery dataset, which may have inflated these correlation estimates, although prediction accuracy was low and do not alter inference with respect to the brain connectomics results. Further supporting evidence of depression heterogeneity are different genetic correlations between specific depression symptoms and a range of other complex traits, including psychiatric, substance use, and socioeconomic phenotypes (Thorp et al., 2019). For instance, psychomotor changes and low self-esteem were correlated with anorexia nervosa, while suicidal ideation was strongly correlated with anxiety disorders.

It can be argued that a more promising avenue to parsing the genetic architecture of depression is not by investigating common risk variants (with small effect sizes) that are related to diverse processes, but rather the rare risk variants that have more pronounced, debilitating effects. Such rare risk variants likely account for the “missing” heritability that is observed in complex traits (Wainschtein et al., 2019). One way of investigating the effect of
rare risk variants are genomic microduplications or microdeletions, often known as copy number variants (CNVs) which have wide-reaching phenotypic impact. CNVs have for instance been associated with increased risk of developing autism spectrum disorders, intellectual disability (Kirov, Rees, & Walters, 2015), and schizophrenia (Rees et al., 2014). A recent UK Biobank study (n = 407,074 individuals of European ancestry) found that a preselected group of 53 CNVs linked to neurodevelopmental disorders were associated with self-reported depression (Kendall et al., 2019). Of these 53 CNVs, 3 of them were associated with risk of depression (1q21.1 duplication, Prader-Willi syndrome duplication, and 16p11.2 duplication). As the authors point out, it is noteworthy that these CNVs do not overlap with risk variants from a recent depression GWAS (Wray et al., 2018), supporting that standard GWA methods may not detect all the relevant risk variants regardless of statistical power.

It is possible that one of the genetic mechanisms underlying neuropsychiatric traits is the modulation of gene expression rather than just the presence of specific risk variants. Transcriptome data (i.e. the genetic code that is transcribed into RNA molecules) provides an avenue for investigating gene expression modulation. Gandal and colleagues (2018) collected transcriptome data from 700 cortical samples spanning patients with MDD to schizophrenia. In comparison to the other psychiatric disorders, MDD did not reveal synaptic or astroglial pathology. However, MDD showed a specific dysregulation of the hypothalamic-pituitary axis and hormone activity pathways, consistent with some models of MDD (Gold, 2015). Gamazon, Zwinderman, Cox, Denys and Derks (2019) collected transcriptome data from brain and non-brain tissue in 393 individuals profiling genes identified in GWAS of several psychiatric disorders. They identified 31 genes in broad depression, whereby 17 of these overlapped with schizophrenia. They did not find any association with MDD, but they surmise that this is mainly due to power relative to the other GWAS. Interestingly, non-brain tissues colon and whole blood showed the highest estimated true positive rate with broad depression. Although the focus of this thesis has been on disentangling depression heterogeneity mainly through neurobiology, this finding implies that gastrointestinal and immune systems are also crucial in depression, at least in terms of genetic predisposition.
9 Methodological considerations

There are many methodological considerations in all three of the papers and delving into all of them is out of the scope of this thesis. For this reason, I have chosen not to discuss each MRI modality in detail, although they have their own important methodological considerations. Amongst other things, I will not discuss construct validity: whether the BOLD response in fMRI can truly be considered a proxy for neuronal activity, or to what extent the diffusion weighted metrics used in paper III reflect white matter microstructural properties or integrity. Apart from methodological considerations regarding participants and inventory measures, the focus will rather be on methods more specific to the papers rather than neuroimaging as a field, in part because these have been discussed in detail previously.

9.1 Participants, clinical and cognitive measures

A common limitation across all three papers is that the symptom loads for depression and anxiety in papers I to III, and neuroticism in paper III were significantly positively skewed. In general, this may undermine any associations between these psychological traits and the various brain measures investigated. This is very much linked with the type of participants that were part of all three papers. Although we had a fairly large number of patients with depression in papers I and III, there were few severely depressed patients. An immediate reason for this is that patients with severe depression are less likely to volunteer in research projects, in particular when projects are fairly intensive, such as requiring separate data collection sessions and long MRI scanning sessions. Although the sample size in the UK Biobank is unprecedented, it is not fully representative of the UK population (Fry et al., 2017) in part because the participation is voluntary. Relevant to paper II, there were only generally healthy individuals in the analyses, which may not extend to (severely or clinically) depressed populations. Other factors that may influence the generalizability of the findings are the specific age (40 to 70), and socioeconomic status (generally middle-class and highly educated). The geographical specificity of the UK Biobank may actually bias genetic analyses. Haworth and colleagues (2019) demonstrated that single genetic variants and genetic scores from the UK Biobank based on multiple variants are linked with birth location, and could not simply be adjusted for in subsequent analyses.

As brought up in the introduction, the choice and use of specific clinical inventories is not trivial. Although the BDI (used in papers I and III) and PHQ (used in paper II) cover the
DSM symptoms, there are many other non-DSM symptoms that are not. A general advantage of the clinical inventories used in papers I to III is that they measure gradual assessment for each item, rather than a dichotomous (yes/no). Interestingly, one study found a significantly higher sum score (6 points) in a sample of healthy adults when the BDI-II was administered as a self-report (as originally intended to) compared to clinician-rated interview administration (Stepankova Georgi, Horakova Vlckova, Lukavsky, Kopecek, & Bares, 2018). They speculate that this difference is due to the participants comparing themselves with (1) their prime years, (2) what they believe other people normally feel or (3) what is normal for people of their age. This is noteworthy, because in terms of point 1, the instructions on the BDI-II state that the questions should be based on the last two weeks. However, with regards to points 2 and 3, there is no explicit specification given in the BDI-II. Another plausible factor is difficulty understanding the questions, which can be clarified upon if it is administered by a (trained) clinician.

Although the BDI aims to probe symptoms of depression within the last two weeks of administration, symptoms of depression and anxiety fluctuate regularly, possibly on a daily basis. It is interesting that there is a discrepancy in the time frame that symptoms are probed for BDI and BAI, whereby the latter is for the last week. A further issue is that the BDI and BAI were generally administered within 1 to 2 weeks of the MRI session rather than the day of the MRI session, further increasing the temporal lag. This temporal lag is one of the major limitations of the data from the clinical inventories in the UK Biobank used in paper. In this case, the PHQ and GAD were only available from the online follow-up, which is several months after the MRI session, and varies between individuals. This likely has an impact on the accuracy of association with the brain measures, in particular with resting-state fMRI measures which itself fluctuates at the individual level (Gordon et al., 2017; Shine, Koyejo, & Poldrack, 2016).

The measure used for years of educational attainment in paper II was actually a proxy based on a conversion from educational qualification. This means instead of having a continuous distribution, we had 6 discrete that were converted into years. Irrespective of it being a proxy, it was skewed, with an over-representation of highly educated individuals. Despite this, Okbay and colleagues (2016) found that 72 of the 74 SNPs in their discovery GWA dataset (which was based on actual years of educational attainment) had the same sign, and the genetic correlation was at 0.95.

The measure for fluid intelligence used in paper II has several issues which are discussed in detail by Kievit, Fuhrmann, Borgeest, Simpson-Kent and Henson (2018). Firstly,
it is only based on few items, lower than more representative measures of fluid intelligence, and only lasting for a couple of minutes. The specificity combined with the low number of items puts into question to what extent it captures all facets of fluid intelligence. An example pointed out by Kievit and colleagues (2018), item two, “which number is the largest?” is better characterized as crystallized rather than fluid intelligence. Secondly, although each question is multiple choice based, the result is dichotomous: either a correct or incorrect response, meaning that it is vulnerable to both floor and ceiling effects. Internal consistency (Cronbach’s alpha) in another study using the same measure in the UK Biobank was 0.62 (Davies et al., 2016). This level of internal consistency is lower than more representative measures of fluid intelligence, for instance, 0.83 for the fluid intelligence component of the Reynolds Intellectual Assessment Scale (Hagmann-von Arx, Gygi, Weidmann, & Grob, 2016). Despite these limitations, the SNP-based heritability of 31% in an earlier GWAS (Davies et al., 2016) was very consistent with estimates from other studies using similar general cognitive (Davies et al., 2015) and intelligence based measures (Davies et al., 2011; Marioni et al., 2014) as pointed out by Davies and colleagues (2016).

9.2 Dimensionality and brain parcellation

Selecting the optimal model order (i.e. number of components to extract) is non-trivial in ICA and LICA (and similar decomposition methods) for resting-state fMRI and other modalities, with no gold standard. The trade-off is between anatomical sensitivity in lower model orders and anatomical specificity in higher model orders (Elseoud et al., 2011), although very high model orders may result in decreased stability of the IC estimates (Li, Adali, & Calhoun, 2007). Another way of determining the optimal number of components is discriminability (prediction accuracy) in a machine learning framework. The current best estimates based on classifying specific psychiatric disorders from controls (using resting-state fMRI-FC) and predicting psychological traits is between 50 (Sripada et al., 2019) to 150 (Dadi et al., 2019; Sripada et al., 2019). These model orders are noticeably higher than what was chosen in papers I and II. One of our other main concerns was the interpretability of the derived components, and we suspect that the sample size in paper I was not sufficient to estimate robust components at higher model orders. The optimal model order is perhaps more complicated in LICA decomposition, as it combines several modalities. This is why we did the analyses at two model orders: 40 and 80, whereby we found similar feature importance rankings in both model orders. However, prediction accuracy was slightly lower at the higher
model order, but we speculate that this is due to the presence of more noise components, supported by the number of components dominated by a single subject which we discarded (13 in total). Indeed, 10 of the 13 components had a subject dominance threshold of over 60%.

The choice of the many different types of parcellation schemes from resting-state fMRI and other brain-imaging modalities has an impact on the results. There are several gross categories of parcellation schemes used in resting-state fMRI that are not mutually exclusive: pre-defined vs. data-driven atlases and functional vs. structural atlases. The Automated Anatomical Labelling (Tzourio-Mazoyer et al., 2002) with is a commonly used pre-defined structural atlas, whereas the Power atlas (Power et al., 2011) is a commonly used pre-defined functional atlas. Data-driven atlases include k-means clustering and agglomerative hierarchical clustering (Michel et al., 2012). Recent evidence suggests that functionally derived and data-driven methods are the best choices in terms of machine learning prediction accuracy, of which ICA fulfils both of these criteria (Dadi et al., 2019; Sala-Llonch, Smith, Woolrich, & Duff, 2019). Although not specifically assessed in these two studies, an advantage of ICA is the possibility to regress out the time series of identified noise components from the remaining components, and as noted by Sala-Llonch and colleagues (2019), ICA reduces redundancy across nodes. As pointed out by Dadi and colleagues (2019), ICA is able to capture uncertainty in regions through soft assignment in contrast to “hard” clustering methods. A disadvantage of structural based atlases is that there may be a degree of overlap in functional properties across nodes, which in particular impacts sFC using partial correlations (Sala-Llonch et al., 2019).

9.3 Estimation of dFC

There are a multitude of methods for conceptualizing dFC. The most common method is correlation-based sliding window analysis, which in this context, is essentially measuring the variation of FC across different time frames of specified length (i.e. windows). However, a major caveat with sliding-window analysis is the choice of window length. On the one hand, small temporal windows may introduce spurious fluctuations (e.g. Hutchison et al., 2013), but on the other hand, long windows hinder detection of temporal variations (e.g. Glerean et al., 2012). Despite this, Pedersen, Omidvarnia, Walz, Zalesky and Jackson (2017) demonstrates that these two methods detect comparable temporal properties when the window lengths were relatively short (< 60 s). A general limitation of most dFC methods, but
in particular related to sliding-window methods, is reduced ability to detect nonstationarities in resting-state fMRI scans of less than 10 minutes (Hindriks et al., 2016), which was the case for papers I to III. However, in the case of paper II, a multiband sequence was used for the data acquisition, such that the TR is substantially lower than traditional sequences such as in paper I, providing many more timepoints, which should in theory allow for robust measurement of temporal variation. Nonetheless, phase-based synchrony may still not be the optimal method of measuring dFC. It is thus possible that we would have found evidence of altered patterns of dFC associated with depression in paper I, as indicated by previous studies (e.g. Demirtas et al., 2016; Kaiser et al., 2016), if we would have used a different method.

One major caveat of this phase-based method is its susceptibility to frequency-based noise, although this was mitigated by narrowband filtering. One promising alternative are autoregressive models as used by Liégeois and colleagues (2019), which can exploit the temporal ordering of the fMRI time-series (Liégeois, Laumann, Snyder, Zhou, & Yeo, 2017).

As previously mentioned, the phase-based dFC method used in papers I and II is sensitive to frequency-based noise, and thus requires narrowband filtering. We specifically chose the range 0.04-0.07 Hz as it avoids both low-frequency drift, cardiac and respiratory variations and high frequency noise, and thus contains more reliable and functionally relevant information in gray matter (Glerean et al., 2012) which has been used in other studies using the same phase-based dFC (e.g. Córdova-Palomera et al., 2017). Related to our choice of frequency band, Thompson and Fransson (2015) found that brain sFC patterns and graph-theoretical properties of the brain are frequency dependent, providing evidence that time-scales effect both network integration and segregation. This is in line with our findings in paper II where we found that connectivity profiles of fluid intelligence, years of educational attainment and age differed when sFC was based on the full signal or bandpass filtered within 0.04-0.07 Hz.

9.4 Polygenic risk scores

The lack of any robust association between the 20 polygenic scores spanning psychiatric domains including depression, cognitive traits, and neuroticism facets suggests that polygenic scores may not encapsulate the genetics of these traits to a sufficient extent. Indeed, there is a growing body of research showing that polygenic risk scores across psychiatric domains explain very little trait variance, for instance 2% in case-control status of MDD (McIntosh et
This arguably means that at the moment, polygenic risk scores have low clinical utility, even for highly heritable traits (Zhao & Zou, 2019).

A further complication specific to estimating the full extent of the polygenic architecture is its low discoverability (i.e. effective strength of association) and highly polygenic nature, even relative to other psychiatric disorders. Holland and colleagues (2019) have estimated that one needs at least 100 million patients with MDD for this purpose, much higher than current sample sizes of depression GWAS (e.g. Howard et al., 2019). This is in stark contrast to for instance education, which is estimated to require over a million individuals, in good agreement with the sample size in the most current GWAS by Lee and colleagues (2018).

Another issue is the choice of significance threshold in the estimation of polygenic risk scores. To compensate for this, most studies conduct their analyses over a range of polygenic risk scores, although this can in some cases be impractical for various reasons (e.g. computational load, multiple comparisons issues, etc.). This was addressed in our paper by estimating polygenic risk scores at a conventional threshold ($p \leq 0.05$), a lenient threshold ($p \leq 0.5$), and a novel method where we take the first principal component from a principal component analysis across thresholds. A preliminary study found that the accuracy of a polygenic risk score at a given threshold is determined by the sparsity of the true genetic signal. Further, they determined that polygenic scores estimated at a specific threshold can only outperform polygenic scores based on all SNPs studied when the sample size is substantially larger than the true number of causal SNPs.

There are several methods that could potentially improve polygenic risk scores. One promising method is PRS-CS, which uses a high-dimensional Bayesian regression framework and places a continuous shrinkage prior on SNP effect sizes, and has been shown to outperform traditional estimation methods in the UK Biobank (Ge, Chen, Ni, Feng, & Smoller, 2019). An alternative approach is to use methods that more accurately quantify polygenic overlap across traits before estimation of polygenic scores (Frei et al., 2019).

Beyond the limitations of polygenic risk scores are inherent limitations of GWAS. Although this is out of the scope of this thesis, one such limitation is that GWAS are unable to detect ultra-rare variants associated with a given trait (e.g. Tam et al., 2019).
10 Concluding Remarks

The present thesis contributes to the understanding of depression and elucidating its heterogeneous nature. In paper I we identified five subgroups based on a data-driven clustering of depression and anxiety symptoms. Although they only seem to differ with respect to total symptom severity, closer inspection reveals that the pattern of symptom centrality differed for each subgroup. Further, in an fMRI subsample, we found that these subgroups showed a main effect of brain sFC patterns, and unique patterns associated with the probability of belonging to specific subgroups. We did not find any significant association between brain FC patterns and depression case-control status, nor sum scores of depression or anxiety symptoms, illustrating the heterogeneity of depression. Taking this a step further, using a population cohort from the UK Biobank, we were not able to predict symptom loads for depression, anxiety, or trait level neuroticism using a machine learning approach based on several conceptualizations of brain FC patterns. In contrast we were able to predict age and sex with high accuracy, and also robustly predict years of educational attainment and fluid intelligence, which are relevant for understanding depression. Crucially, we were not able to predict polygenic scores of related traits, which fits in with an increasing number of studies revealing low explained trait variance. Even advanced methods of fusing sensitive structural and functional brain properties across modalities in paper III, there was no significant association with conventional categorical and dimensional measures of depression, and is thus not enough to dissect depression heterogeneity.

The complex nature of depression and other mental disorders means that they cannot be explained by unicausal or oligo-causal theories informed by neurobiology (Paulus & Thompson, 2019) or genetics. In terms of fMRI, more studies investigating the sensitivity and specificity of static and dynamic measures of brain connectivity are needed required, and also their dependencies at specific frequency domains. In a neuroimaging context but potentially extending to other domains, this is precisely why multimodal fusion is a promising method, which as previously stated, takes into account the interactions between brain heterogeneity amongst individuals (Schnack, 2019). Although prediction accuracy for classifying patients from controls in paper III was low, prediction accuracy for age was very high, especially considering the low number of input features in the machine learning context. As a result, multimodal fusion may have potential for uncovering mechanisms across neuroimaging domains for the gap between chronological and biological age (i.e. brain age gap), a
A recurrent implied theme throughout this thesis is the need for large sample sizes. This is one of the main messages in the study by Masouleh and colleagues (2019) which we discussed in terms of replicability, and also Paulus and Thompson (2019). As illustrated by large-scale studies such as the ones conducted by ENIGMA, the effect sizes in neuroimaging studies of depression are overall small (Ho et al., 2019; Schmaal et al., 2017, 2016), but also for mental disorders in general (Paulus & Thompson, 2019). Small sample sizes also impact the robustness and inflate estimates of prediction accuracy in machine learning approaches (Wolfers et al., 2015), which will hamper their clinical utility when applied to other datasets (e.g. Rutledge, Chekroud, & Huys, 2019). One approach is to use openly available large population-based cohorts such as the UK Biobank, with vast arrays of neurobiological and genetic data. For richer and more clinically relevant measures of depression, large-scale meta-analyses and multisite studies such as the ones conducted by the ENIGMA and REST-meta-MDD (Yan et al., 2019) consortiums will be of the utmost importance. A major caveat in case-control designs is the implicit assumption that the patients, but also the controls are both homogenous and representative of their respective populations (Feczko et al., 2019). As a result, simply increasing sample size will not solve the issues of parsing depression heterogeneity if they are only conducted in a case-control fashion, and advanced neuroimaging is not sufficient by itself. I believe one of the keys to this is more precise methods of stratifying individuals with depression are needed, whether it be data-driven clustering similar to what we did in paper I, a transdiagnostic approach, or promising methods such as normative modelling, brain age prediction or others that have not be discussed in this thesis. Such approaches would arguably lead to better avenues for individualized treatment and possibly prevention of depression.
11 References


and the effect on health-related quality of life. PeerJ, 1, e98. https://doi.org/10.7717/peerj.98


Papers I – III
Title: Predicting cognitive and mental health traits and their polygenic architecture using large-scale brain connectomics

Short title: Connectome mapping of human traits and genetics

Luigi A. Maglanoc\textsuperscript{a,b}, Tobias Kaufmann\textsuperscript{b}, Dennis van der Meer\textsuperscript{b,c}, Andre F. Marquand\textsuperscript{d,e}, Thomas Wolters\textsuperscript{d}, Rune Jonassen\textsuperscript{f}, Eva Hilland\textsuperscript{b,g}, Ole A. Andreassen\textsuperscript{b}, Nils Inge Landrø\textsuperscript{a}, Lars T. Westlye\textsuperscript{b,h}

\textsuperscript{a}Clinical Neuroscience Research Group, Department of Psychology, University of Oslo, Oslo, Norway
\textsuperscript{b}NORMENT, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Norway
\textsuperscript{c}School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, The Netherlands
\textsuperscript{d}Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
\textsuperscript{e}Department of Neuroimaging, Institute of Psychiatry, King’s College London, London, UK
\textsuperscript{f}Department of Health Sciences, Oslo Metropolitan University, Oslo, Norway
\textsuperscript{g}Division of Psychiatry, Diakonhjemmet Hospital, Oslo, Norway
\textsuperscript{h}Department of Psychology, University of Oslo, Oslo, Norway

Corresponding author(s): Luigi A. Maglanoc, MSc, University of Oslo, Department of Psychology; PO Box 1094 Blindern, 0317 Oslo, Norway (luigi.maglanoc@psykologi.uio.no / luigi_maglanoc@hotmail.com)

Classification: BIOLOGICAL SCIENCES: Psychological and Cognitive Sciences

Keywords: fMRI, functional connectivity, human traits, polygenic scores, machine learning
Abstract

Cognitive abilities and mental disorders are complex traits sharing a largely unknown neuronal basis and aetiology. Their genetic architectures are highly polygenic and overlapping, which is supported by heterogeneous phenotypic expression and substantial clinical overlap. Brain network analysis provides a non-invasive means of dissecting biological heterogeneity yet its sensitivity, specificity and validity in clinical applications remains a major challenge. We used machine learning on static and dynamic temporal synchronization between all brain network nodes in 10,343 healthy individuals from the UK Biobank to predict (i) cognitive and mental health traits and (ii) their genetic underpinnings. We predicted age and sex to serve as our reference point. The traits of interest included individual level educational attainment and fluid intelligence (cognitive) and dimensional measures of depression, anxiety, and neuroticism (mental health). We predicted polygenic scores for educational attainment, fluid intelligence, depression, anxiety, and different neuroticism traits, in addition to schizophrenia. Beyond high accuracy for age and sex, permutation tests revealed above chance-level prediction accuracy for educational attainment and fluid intelligence. Educational attainment and fluid intelligence were mainly negatively associated with static brain connectivity in frontal and default mode networks, whereas age showed positive correlations with a more widespread pattern. In comparison, prediction accuracy for polygenic scores was at chance level across traits, which may serve as a benchmark for future studies aiming to link genetic factors and fMRI-based brain connectomics.
Significance

Although cognitive abilities and susceptibility to mental disorders reflect individual differences in brain function, neuroimaging is yet to provide a coherent account of the neuronal underpinnings. Here, we aimed to map the brain functional connectome of (i) cognitive and mental health traits and (ii) their polygenic architecture in a large population-based sample. We discovered high prediction accuracy for age and sex, and above-chance accuracy for educational attainment and intelligence (cognitive). In contrast, accuracies for dimensional measures of depression, anxiety and neuroticism (mental health), and polygenic scores across traits, were at chance level. These findings support the link between cognitive abilities and brain connectomics and provide a reference for studies mapping the brain connectomics of mental disorders and their genetic architectures.
Introduction

Mental health and cognitive abilities are both linked to brain function. Mental disorders such as depression and schizophrenia have joint high prevalence, early onset, and often a persistent nature (1). As such, identifying neuroimaging-based biomarkers which may facilitate early risk detection and interventions is a global aim potentially implicating millions worldwide (2–4). Twin and family studies have documented a strong genetic contribution across a range of human traits and mental disorders (5), and high polygenicity based on genome-wide associations studies (GWAS)(6–10). Educational attainment and fluid intelligence are two related and heritable phenotypes associated with a range of behaviors and outcomes (11) such as occupational attainment and social mobility (12), brain measures (13–15), and health (16, 17). Further, they show strong genetic correlations across cognitive domains and levels of ability (18).

Both the structural and functional architecture of the brain are heritable traits (19), and previous studies reporting aberrations in fMRI measures of brain connectivity in mental disorders are abundant (20–22). However, the generalizability, robustness and clinical utility of the findings have been questioned (23–25). This is likely due to differences in power (26), sample characteristics and analytical methods (27), but also the vast clinical heterogeneity of mental disorders (28–30). Further, although a substantial heritability of brain measures has been demonstrated (19, 31, 32), the sensitivity of fMRI features to the polygenic risk for mental disorders and associated personality traits is largely unknown.

Studies assessing the heritability and clinical associations with fMRI resting-state brain functional connectivity (FC) have typically targeted estimates of static FC (sFC), defined as the average temporal correlation between two brain regions. Less is known about the associations with dynamic properties of FC (dFC), conceptualized as the fluctuation in temporal correlations between two brain regions. Further, fMRI-based FC reflects the joint
contribution of partly independent sources oscillating at specific frequency bands (33), which are tied to a range of neural processes (34) and cognitive functions (35). Establishing the sensitivity and differential associations of imaging-based indices of brain function and connectivity to the genetic architecture of cognitive abilities and mental disorders may inform nosological and mechanistic studies and improve diagnostics, prevention and treatment.

Here, we tested the ability to detect mappings between static and dynamic measures of brain connectivity with (i) cognitive and mental health traits and (ii) genetic architecture of related traits based on polygenic score. These were compared with mappings of age and sex to help establish the results as a benchmark for clinical functional imaging. We used resting-state fMRI data from 10,343 healthy individuals from the UK Biobank (UKB) in a multivariate machine learning approach. The cognitive traits were years of educational attainment and fluid intelligence, and the mental health traits were dimensional measures of depression, anxiety and neuroticism. The polygenic scores in the same individuals included educational attainment, fluid intelligence, depression, anxiety, and 13 neuroticism traits. In addition, we included polygenic scores for schizophrenia, which is a disorder with high relevance for brain function, but currently not possible to study directly in the UKB due to the low number of cases with available MRI data. We employed cross-validation and evaluation of model performance to reduce bias and overfitting, in addition to permutation testing for statistical inference.

Results

Multivariate prediction of phenotypes

Fig. 1 shows the distribution of the phenotypes and cross-validated results of the overall best feature set (sFC) using 10-fold internal cross-validation with 100 repetitions (on 80% of the total sample). Permutation testing revealed above chance-level prediction accuracy for
phenotypic level educational attainment, fluid intelligence, age, (see Table 1) and sex (accuracy = 77.3%, corrected $p < 0.0027$, sensitivity = 87.7%, specificity = 32.4%). In comparison, predictions of dimensional measures of depression, anxiety and neuroticism anxiety performed at chance level (see Table 1). Fig. S1 and S2 provides the cross-validated results of the other FC feature sets. The feature set with all three FC-types (sFC, bandpass filtered sFC, and dFC) resulted in marginally higher prediction accuracy than the sFC feature set only for age ($t = 85.613, df = 170.83, p < 0.001$) and sex ($t = 75.785, df = 183.04, p < 0.001$). The model validation results (using 80% of the total sample to predict the leftover 20%) are very similar to the cross-validated results (see Fig. S3). Sensitivity analyses predicting age using the sFC feature set showed that there was very little difference when regressing scanner site out of the edges or excluding participants scanned in Newcastle (See methods and Fig. S4A). Further sensitivity analyses predicting fluid intelligence using sFC showed similar results when regressing out age (linear and quadratic), sex and head motion from the edges in the sFC feature set (see Fig. S4B).

Fig. 2 shows the grouping of brain networks based on hierarchical clustering and the top 20 edges in the sFC feature set for trait-level educational attainment and fluid intelligence based on feature importance, using correlation-adjusted marginal correlation (CAR) scores. Briefly, we observed mainly negative association with sFC for educational attainment, particularly within the cluster of default mode network (DMN) and frontal network nodes, and a similar brain functional pattern for fluid intelligence. In contrast, the CAR scores revealed largely positive associations between sFC and age, mainly within the cluster of motor/somatosensory and attentional networks, and within the cluster of default mode and frontal networks. Fig. S5 shows the top 40 edges that were positively associated with each phenotype, based on the feature set using all three FC-types, suggesting that these are largely non-overlapping. In short, based on feature importance, we observed roughly equal amounts
of edges across all three FC-types for educational attainment. We observed only two edges based on bandpass filtered sFC between visual networks and the cluster of default mode and frontal networks for fluid intelligence and only three edges based on dFC for age. Table S1 shows the correlation between the FC-types for age, fluid intelligence and educational attainment.

Multivariate prediction of polygenic scores

Fig. S6 shows the distribution of each polygenic score, and Fig. S7 shows a correlation plot with a dendrogram based on hierarchical clustering across all polygenic scores. Model performance for all polygenic scores (depression, anxiety, schizophrenia, neuroticism traits, educational attainment and intelligence) had negative explained variance based on the coefficient of determination ($R^2$) across feature sets (Fig. S8). As such, we only performed permutations testing for the polygenic scores that showed higher correlation-based model performance, including educational attainment, fluid intelligence and worry (Table 1). Fig. 3 provides an illustrative example of relative high model performance (phenotypic level educational attainment) and low model performance (educational attainment polygenic scores) showing both cross-validated results on empirical and permuted data. The model validation results are very similar to the cross-validated results (Fig. S9). There was very little difference when controlling for population stratification by regressing out the first ten genetic principal components for the sFC feature set predicting educational attainment polygenic risk scores (Fig. S4C). Cross-validated results were very similar using a lower polygenic score threshold ($p \leq 0.05$: Fig. S10), and after performing principal component analysis (PCA) across p-value thresholds on each of the above individual polygenic scores (36) (Fig. S11). See Supplemental Results for the correlation between phenotypes for specific feature sets, and phenotypes with their corresponding polygenic scores.
Fig. 1. (A) The distribution of all the phenotypes we used in the multivariate analyses. (B) The coloured density distributions represent the cross-validated results (across 100 repetitions) of the best FC feature set (sFC: the mean is denoted by the white circles) for a given phenotype, using R² as our metric of model performance. The grey distributions represent the mean model performance for each of the 10,000 permuted datasets. Although RMSE is our main metric of model performance, R² allows for comparison across phenotypes and polygenic risk scores. R² is negative for some of the models because we are testing it out-of-sample, meaning that the model can be arbitrarily bad. (C) Associations between raw and predicted scores for phenotypic level intelligence (left) and polygenic score (right). Education pertains to the years of educational attainment.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>RMSE</th>
<th>p-value</th>
<th>Correlation</th>
<th>MAE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive and mental health trait</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational attainment</td>
<td>4.739</td>
<td>&lt;0.0027</td>
<td>0.1634</td>
<td>4.348</td>
<td>0.0164</td>
</tr>
<tr>
<td>Fluid intelligence</td>
<td>2.043</td>
<td>&lt;0.0027</td>
<td>0.2014</td>
<td>1.638</td>
<td>0.0317</td>
</tr>
<tr>
<td>Age</td>
<td>6.482</td>
<td>&lt;0.0027</td>
<td>0.4947</td>
<td>5.317</td>
<td>0.2428</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>0.984</td>
<td>≥1</td>
<td>0.0924</td>
<td>0.8201</td>
<td>-0.0078</td>
</tr>
<tr>
<td>Depression</td>
<td>3.383</td>
<td>≥1</td>
<td>0.0947</td>
<td>2.353</td>
<td>-0.0126</td>
</tr>
<tr>
<td>Anxiety</td>
<td>3.222</td>
<td>≥1</td>
<td>0.0616</td>
<td>2.270</td>
<td>-0.0220</td>
</tr>
<tr>
<td>Polygenic scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational attainment</td>
<td>0.997</td>
<td>≥1</td>
<td>0.0941</td>
<td>0.7926</td>
<td>-0.0061</td>
</tr>
<tr>
<td>Fluid intelligence</td>
<td>1.002</td>
<td>≥1</td>
<td>0.0553</td>
<td>0.8009</td>
<td>-0.0163</td>
</tr>
<tr>
<td>Worry</td>
<td>1.007</td>
<td>≥1</td>
<td>0.0670</td>
<td>0.8170</td>
<td>-0.0138</td>
</tr>
</tbody>
</table>

Table 1: Model performance measures for all the phenotypes and polygenic scores with high correlation-based model performance using the sFC feature set. Worry is one of the 13 neuroticism traits. Correlation was based on spearman’s rho for the phenotypes and Pearson’s r for the polygenic scores. The p-value is calculated using permutation testing (based on RMSE), with an additional Bonferroni correction for 7 phenotypes and 20 polygenic score models.
Fig. 2. The top 20 edges based on CAR-scores for the best FC feature set (sFC). Cross-fold validation for age, phenotypic years of educational attainment and fluid intelligence. The thickness of the lines represents the relative importance of each edge for that specific feature set. Red lines indicate that the feature is positively associated with the phenotype when considering all the features. Blue lines indicate that the feature is negatively associated with the phenotype when considering all the features. The brain networks are grouped into 3 clusters which roughly represent the motor/somatosensory and attentional networks (red), default mode and frontal networks (green), visual components (blue).

Fig. 3. Cross-validated results comparing phenotypic level years of educational attainment and polygenic scores. For each plot, the blue represents model performance for each repetition (100) of the empirical data, and orange represents the mean model performance for each of the 10,000 permuted datasets (orange) based on (A) RMSE, (B) Correlation, (C) mean absolute error (MAE), and (D) R². The lines denote the mean of the respective distributions.
Discussion

Recent advances have led to an increasing interest in mapping the brain connectomic landscape underlying (i) cognitive and mental health traits and (ii) their genetic underpinnings. To this end, we conducted a large-scale integration of advanced measures of fMRI-based brain functional connectivity in a population cohort of 10,343 healthy individuals. First, we show that individual differences in educational attainment and fluid intelligence (cognitive), which have high relevance for life satisfaction and real-world outcomes, are partly reflected in the temporal organization of the brain. Furthermore, we demonstrate high prediction accuracy for age and sex, which provides a reference point for brain connectomic mapping. In contrast, we found chance-level performance for individual level prediction of dimensional measures of depression, anxiety and neuroticism (mental health). Second, we found chance-level performance for polygenic scores of depression, anxiety, schizophrenia, neuroticism traits, educational attainment and intelligence based on large-scale GWAS.

We found significant prediction accuracy for educational attainment and the corresponding high feature importance for the DMN and frontal networks. We also obtained significant prediction accuracy of brain FC patterns for fluid intelligence, largely driven by negative correlations between intelligence and sFC amongst frontal and DMN networks. This indicates relative dedifferentiation between the DMN and frontal regions in individuals with higher intelligence. This pattern was largely overlapping with the results for educational attainment, consistent with a study using a subset of the current sample (14), and also reflected in a strong positive correlations between the estimated importance of the feature sets from the two prediction tasks.
Highest model performance across all traits was observed for age, in line with several studies showing high age-sensitivity of functional imaging features (37–40). In general, compared to educational attainment and intelligence, we observed a more distributed pattern of the edge-wise feature importance scores for the age prediction. This suggests that individual differences in cognitive abilities may be more confined to specific parts of the brain connectome than the effects of aging, which are more pervasive.

To date, there is a lack of functional imaging studies comparing the sensitivity and predictive accuracy of dynamic and static measures of brain connectivity. Our comparisons revealed similar relative trait ranking in prediction accuracies for the different edge conceptualizations. However, corresponding CAR scores revealed differential feature importance rankings, suggesting that the different edge definitions may capture partly non-overlapping variance. For age, whereas previous studies have shown that dFC decreases with aging (41, 42), we obtained noticeably poorer prediction accuracy for dFC than the other feature sets, implying that age is better characterized by average FC rather than dynamic FC.

Overall, our results suggest higher dFC within a frontal-DMN cluster in individuals with higher intelligence and educational attainment. However, we must be careful in over-interpreting the dFC and bandpass filtered sFC results, as they performed worse than the feature set with only sFC edges. One exception was that the feature set with all FC-types performed marginally better when predicting age and classifying sex compared to sFC alone, which may suggest that there is added predictive ability when the signal is strong. A recent study found that the discriminative properties of FC vary across parcellation schemes and frequencies, with ICA-based parcellations revealing greater discriminability at high frequencies compared to other parcellations (43).

We were not able to predict dimensional measures of depression or anxiety, consistent with a smaller independent study (44) and a recent large-scale multi-site study
or trait neuroticism. One explanation for the lack of brain FC associations, particularly with regards to depression is its diversity in symptom profiles (28), and phenomenological overlap with other clinical domains (46). Some studies have attempted to identify putative subgroups of patients by using data-driven clustering based on symptoms (44, 47) or brain functional connectivity patterns (48, 49). In line with recent work questioning the robustness and generalizability of such clustering approaches (23), our current results do not support a simple link between fMRI-based brain connectivity and sub-clinical manifestations of mental disorders.

To the best of our knowledge, this is the first study that uses a machine learning approach to predict polygenic scores based on resting-state brain FC measures. A recent study demonstrated significant prediction accuracy for polygenic scores for autism using grey matter volumes, but not for attention deficit-hyperactivity disorder, bipolar disorder or schizophrenia (50). However, the sample size was small compared to the current study. Large sample sizes are needed to establish reliable estimates for the neuronal correlates of measures with weak associations (51). Relatedly, the current findings illustrate that the correlation between raw and predicted scores may overestimate model performance, at least when model performance is relatively low. Unlike $R^2$, MAE and RMSE, correlation coefficients are affected by scaling and location of the predicted scores relative to the raw scores, and, in contrast to MAE and RMSE, assumes linearity and homoscedasticity. As a result, future studies using multivariate analyses to predict continuous measures should carefully consider the choice of metric used for evaluating model performance, and reporting converging evidence of model fit and robustness may increase the reliability of the findings.

Despite the overall low predictive accuracy of the polygenic scores, we observed a relatively high positive correlation between the edge-wise feature importance of trait level educational attainment and intelligence with their respective polygenic scores, indicating
some shared signal. It is conceivable that any genetic pleiotropy between complex human traits and the brain is stronger for other imaging modalities. Indeed, some studies have demonstrated an association between cortical structure and polygenic scores for schizophrenia (52, 53), which are supported by a recent large-scale UKB study reporting thinner fronto-temporal cortex with higher polygenic risk (36), suggesting shared mechanisms. However, the variance explained by polygenic scores, such as those for schizophrenia, in brain measures like gray matter volume, tend to be low, around a few percent (54).

The low predictive accuracy for polygenic scores was observed both when applying a liberal and a more conservative initial p-value threshold when computing the polygenic score, and when employing a PCA-based approach calculated across a wide range of thresholds. Hence, the low predictive accuracy cannot simply be explained by the number of independent single nucleotide polymorphisms (SNP) feeding into the cumulative score. Relatedly, the poor predictive accuracy may also be partly explained by the low amount of explained trait variance accounted for by the polygenic score. For example, in the most recently published GWAS the variance explained for educational polygenic scores on phenotypic educational attainment was 12% (6). In the same study, polygenic scores for cognitive performance explained 7-10% of the variance in cognitive performance. Further, based on case-control differences, the polygenic score for schizophrenia (8), MDD (9), and anxiety (55) explained at most 7%, 1.9%, 2.3% and 2.1% of variance in liability, respectively. The low predictive accuracy of the polygenic risk scores may partly be due to the heterogeneity of complex human traits and disorders, which may disguise true genetic pleiotropy between brain features and the polygenic architecture of complex traits. Dimensional approaches such as normative modeling (56, 57) or brain age prediction (58) may reveal stronger shared signal. Many studies have shown that complex traits and disorders have substantial genetic overlap.
(59, 60) which may dilute the signal-to-noise ratio and specificity of polygenic scores. Future studies applying GWAS results with more power and novel techniques for boosting the shared genetic signal between traits (61, 62) may reveal stronger evidence of pleiotropy between brain FC and the polygenic architecture of complex traits. Taken together with other inherent limitations (see Supplements), the converging pattern across metrics of model performance indicate chance-level performance of brain FC for polygenic scores.

In conclusion, based on fMRI data from 10,343 individuals we have demonstrated high prediction accuracy of fMRI-based brain connectivity for age and sex, as well as above chance accuracy for educational attainment and intelligence (cognitive). In contrast, we obtained chance-level prediction accuracy for dimensional measures of depression, anxiety and neuroticism (mental health), in addition to their genetic underpinnings based on the respective polygenic scores and polygenic risk for schizophrenia. These novel findings support the link between cognitive abilities and the brain organization of the brain functional connectome and provide a reference for imaging studies mapping the brain mechanisms of clinical and genetic risk for mental disorders.

Materials and Methods

Sample

We used the October 2018 release of UKB imaging data (63), comprising 12,213 individuals with resting-state fMRI scans that passed the quality control performed by the UKB (64). The study was funded by the Research Council of Norway, the South-Eastern Norway Regional Health Authority, and approved by the UKB Access Committee (Project No. 27412) and the South-Eastern Norway Regional Committees for Medical and Health Research Ethics (REC). All participants provided informed consent prior to enrolment. The exclusion criteria across all phenotypes and polygenic scores were individuals with an ICD-10 based mental or a
neurological disorder (N = 210) and non-Caucasians (N = 1659) yielding a total of 10,343 individuals. Most of these participants were MRI-scanned in Manchester, but some were MRI-scanned in Newcastle (N = 354).

Phenotype data

Neuroticism was calculated based on a previous implementation (65), while educational attainment was based on the “qualifications” variable (UKB field: 6138) following a recent imputation procedure (66). Fluid intelligence (UKB field: 20016) was defined as the sum of the number of correct answers on the verbal-numerical reasoning test (13 items). Data for all the aforementioned phenotypes were taken from instance 2 (imaging visit). Symptom load for depression and anxiety were based on the sum score from the 9-item Patient Health Questionnaire (PHQ-9) and the 7-item Generalized Anxiety Disorder (GAD-7) questionnaire, respectively. Both PHQ-9 and GAD-7 were taken from the online follow-up. Table S2 shows the specific number of individuals for each phenotype and all polygenic scores.

Polygenic risk scores

Procedures for DNA collection and genotyping in UKB have been described previously (67). Polygenic scores were calculated using PRSice v.1.25 (68) based on GWAS results for broad depression, probable MDD (69), diagnostic MDD (9), item-level and sum neuroticism (65), anxiety (70), and schizophrenia (8). In the case of polygenic scores based on item-level and sum neuroticism, we performed our own GWAS (see Supplemental Material) to avoid overlap in participants in the discovery and target dataset (i.e. fMRI sample). We used the same reasoning to run our own GWAS (see Supplemental) to derive polygenic scores for educational attainment (based on years of schooling) and fluid intelligence, following similar procedures to previous studies (6, 7). For the main polygenic score analyses, a liberal SNP
inclusion threshold was set at $p \leq 0.5$, as this typically explains more variance in the clinical phenotype (71). For sensitivity analyses, we repeated the same analyses using a more conservative threshold of $p \leq 0.05$. As an additional sensitivity analysis, we ran a PCA on the computed polygenic scores from thresholds $p \leq 0.001$ to $p \leq 0.5$ (at 0.001 intervals) based on a recent implementation (36).

**Image Acquisition and pre-processing**

Detailed description of the image acquisition, pre-processing, group-level ICA and dual regression can be found in a previous study (64). The majority of the MRI data used were obtained in Cheadle Manchester on a Siemens Skyra 3.0 T scanner (Siemens Medical Solutions, Germany) with a 32-channel head coil. A small number of scans ($N = 354$) were obtained at an identical scanner in Newcastle. FSL ([http://fsl.fmrib.ox.ac.uk/fsl](http://fsl.fmrib.ox.ac.uk/fsl)) was used for fMRI data preprocessing. Briefly this involved motion correction, high-pass temporal filtering, echo-planar image un warping, gradient distortion correction un warping, and removal of structured artifacts. Estimated mean relative in-scanner head motion (volume-to-volume displacement) was computed with MCFLIRT. Group-level ICA was carried using MELODIC based on 4,162 datasets with a dimensionality of 25. Four of these ICs were identified as noise and discarded, leaving a total of 21 ICs for analyses. Dual regression was performed to generate subject specific spatial maps and corresponding time series.

**Functional connectivity measures**

All FC measures were computed in MATLAB. SFC was computed both with the unfiltered and bandpass filtered time-series within 0.04-0.07 Hz. For both, a node-by-node connectivity matrix was created using partial correlations between the subsequent time-series (72),
resulting in 210 unique edges. These partial correlations were L1-regularized, with estimated regularization strength (lambda) at the subject level (73–75).

For dFC we used a phase-based method (76) within the 0.04-0.07 Hz frequency band (77). Briefly, this method is sensitive to the degree of coupling and de-coupling between pairs of brain networks across the scanning session, based on applying the Hilbert transform and subsequently the Kuramoto order on the node time-series.

**Statistical analysis**

All statistical analyses were performed in R version 3.4.2 (78). We z-normalized all polygenic scores prior to the multivariate analyses.

**Multivariate analysis of brain FC**

In our primary analyses, we tested to what extent various combinations of static and dynamic FC between all nodes in the an extended brain network could predict the phenotypes, by using shrinkage linear regression (79) implemented in the R-package ‘care’ ([http://strimmerlab.org/software/care](http://strimmerlab.org/software/care)). See the original paper for details on the optimization of shrinkage parameters. Firstly, 80% of the data was used as the training set while the remaining 20% of the data was used as the left-out test set. We ran 10-fold internal cross-validation on the training set (i.e. based on iteratively using 90% of the sample to predict the remaining 10%), repeated 100 times on randomly partitioned data. We computed RMSE as our main measure of model performance for all phenotypes, but also MAE and R² (explained variance or model fit). R² was mainly used to visualize and compare model performance across models. We also used spearman’s rho as a measure for model performance for predicting these phenotypes. For all phenotypes, we computed Spearman’s rho, but Pearson correlation coefficient (r) for age. For sex, we used logistic regression for classification and
accuracy, sensitivity and specificity as markers of model performance. The same model parameters were used in the model validation, by using the whole training set to predict the phenotypes in the left-out test set. Statistical significance for phenotypes were assessed for the best feature set with a positive $R^2$ using permutation-based testing (10,000 permutations) based on RMSE, with an additional Bonferroni correction for 7 phenotypes and 20 polygenic score models. The mean model performance from cross-validation of the empirical data was used as our point estimate. For the phenotype models that were statistically significant, we determined the relative importance of each edge by computing the mean CAR scores (80) in cross-validation. We also assessed the correlation amongst these models based on the CAR-scores. To assess the confounding effects of age, sex and head motion, we repeated the analyses for fluid intelligence by regressing these confounders from the edges in the sFC feature set. We used the same framework to predict polygenic scores, computing Pearson’s $r$ in addition to the other measures of model performance, and model validation for the main PGRS analyses ($p \leq 0.5$). To assess the effect of population stratification, we regressed out the first 10 genetic principal components genetic principal components from the edges of the best feature set for one of the polygenic score models. We also assessed the correlation amongst phenotypes and corresponding polygenic scores based on the CAR-scores in the cases where the phenotype model was statistically significant.
Acknowledgments

The authors were funded by the Research Council of Norway (213837, 229129, 204966, 249795, 251134), the South-Eastern Norway Regional Health Authority (2014097, 2015073, 2016083, 2017112), the Department of Psychology, University of Oslo and the KG Jebsen Stiftelsen (223723). This research has been conducted using the UK Biobank Resource (access code 27412). The permutation testing was performed using resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway.
References


Title: Multimodal fusion of structural and functional brain imaging in depression using linked independent component analysis

Short title: Multimodal brain imaging in depression

Luigi A. Maglanoc\textsuperscript{a,b}, Tobias Kaufmann\textsuperscript{b}, Rune Jonassen\textsuperscript{c}, Eva Hilland\textsuperscript{a,d}, Dani Beck\textsuperscript{b,e}, Nils Inge Landrø\textsuperscript{a}, Lars T. Westlye\textsuperscript{b,e}

\textsuperscript{a}Clinical Neuroscience Research Group, Department of Psychology, University of Oslo, Oslo, Norway

\textsuperscript{b}NORMENT, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Norway

\textsuperscript{c}Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway

\textsuperscript{d}Division of Psychiatry, Diakonhjemmet Hospital, Oslo, Norway

\textsuperscript{e}Department of Psychology, University of Oslo, Oslo, Norway

* Corresponding authors: Luigi A. Maglanoc (luigi_maglanoc@hotmail.com) & Lars Westlye (lt.westlye@psykologi.uio.no), Department of Psychology, University of Oslo, Pb. 1094, Blindern, 0317 OSLO, Norway

Keywords: depression, multimodal MRI, linked independent component analysis, machine learning, heterogeneity
Abstract

Background: Previous structural and functional neuroimaging studies have implicated distributed brain regions and networks in depression. However, there are no robust imaging biomarkers that are specific to depression, which may be due to clinical heterogeneity and neurobiological complexity. A dimensional approach and fusion of imaging modalities may yield a more coherent view of the neuronal correlates of depression.

Methods: We used linked independent component analysis to fuse cortical macrostructure (thickness, area, gray matter density), white matter diffusion properties and resting-state fMRI default mode network amplitude in patients with a history of depression (n = 170) and controls (n = 71). We used univariate and machine learning approaches to assess the relationship between age, sex, case-control status, and symptom loads for depression and anxiety with the resulting brain components.

Results: Univariate analyses revealed strong associations between age and sex with mainly global but also regional specific brain components, with varying degrees of multimodal involvement. In contrast, there were no significant associations with case-control status, nor symptom loads for depression and anxiety with the brain components, nor any interaction effects with age and sex. Machine learning revealed low model performance for classifying patients from controls and predicting symptom loads for depression and anxiety, but high age prediction accuracy.

Conclusion: Multimodal fusion of brain imaging data alone may not be sufficient for dissecting the clinical and neurobiological heterogeneity of depression. Precise clinical stratification and methods for brain phenotyping at the individual level based on large training samples may be needed to parse the neuroanatomy of depression.
Introduction

With an estimated prevalence of 4.4%, depression affects more than 300 million worldwide (World Health Organization, 2017) and is a substantial contributor to disability and health loss (Friedrich, 2017). Identifying useful imaging based and other biomarkers to aid detection of individuals at risk for depression and facilitating individualized treatment is a global aim (Cuthbert & Insel, 2012; Insel, 2014, 2015).

A host of studies across a range of neuroimaging modalities have implicated various brain regions and networks in depression. Meta-analyses of structural magnetic resonance imaging (MRI) studies have suggested thinner orbitofrontal (OFC) and anterior cingulate cortex (ACC) in patients with depression compared to healthy controls (Lai, 2013; Schmaal et al., 2017; Suh et al., 2019). A large-scale meta-analysis comprising 2148 patients and 7957 controls from 20 different cohorts reported slightly smaller hippocampal volumes in patients with depression compared to controls (Schmaal et al., 2016), but the overall pattern of results suggested substantial heterogeneity and otherwise striking similarity across groups for all other investigated subcortical structures (Fried & Kievit, 2016). A meta-analysis of diffusion tensor imaging (DTI) studies including 641 patients and 581 healthy controls reported fractional anisotropy (FA) reductions in the genu of the corpus callosum and the anterior limb of the internal capsule (Chen et al., 2016), implicating interhemispheric and frontal-striatal-thalamic connections among the neuronal correlates of depression. Supporting the relevance of brain connectivity in mood disorders, resting-state fMRI studies have reported aberrant connectivity within the default mode network (DMN) in patients with depression compared to healthy controls (Kaiser, Andrews-Hanna, Wager, & Pizzagalli, 2015; Mulders, van Eijndhoven, Schene, Beckmann, & Tendolkar, 2015; Yan et al., 2019).

However, despite meta-analytical evidence suggesting brain aberrations in large groups of patients with depression, the reported effect sizes are small and the direct clinical
utility is unclear (Müller et al., 2017; Paulus & Thompson, 2019). One explanation for the lack of robust imaging-based markers in depression may be that previous studies have either focused on a single imaging modality or have analyzed different imaging modalities along separate pipelines and thus failed to model the common variance across features. In contrast, linked independent component analysis (LICA: Groves, Beckmann, Smith, & Woolrich, 2011; Groves et al., 2012) offers an integrated approach by fusing different structural and functional imaging modalities (Groves et al., 2012). LICA identifies modes of variation across modalities and disentangles independent sources of variation that may account both for large and small parts of the total variance, that may otherwise be overlooked by conventional approaches. By decomposing the imaging data into a set of independent components, LICA enables an integrated perspective that may improve clinical sensitivity compared to unimodal analyses (Alnæs et al., 2018; Doan, Engvig, Persson, et al., 2017; Francx et al., 2016; Wu et al., 2019).

Apart from the predominantly unimodal approaches in previous imaging studies, large individual differences and heterogeneity in the configuration and load of depressive symptoms represent other factors that could explain the lack of robust imaging markers. Symptom-based approaches have revealed more than 1000 unique symptom profiles among 3703 depressed outpatients based on only 12 questionnaire items (Fried & Nesse, 2015), suggesting large heterogeneity. Additionally, depression is highly comorbid with anxiety, with reported rates exceeding 50% (Johansson, Carlbring, Heedman, Paxling, & Andersson, 2013; Lamers et al., 2011). Furthermore, depression can be conceptualized along a continuum including individuals of the general, healthy population that may experience transient symptoms to varying degrees, and thus warrants a dimensional approach.

The main aim of the current study was to determine whether fusion of neuroimaging modalities would capture modes of brain variations which discriminate between patients with
a history of depression (n = 170) and healthy controls with no history of depression (n = 71),
and which are sensitive to current symptoms of depression and anxiety across groups. To this
end, we used LICA to combine measures of cortical macrostructure (cortical surface area and
thickness, and grey matter density), white matter diffusion properties (DTI-based FA, MD
and RD), and resting-state fMRI DMN amplitude.

There is evidence of sex and age differences in the prevalence and clinical
characteristics of depression, including lower age at onset of first major depressive episode in
women compared to men (Marcus et al., 2005), and longer duration of illness and different
symptoms in older compared to younger patients (Husain et al., 2005), which may reflect
differential neuronal correlates. Therefore, we tested for main effects of age and sex and their
interactions with the resulting brain components’ subject weights on group and symptoms.

In addition, we assessed the overall clinical sensitivity of all measures combined using
machine learning to classify patients and controls and to predict symptom loads for
depression and anxiety, which we compared with age prediction. Based on the above
reviewed studies and current models we anticipated 1) that brain variance related to
depression would be captured in components primarily reflecting the previously extended
functional neuroanatomy of depression, including limbic and fronto-temporal networks and
their connections. Irrespective of having a history of depression, we hypothesized 2) several
strong age and sex differences, reflecting well documented age and sex-related variance in
brain structure, including global thickness and volume reductions with increasing age, and
larger brain volume and surface area in men compared to women. To the extent that having a
history of depression interacts with sex and age-related processes in the brain, we
hypothesized 3) interactions between the age-related trajectories and sex differences
identified above with case-control status or symptoms of depression. To increase robustness
and generalizability we corrected for multiple comparisons across all univariate analyses and performed cross-validation and robust model evaluation in the machine learning analyses.

**Materials and Methods**

**Sample**

Patients (n = 194) were primarily recruited from outpatient clinics, while healthy controls (n = 78) were recruited through posters, newspaper advertisements and social media. The patient group was drawn from two related clinical trials (ClinicalTrials.gov ID NCT0265862 and NCT02931487). All participants were evaluated with the Mini International Neuropsychiatric Interview (M.I.N.I 6.0: Sheehan et al., 1998). Exclusion criteria for all participants were MRI contraindications and a self-reported history of neurological disorders. The study was approved by the Regional Ethical Committee of South-Eastern Norway (REK Sør-Øst), and we obtained a signed informed consent from all the participants. Symptom loads for depression and anxiety were evaluated using the Becks Depression Inventory (BDI-II; Beck, 1996) and the Becks Anxiety Inventory (BAI; Beck & Steer, 1993) respectively. The demographics for the final sample (after exclusions, see below) are shown in table 1. The range of symptom load for depression and anxiety for the control group was from 0 to 20 and 0 to 19 respectively, while the range for the patient group was from 0 to 51 and 0 to 45 respectively (see Figure 1 for the distributions).

**Image Acquisition**

MRI data was obtained on a 3T Philips Ingenia scanner (Phillips Healthcare) at the Oslo University Hospital using a 32-channel head coil. The same protocol was used for all participants, but there was a change in the phase-encoding direction during the course of
study recruitment and data collection which affected the T1-weighted data for 4 controls and 95 patients, and resting-state fMRI data for 64 patients.

T1-weighted data was collected for 74 controls and 194 patients using a 3D turbo field echo (TFE) scan with SENSE using the following parameters: acceleration factor = 2; repetition time (TR)/echo time (TE)/flip angle (FA): 3000 ms/3.61 ms/8°; scan duration: 3 min 16 s, 1 mm isotropic voxels.

Diffusion weighted data was collected for 72 controls and 184 patients using a dual spin echo, single-shot EPI sequence with the following parameters was used: TR/TE = 7200/86.5ms, FOV = 224 × 224 mm², 112 × 112 matrix, 2.0mm isotropic voxels; 32 volumes with non-collinear directions (b = 1000s/mm²). Additionally, we acquired two b = 0 volumes with opposite phase polarity (blip up/down volumes).

Resting-state fMRI data was collected using a T2* weighted single-shot gradient echo EPI sequence was acquired for 72 controls and 178 patients with the following parameters: TR/TE/FA = 2500ms/30ms/80°; 3.00 mm isotropic voxels; 45 slices, 200 volumes; scan time ≈ 8.5 min. Participants were instructed to have their eyes open, and refrain from falling asleep.

Structural MRI preprocessing

Vertex-wise cortical thickness and surface area measures (Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999) were estimated based on the T1-weighted scans using FreeSurfer (http://surfer.nmr.mgh.harvard.edu) (Fischl et al., 2002). Details are described elsewhere (Dale et al., 1999; Fischl et al., 1999) but in short, after gray/white boundary and pial reconstruction, cortical thickness was defined as the shortest distance between the surfaces vertex-wise (Dale et al., 1999), before resampling to the Freesurfer common template (fsaverage, 10,242 vertices; Fischl et al., 1999). The vertex-wise expansion or
compression was used to calculate vertex-wise maps of arealization. None of the thickness nor surface area data for healthy control (n = 74) nor patients (n = 194) were excluded after visual QC.

Voxel-based morphometry

Grey matter density maps (GMD) were created based on voxel-based morphometry (VBM) using the computational anatomy toolbox (CAT12: http://www.neuro.uni-jena.de/cat/) within SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). This involved brain-extraction, gray matter-segmentation, and then registration to MNI152 standard space. The resulting images were averaged and flipped along the x-axis to create a left-right symmetric, study-specific grey matter template. The modulated gray matter maps were smoothed with a sigma of 4 mm (FWHM = 9.4 mm). None of the GMD data for healthy control (n = 74) nor patients (n = 194) were excluded after visual QC.

DTI preprocessing

Processing steps included correction for motion and geometrical distortions based on the two b = 0 volumes and eddy currents by using FSL topup (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TOPUP) and eddy (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/eddy). We also used eddy to automatically identify and replace slices with signal loss within an integrated framework using Gaussian process (Andersson & Sotiropoulos, 2016), which substantially improved the temporal signal-to-noise ratio (tSNR: Roalf et al., 2016) (t = 24.139, p < 0.001, Cohen’s d = 2.13). We fitted a diffusion tensor model using dtifit in FSL to generate maps of fractional anisotropy (FA), mean diffusivity (MD) and radial diffusivity (RD). Based on manual QC, we excluded 3 subjects due to insufficient brain coverage, and 3 subjects due to poor data quality. One
additional subject was flagged with a tSNR of > 2 SD lower than the mean and discarded after additional manual QC. This yielded a total number of DTI scans of 71 healthy controls and 178 patients.

Resting-state fMRI preprocessing

Resting-state fMRI data was processed using the FSL’s FMRI Expert Analysis Tool (FEAT). This included co-registration with T1 images, brain extraction, motion correction (MCFLIRT: Jenkinson, Bannister, Brady, & Smith, 2002), spatial smoothing (FWHM = 6 mm), high pass filtering (100s), standard space registration (MNI-152) with FLIRT, and single-session independent component analysis (ICA; MELODIC). Automatic classification and regression of noise components was done using ICA-based Xnoiseifier (FIX: Griffanti et al., 2014; Salimi-Khorshidi et al., 2014), with a threshold of 60. FIX substantially improved tSNR ($t = 20.89$, $p < 0.001$, Cohen’s $d = 1.95$), and no fMRI scans from healthy controls ($n = 72$) nor from patients ($n = 178$) were excluded. Group-level ICA with model order fixed at 40 was performed on a balanced subset of healthy controls and patients ($N = 72$ from each group), which has been used in a previous study (Maglanoc et al., 2019). Dual regression (Nickerson, Smith, Öngür, & Beckmann, 2017) was used to estimate spatial maps and corresponding time-series of all components. We then identified an IC representing the canonical DMN (Supplemental Figure 1) and used the individual DMN spatial maps from dual regression in multimodal decomposition using LICA.

LICA

We used FMRIB’s LICA (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLICA) to perform data-driven multi-modal fusion, which evaluates shared inter-subject variations across the brain imaging measures (Groves et al., 2011, 2012). This produces spatial maps based on the commonalities
across features (e.g. GMD, DTI measures, DMN maps) and subjects, and corresponding subject weights (i.e. the degree to which a subject contributes to a LICA component). We included complete data from 70 patients and 171 controls in the decomposition. We chose a relatively low model order of 40 based on previous recommendation of estimating robust components (Wolfers et al., 2017), and the biological meaningfulness of the spatial maps. For transparency and comparison, we also performed similar analysis using a higher dimensionality (80, more details in Supplemental). For both model orders, we discarded components which were highly driven by one subject (threshold: > 20%) yielding a total of 40 and 67 components respectively (Supplemental Figure 2). One component was strongly associated with phase encoding direction (IC4 in both decompositions, \(t = 33.07\) and \(t = 32.47, p < 0.001\) respectively, Supplemental Figures 3 and 4) but not removed from the analyses because of the biologically meaningful spatial patterns.

### Table 1. Demographics of the final sample.

<table>
<thead>
<tr>
<th></th>
<th>Controls (n = 70)</th>
<th>Patients (n = 171)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (female, %)</td>
<td>46 (66)</td>
<td>120 (70)</td>
<td>0.599</td>
</tr>
<tr>
<td>Age (mean, SD)</td>
<td>41.8 (13.1)</td>
<td>38.7 (13.3)</td>
<td>0.092</td>
</tr>
<tr>
<td>Education level ISCED (mean)</td>
<td>6.0 (1.0)</td>
<td>5.9 (1.2)</td>
<td>0.932</td>
</tr>
<tr>
<td>Depression symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDI-II (mean, SD)</td>
<td>1.6 (3.0)</td>
<td>11.6 (10.4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Anxiety symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAI (mean, SD)</td>
<td>1.7 (2.8)</td>
<td>8.1 (8.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUDIT (mean, SD)</td>
<td>4.8 (3.3)</td>
<td>6.3 (5.0)</td>
<td>0.127</td>
</tr>
<tr>
<td>DUDIT (mean, SD)</td>
<td>0.5 (1.9)</td>
<td>0.8 (2.5)</td>
<td>0.123</td>
</tr>
<tr>
<td>Left Handedness (N)</td>
<td>7</td>
<td>6</td>
<td>0.087</td>
</tr>
<tr>
<td>History of anxiety disorder (N)</td>
<td>1</td>
<td>50</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>History of (hypo)mania (N)</td>
<td>0</td>
<td>23</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>History of other Axis-I disorders (N)</td>
<td>0</td>
<td>23</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No. of major depressive episodes (mean, SD)</td>
<td>0</td>
<td>4.4 (5.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Currently medicated (SSRI, N)</td>
<td>0</td>
<td>52</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Fig. 1. Histogram of symptom loads of (A) depression based on BDI-II and (B) anxiety based on BAI.

Statistical analysis

Statistical analyses were performed in R version 3.5.1 (R Core Team, 2018) and Matlab 2014A (The MathWorks). We used linear models to test for main effects of clinical characteristics (case-control status, symptoms), age, and sex on each LICA subject weight with each IC as the dependent variable. In additional models we tested for interactions between age or sex and clinical characteristics (case-control status, symptoms) on each IC. For the analyses involving symptoms, one healthy control was removed due to missing data. We included phase encoding direction as an additional covariate in all the univariate analyses, and we controlled the false discovery rate (FDR) across tests using p.adjust in R.

Machine learning approach

For group classification we submitted all LICA subject weights to shrinkage discriminant analysis (Ahdesmäki & Strimmer, 2010) in the R-package ‘sda’
For the main analyses we used the residuals of each component’s subject weight after regressing out age and sex, and additionally, phase encoding for IC4. As a supplemental analysis, we used the residuals of the subject weights after regressing out age, sex and phase encoding direction from all the ICs. For robustness and to reduce overfitting, we performed cross-validation with 10 folds across 100 iterations. We calculated area under the receiver operating curve (AUC) as our main measure of model performance using the R-package ‘pROC’ (Robin et al., 2011), but also accuracy, sensitivity and specificity. The relative feature importance was determined by calculating correlation-adjusted t-scores (CAT scores: Ahdesmäki & Strimmer, 2010). We determined statistical significance based on AUC using permutation-based testing across 10,000 iterations. We used the same framework to predict depression and anxiety symptoms, but by implementing shrinkage linear estimation (Schäfer & Strimmer, 2005) in the R-package ‘care’ (http://strimmerlab.org/software/care). Here, we computed root mean squared error (RMSE) between the raw and predicted scores as our main measure of model performance, but also mean absolute error (MAE), spearman’s rho, and $R^2$. In this case, the relative feature importance was determined by computing the mean correlation-adjusted marginal correlation (CAR) scores (Zuber & Strimmer, 2011). Here, statistical significance was based on RMSE and permutation testing. As a comparison, we also predicted age using the same framework, using residuals of the subject weights after regressing out phase encoding direction in the methods shown above (using Pearson’s r instead of spearman’s rho).

**Results**

**LICA**

Figure 2A shows the degree of fusion across MRI measures for each component. Figure 2B shows the percentage of the total variance explained by each IC. Most of the components
were characterized by region-specific features that were mainly bilateral, with the exception of 4 global components shown in Figure 3. Briefly, there was no substantial fusion between DMN maps and the other modalities, except for IC26 (Supplemental Figure 5).

**Fig 2.** (A) The degree of fusing across MRI measures. (B) Explained percentage variance of each IC
Fig 3. ICs that are mainly dominated by global features. For each IC, only measures that have an interpretable spatial pattern are presented. A z-score threshold of $\geq 3$ was used for illustration. For visualization of the skeleton-based ICs, we used tbss_fill. IC1: global GMD and surface area. IC2: global white matter microstructure. IC5: DMN amplitude. IC7: global thickness.
**Univariate analyses**

Table 2 shows results from linear models testing for main effects of group, age, sex, and symptom load for depression and anxiety on each IC. Supplemental Table 1 shows results from linear models testing for interactions between group or symptom loads for depression or anxiety with age or sex. Briefly, after corrections for multiple comparisons, the analysis revealed no significant associations between ICs and group, nor symptom load for depression and anxiety. There were significant main effects of age and sex on 10 and 5 LICA components (see Figure 4), respectively, but no significant main effects of group, with similar results for the decomposition with 80 components (see Supplemental Tables 2 and 3). Figure 3 shows the global LICA components associated with age and sex (see Supplemental Figure 6 for associations with additional LICA components). Age was negatively associated with IC1, indicating lower GMD and cortical surface area globally with increasing age, positively associated with IC2, indicating lower FA globally with increasing age, negatively associated with IC5, indicating lower DMN amplitude with increasing age, and negatively associated with IC7, indicating thinner cortex globally with increasing age. The analyses revealed main effects of sex on IC1, indicating larger global surface area and higher GMD in men compared to women, and IC5, indicating that men had higher DMN amplitude. The analyses revealed no significant interaction effects between group or symptom loads for depression or anxiety with age or sex with any of the ICs.
<table>
<thead>
<tr>
<th>IC</th>
<th>Age (t, p)</th>
<th>Sex (t, p)</th>
<th>Group (t, p)</th>
<th>BDI-II (t, p)</th>
<th>BAI (t, p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC0</td>
<td>7.69 (&lt;0.001)</td>
<td>1.48 (0.287)</td>
<td>0.98 (0.883)</td>
<td>0.96 (0.853)</td>
<td>-0.04 (0.973)</td>
</tr>
<tr>
<td>IC1</td>
<td>-8.46 (&lt;0.001)</td>
<td>10.85 (&lt;0.001)</td>
<td>0.32 (0.883)</td>
<td>-0.59 (0.969)</td>
<td>0.308 (0.965)</td>
</tr>
<tr>
<td>IC2</td>
<td>-10.21 (0.202)</td>
<td>-6.95 (&lt;0.001)</td>
<td>0.57 (0.883)</td>
<td>-0.42 (0.969)</td>
<td>-0.29 (0.965)</td>
</tr>
<tr>
<td>IC3</td>
<td>-1.70 (0.202)</td>
<td>-1.87 (0.208)</td>
<td>0.86 (0.883)</td>
<td>1.10 (0.853)</td>
<td>1.40 (0.965)</td>
</tr>
<tr>
<td>IC4</td>
<td>-4.88 (&lt;0.001)</td>
<td>1.41 (0.307)</td>
<td>0.86 (0.883)</td>
<td>-0.14 (0.976)</td>
<td>0.33 (0.965)</td>
</tr>
<tr>
<td>IC5</td>
<td>-6.34 (&lt;0.001)</td>
<td>3.45 (0.005)</td>
<td>-1.25 (0.883)</td>
<td>-0.42 (0.969)</td>
<td>-0.75 (0.965)</td>
</tr>
<tr>
<td>IC6</td>
<td>-3.89 (0.001)</td>
<td>0.28 (0.841)</td>
<td>1.34 (0.883)</td>
<td>-2.23 (0.516)</td>
<td>-2.49 (0.536)</td>
</tr>
<tr>
<td>IC7</td>
<td>-3.56 (0.002)</td>
<td>-0.54 (0.759)</td>
<td>1.00 (0.883)</td>
<td>0.51 (0.969)</td>
<td>1.00 (0.965)</td>
</tr>
<tr>
<td>IC8</td>
<td>-1.64 (0.215)</td>
<td>0.35 (0.83)</td>
<td>1.32 (0.883)</td>
<td>0.31 (0.976)</td>
<td>1.06 (0.965)</td>
</tr>
<tr>
<td>IC9</td>
<td>0.81 (0.551)</td>
<td>-1.87 (0.208)</td>
<td>0.62 (0.883)</td>
<td>0.55 (0.969)</td>
<td>0.35 (0.965)</td>
</tr>
<tr>
<td>IC10</td>
<td>-0.61 (0.62)</td>
<td>-1.76 (0.245)</td>
<td>-0.21 (0.925)</td>
<td>0.47 (0.969)</td>
<td>-0.07 (0.973)</td>
</tr>
<tr>
<td>IC11</td>
<td>1.34 (0.33)</td>
<td>0.40 (0.83)</td>
<td>-1.26 (0.883)</td>
<td>0.11 (0.976)</td>
<td>-0.99 (0.965)</td>
</tr>
<tr>
<td>IC12</td>
<td>-2.06 (0.101)</td>
<td>0.84 (0.578)</td>
<td>0.37 (0.883)</td>
<td>-1.40 (0.853)</td>
<td>-0.94 (0.965)</td>
</tr>
<tr>
<td>IC13</td>
<td>-5.90 (&lt;0.001)</td>
<td>-4.24 (&lt;0.001)</td>
<td>-0.04 (0.968)</td>
<td>-0.32 (0.976)</td>
<td>-0.37 (0.965)</td>
</tr>
<tr>
<td>IC14</td>
<td>-1.11 (0.397)</td>
<td>1.17 (0.404)</td>
<td>0.21 (0.925)</td>
<td>0.08 (0.976)</td>
<td>0.62 (0.965)</td>
</tr>
<tr>
<td>IC15</td>
<td>0.10 (0.952)</td>
<td>5.25 (&lt;0.001)</td>
<td>-0.85 (0.883)</td>
<td>-0.42 (0.969)</td>
<td>-1.08 (0.965)</td>
</tr>
<tr>
<td>IC16</td>
<td>0.64 (0.62)</td>
<td>3.09 (0.015)</td>
<td>1.24 (0.883)</td>
<td>0.74 (0.969)</td>
<td>0.20 (0.965)</td>
</tr>
<tr>
<td>IC17</td>
<td>-0.97 (0.478)</td>
<td>1.27 (0.372)</td>
<td>-0.38 (0.883)</td>
<td>0.11 (0.976)</td>
<td>-0.42 (0.965)</td>
</tr>
<tr>
<td>IC18</td>
<td>0.51 (0.677)</td>
<td>-0.33 (0.83)</td>
<td>-0.90 (0.883)</td>
<td>-0.54 (0.969)</td>
<td>-0.51 (0.965)</td>
</tr>
<tr>
<td>IC19</td>
<td>-2.68 (0.026)</td>
<td>1.25 (0.372)</td>
<td>2.84 (0.198)</td>
<td>-1.79 (0.6)</td>
<td>0.51 (0.965)</td>
</tr>
<tr>
<td>IC20</td>
<td>-2.47 (0.044)</td>
<td>0.15 (0.925)</td>
<td>0.72 (0.883)</td>
<td>1.02 (0.853)</td>
<td>0.11 (0.973)</td>
</tr>
<tr>
<td>IC21</td>
<td>0.88 (0.527)</td>
<td>0.69 (0.653)</td>
<td>0.97 (0.883)</td>
<td>-0.95 (0.853)</td>
<td>-0.42 (0.965)</td>
</tr>
<tr>
<td>IC22</td>
<td>-0.63 (0.62)</td>
<td>-1.71 (0.254)</td>
<td>0.97 (0.883)</td>
<td>-0.14 (0.976)</td>
<td>-0.12 (0.973)</td>
</tr>
<tr>
<td>IC23</td>
<td>2.21 (0.075)</td>
<td>-1.62 (0.286)</td>
<td>0.45 (0.883)</td>
<td>-1.09 (0.853)</td>
<td>-1.29 (0.965)</td>
</tr>
<tr>
<td>IC24</td>
<td>1.18 (0.38)</td>
<td>-0.05 (0.96)</td>
<td>0.61 (0.883)</td>
<td>0.57 (0.969)</td>
<td>0.36 (0.965)</td>
</tr>
<tr>
<td>IC25</td>
<td>-0.27 (0.832)</td>
<td>1.48 (0.287)</td>
<td>0.54 (0.883)</td>
<td>-0.35 (0.976)</td>
<td>0.37 (0.965)</td>
</tr>
<tr>
<td>IC26</td>
<td>1.20 (0.38)</td>
<td>-0.44 (0.83)</td>
<td>0.60 (0.883)</td>
<td>-1.18 (0.853)</td>
<td>-0.40 (0.965)</td>
</tr>
<tr>
<td>IC27</td>
<td>0.09 (0.932)</td>
<td>0.32 (0.83)</td>
<td>-0.88 (0.883)</td>
<td>-1.45 (0.853)</td>
<td>-1.53 (0.965)</td>
</tr>
<tr>
<td>IC28</td>
<td>0.48 (0.68)</td>
<td>1.52 (0.287)</td>
<td>0.52 (0.883)</td>
<td>1.01 (0.853)</td>
<td>0.69 (0.965)</td>
</tr>
<tr>
<td>IC29</td>
<td>0.80 (0.551)</td>
<td>-1.98 (0.198)</td>
<td>-0.07 (0.968)</td>
<td>0.98 (0.853)</td>
<td>0.27 (0.965)</td>
</tr>
<tr>
<td>IC30</td>
<td>3.64 (0.002)</td>
<td>0.72 (0.653)</td>
<td>0.52 (0.883)</td>
<td>0.63 (0.969)</td>
<td>-0.03 (0.973)</td>
</tr>
<tr>
<td>IC31</td>
<td>-2.99 (0.011)</td>
<td>1.52 (0.287)</td>
<td>0.36 (0.883)</td>
<td>-0.11 (0.976)</td>
<td>-0.22 (0.965)</td>
</tr>
<tr>
<td>IC32</td>
<td>-1.72 (0.202)</td>
<td>-2.44 (0.078)</td>
<td>0.38 (0.883)</td>
<td>1.94 (0.542)</td>
<td>1.27 (0.965)</td>
</tr>
<tr>
<td>IC33</td>
<td>-5.16 (&lt;0.001)</td>
<td>-1.00 (0.488)</td>
<td>0.33 (0.883)</td>
<td>-0.001 (0.999)</td>
<td>0.65 (0.965)</td>
</tr>
<tr>
<td>IC34</td>
<td>-0.72 (0.587)</td>
<td>2.30 (0.1)</td>
<td>0.10 (0.965)</td>
<td>0.57 (0.969)</td>
<td>0.20 (0.965)</td>
</tr>
<tr>
<td>IC35</td>
<td>1.25 (0.373)</td>
<td>-0.05 (0.96)</td>
<td>0.15 (0.952)</td>
<td>1.16 (0.853)</td>
<td>1.73 (0.965)</td>
</tr>
<tr>
<td>IC36</td>
<td>1.44 (0.291)</td>
<td>-1.54 (0.287)</td>
<td>0.39 (0.883)</td>
<td>-2.26 (0.516)</td>
<td>-0.74 (0.965)</td>
</tr>
<tr>
<td>IC37</td>
<td>-1.45 (0.291)</td>
<td>-2.70 (0.044)</td>
<td>-1.01 (0.883)</td>
<td>0.06 (0.976)</td>
<td>-0.59 (0.965)</td>
</tr>
<tr>
<td>IC38</td>
<td>2.25 (0.073)</td>
<td>1.13 (0.418)</td>
<td>0.52 (0.883)</td>
<td>-2.08 (0.516)</td>
<td>-1.42 (0.965)</td>
</tr>
<tr>
<td>IC39</td>
<td>-1.16 (0.38)</td>
<td>0.98 (0.488)</td>
<td>-1.03 (0.883)</td>
<td>-0.25 (0.976)</td>
<td>-0.74 (0.965)</td>
</tr>
</tbody>
</table>
**Fig. 4.** Significant effects of age and sex on ICs ($p < 0.01$). A: Scatter plots of the significant (linear) effects of age on ICs, sorted by the strength of the association. For visualization purposes we plotted LOESS and separated based on case-control status. The IC subject weights in the plot have been residualized for group, sex and phase encoding direction. B: Violin plots showing the distribution of the subject weights within men and women for each of the components showing a significant main effect of sex. The subject weights in the plot have been residualized for group, age and phase encoding direction.

**Machine learning analyses**

Figure 5 shows the results of the machine learning analyses, with the spatial maps of a select few top features for each model shown in Supplemental Figure 7. Model performance was low for classifying patients and controls using residualized IC features (AUC = 0.5702, $p = 0.06135$, accuracy = 0.6169, sensitivity = 0.4292, specificity = 0.3009). The feature importance based on CAT-scores identified IC19 as the most important feature for classifying group. IC19 represents a covarying pattern of high GMD in most cerebellar regions, low GMD in cerebellar crus II, and high GMD in the angular gyri.

Model performance was low for predicting depression symptoms using residualized IC features (RMSE = 10.72, $p = 0.9236$, MAE = 8.498, $R^2 = -0.3302$, spearman’s rho =
IC0 had the highest feature importance based on CAR-scores (positive association). IC0 is characterized by a complex covarying pattern including low GMD in temporal regions, the thalamus and cingulate, and low thickness in the cingulate and fronto-temporal regions. In terms of white matter diffusion properties, IC0 is characterized by high FA in several pathways including the posterior thalamic radiation and low FA in the anterior thalamic radiation and fornix, with mostly the reverse pattern for MD and RD.

Model performance was low when predicting anxiety symptoms using residualized IC features (RMSE = 8.181, $p = 0.8946$, MAE = 6.262, $R^2 = -0.424$, Spearman’s rho = -0.064). IC6 had the highest feature importance based on CAR-scores (negative association). IC6 is mainly characterized by a complex covarying pattern of high FA in the splenium of the corpus callosum, high FA and low MD and RD in the fornix, high MD and RD in the thalamus, in addition to high GMD in the thalamus, and low GMD in hippocampal and amygdala regions. Model performance was slightly lower when regressing out phase encoding from all the IC features, and also suggested a different order of feature importance (see Supplemental Results and Supplemental Figure 8). Using the decomposition with higher model order revealed similar results in terms of feature importance, albeit slightly lower model performance for group classification, and symptom prediction (see Supplemental Results and Supplemental Figure 9). In contrast, model performance was high when predicting age (RMSE = 6.764, $p < 0.0001$, MAE = 5.530, $R^2 = 0.712$, $r = 0.861$) using residualized features, with feature importance generally in line with the univariate results (see Figure 6). Model performance for predicting age was high but slightly lower when regressing out phase encoding from all the IC features (see Supplemental Results and Supplemental Figure 10), and when using the decomposition with higher model order (see Supplemental Results and Supplemental figure 11).
Fig. 5. The results of the main analyses of the machine learning approach using 10-fold cross-validation across 100 repetitions for (A) classifying group (B) prediction symptom load for depression and (C) symptom load for anxiety. Here, phase encoding direction was only regressed out of the subject weights in IC4, while age and sex were regressed out from the subject weights of all the ICs. The figures on the left show prediction accuracy based on various model performance metrics. The barplots on the right show the most important features for each model based on CAT-scores (A) or CAR-scores (B and C).
Fig. 6. Age prediction regressing out phase encoding from the subject weights in IC4. (A) model performance results and (B) feature importance

Discussion

The distributed functional and structural neuroanatomy of complex traits and disorders warrants integrated perspectives and analytical approaches. To this end, we probed the neuronal correlates of depression using multimodal fusion across cortical macrostructure, white matter diffusion properties and DMN amplitude based on resting-state fMRI. LICA yielded 40 components with various degree of multimodal involvement and different anatomical distributions, including both global and regionally specific patterns. Univariate analyses revealed strong associations with age and sex for several components’ subject weights after multiple comparison correction, but no robust group differences between patients with a history of depression and healthy controls, and no significant interactions between group and sex nor between group and age. Likewise, we observed no robust associations with symptom loads for depression or anxiety, nor interactions with age or sex. In line with the univariate analyses, the machine learning approach revealed overall low prediction accuracy for group classification and prediction of symptom loads for depression and anxiety, but high prediction accuracy for age.
Our univariate analyses revealed no main effects of history of depression or symptom load for depression on any of the LICA components. The machine learning analyses here revealed overall low predictive value both for case-control status and symptoms of depression and anxiety, which is generally in line with the univariate analyses and an increasing body of literature suggesting small differences in brain structure between patients with MDD and healthy controls (Schmaal et al., 2017, 2016; Varoquaux, 2018; Wolfers, Buitelaar, Beckmann, Franke, & Marquand, 2015). While considering the overall low performance, the most important feature for classifying patients with a history of depression from healthy controls was a component encompassing covarying patterns of both high and low GMD in cerebellar regions (IC19). There is some evidence that depression is linked to cerebellar regions that communicate with networks related to cognitive and introspective processing (Depping, Schmitgen, Kubera, & Wolf, 2018). Cerebellar structural characteristics have recently been demonstrated to rank among the most sensitive brain features when comparing adult patients with schizophrenia and healthy controls (Moberget et al., 2018), and also for predicting psychiatric symptoms in youths (Moberget et al., 2019). The most important feature for predicting depression symptoms was IC0, which involves complex covarying patterns of low GMD and cortical thickness in mainly temporal but also frontal regions. This pattern is largely in line with previous research (Lai, 2013; Schmaal et al., 2017; Suh et al., 2019). IC0 also encompassed high FA and low MD and RD in interhemispheric connections and frontal-striatal thalamic pathways, in line with previous studies, albeit in the opposite direction (Chen et al., 2016). One study reported a positive association between symptom load for depression and FA in the thalamus (Osoba et al., 2013), and another study suggested this association (although in the opposite direction) is related to late onset MDD, especially in the corpus callosum (Cheng et al., 2014). However, while the implicated brain patterns are in
line with previous reports, the overall poor model performance for predicting symptom load for depression warrants caution when interpreting this finding.

Higher age was related to lower global cortical thickness (IC7), in line with previous studies (e.g. Fjell et al., 2015). As hypothesized, higher age was also associated with lower global volume and smaller surface area (IC1), similar to previous studies using LICA (Doan, Engvig, Zaske, et al., 2017; Douaud et al., 2014), and a consistent finding in lifespan studies. Additionally, advancing age was negatively associated with IC2, indicating decreased FA globally, but also increased RD and to some extent MD, consistent with the aging literature (Davis et al., 2009; Sexton et al., 2014; Westlye et al., 2010). Higher age was associated with IC5, reflecting age-related decreases in DMN amplitude, in line with previous research (Damoiseaux et al., 2008; Mevel et al., 2013; Mowinckel, Espeseth, & Westlye, 2012; Razlighi et al., 2014; Vidal-Piñeiro et al., 2014). We observed high prediction accuracy for age in the machine learning approach, which shows the potential utility of LICA in estimating the gap between chronological and biological age (i.e. brain age gap).

Men had larger global brain volume and surface area than women (IC1), which is consistent with previous studies (e.g. Ritchie et al., 2018). Additionally, women had higher subject weights in IC13, reflecting lower FA in the corticospinal tract, portions of the superior longitudinal fasiculi and posterior thalamic radiation compared to men, generally in line with a large-scale UK Biobank study (Ritchie et al., 2018). We also found that men had greater DMN amplitude (IC5) than women, which adds to previous inconclusive findings (Mowinckel et al., 2012; Weissman-Fogel, Moayedi, Taylor, Pope, & Davis, 2010) and contrasts a previous report suggesting effects in the opposite direction (Jamadar et al., 2018).

In general, our analyses did not provide support for our hypothesis that a history of depression and symptoms of depression interact with age-related trajectories or sex differences of the LICA components. To the best of our knowledge, although studies have
found specific cortical abnormalities related to adults with MDD, adolescents with MDD (Schmaal et al., 2017) and age at onset of depression (Ho et al., 2019), few or no studies have reported age-by-group interactions. Although there have been some early reports of a sex-by-group interaction in hippocampal volumes (e.g. Frodl et al., 2002), the recent large-scale ENIGMA MDD study reported no sex-by-group interactions in any subcortical volumes (Schmaal et al., 2016). Another recent study identified sex-by-group interactions using VBM, including higher GMD in the left cerebellum of male patients only, and lower GMD in the dorsal medial prefrontal gyrus in female patients only (Yang et al., 2017). However, the sample size was relatively small (less than 100 in the patient and control group each) which may affect the reproducibility of these findings. Despite separate reports of a link between resting-state DMN connectivity and rumination in depression (e.g. Hamilton, Farmer, Fogelman, & Gotlib, 2015) and evidence of sex-differences in rumination among adolescents (e.g. Jose & Brown, 2008), no studies have reported a sex-by-group interaction on DMN functional connectivity or amplitude.

The lack of positive findings in the univariate analyses and low predictive accuracy in the machine learning approach can be attributed to at least two factors. First, as illustrated by the large-scale ENIGMA studies (Ho et al., 2019; Schmaal et al., 2017, 2016), the effect sizes in neuroimaging studies of mental disorders and depression are overall small (Paulus & Thompson, 2019). Similarly, small sample sizes may contribute to over-inflated estimates of prediction accuracy in machine learning approaches (Wolfers et al., 2015). This is one possible explanation why other multimodal fusion studies of depression have reached higher prediction accuracies for classifying patients with depression from healthy controls (He et al., 2017; Ramezani et al., 2014; Yang et al., 2018), with patient groups consisting of no more than 60 individuals. Secondly, and also related to the small effect sizes, mental disorders including depression are clinically highly heterogenous. As an example, Müller and
colleagues (2017) partially attribute the lack of convergence in their meta-analysis of activation-based fMRI experiments involving 1058 MDD patients to clinical heterogeneity. As a result, future research probing the neurobiology of depression should aim for large sample sizes (Rutledge, Chekroud, & Huys, 2019), and more importantly, stratifying patients (Feczko et al., 2019) with depression at the individual level. Pursuant to this, there has been considerable interest in identifying clinically relevant subgroups based on brain imaging, with initially encouraging results (Drysdale et al., 2017). However, the robustness and generalizability of such studies have been brought into question (Dinga et al., 2019), which may be partly due to substantial brain heterogeneity within groups, which has been illustrated in terms of morphometry in schizophrenia (Alnæs et al., 2019). Alternatively, dimensional measures such as brain age prediction (Kaufmann et al., 2018) and normative modelling (Marquand et al., 2019; Marquand, Rezek, Buitelaar, & Beckmann, 2016) have shown promising results in elucidating brain heterogeneity in mental disorders such as schizophrenia (Wolfers et al., 2018) and attention deficit/hyperactivity disorder (Wolfers et al., 2019).

The current findings should be considered in light of relevant limitations associated with statistical power and study design. The relatively low number of severely depressed patients may have influenced the sensitivity and specificity of the machine learning approach, in particular for predicting symptom loads of depression and anxiety. The varied current use of antidepressants in the patient group may have impacted the classification accuracy, although it is undoubtedly difficult to get large samples of non-medicated patients.

One weakness of this study is that the change in phase encoding direction may have introduced systematic differences in the MRI signal. However, only one component (IC4 in both decompositions) was strongly sensitive to phase encoding direction, suggesting that the remaining components were largely unaffected. Furthermore, we accounted for phase encoding direction in both the univariate and the machine learning analyses. This along with
previous studies (Doan, Engvig, Persson, et al., 2017; Doan, Engvig, Zaske, et al., 2017; Groves et al., 2012) provides additional evidence that LICA is a promising tool to account for various scanner effects, particularly relevant for multi-site and longitudinal studies. We used a model order of 40 for LICA decomposition in the main analyses. Even though we did find similar feature importance ranking in the decomposition with higher model order, prediction accuracy was slightly lower across all models. Although there is no consensus on the optimal model order, this may imply that we were modelling more noise components in the higher model order decomposition, which was also supported by the number of discarded components due to dominance by a single subject.

In conclusion, based on fusion of structural, diffusion-weighted and resting-state fMRI data from 241 individuals with or without a history of depression, we identified multimodal and modality specific components that revealed strong associations with age and sex. None of the components showed significant association with categorical or dimensional measures of depression, nor any interaction effects with age and sex. Similarly, machine learning revealed low prediction accuracy for classifying patients from controls and predicting symptom loads. This study supports accumulating evidence of small effect sizes when comparing brain imaging features between patients with a history of depression and healthy controls, and indicates the need for more precise methods of stratifying individuals with depression, as well as large samples sizes.
Acknowledgments

This work was supported by the South-Eastern Norway Regional Health Authority (2014097, 2015073, 2015052), the Research Council of Norway (249795, 229135), and the Department of Psychology, University of Oslo. The permutation testing was performed using resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway.
References


https://doi.org/10.1371/journal.pone.0112307

https://doi.org/10.1186/1741-7015-11-126

https://doi.org/10.1006/nimg.1998.0395


https://doi.org/10.1016/j.neuroimage.2009.01.068

https://doi.org/10.3389/fpsyt.2018.00634

https://doi.org/10.1038/srep45131

https://doi.org/10.1016/j.neuroimage.2017.06.070

https://doi.org/10.1073/pnas.1410378111


https://doi.org/10.1006/nimg.1998.0396


studies. Biological Psychiatry, 80(7), 552–561.
https://doi.org/10.1016/j.biopsych.2015.12.023

https://doi.org/10.1016/j.neurobiolaging.2012.08.018

https://doi.org/10.1038/mp.2017.106

https://doi.org/10.1016/j.biopsych.2019.01.019


https://doi.org/10.1016/j.neubiorev.2015.07.014


https://doi.org/10.1001/jamapsychiatry.2016.2783

https://doi.org/10.3389/ fnins.2017.00115


https://doi.org/10.1001/jamapsychiatry.2018.4540


https://doi.org/10.1016/j.nicl.2014.11.016


