Identification of multimodal mental health signatures in the young population using deep phenotyping

Authors: Niels Mørch[1*], Andrés B. Calderón[1*], Timo L. Kvamme[2], Julie G. Donskov[1,4], Blanka Zana[2], Simon Durand[1,2], Jovana Bjekic[3], Maro G. Machizawa[4], Makiko Yamada[5], Filip A. Ottosson[6], Jonas Bybjerg-Grauholm[6], Madeleine Ernst[6], Anders D. Børglum[1,7], Kristian Sandberg[2], Per Qvist[1,7†]

Affiliations:

[1]Department of Biomedicine, Aarhus University, 8000 Aarhus C, Denmark.
[2]Center of Functionally Integrative Neuroscience, Aarhus University, 8000 Aarhus C, Denmark.
[3]Human Neuroscience department of the Institute for Medical Research, University of Belgrade, 11000 Belgrade, Serbia.
[4]Department of Functional Brain Imaging, Institute for Quantum Medical Science, National Institutes for Quantum Science and Technology, Chiba 263-8555, Japan.
[5]Center for Brain, Mind and KANSEI Sciences Research, Hiroshima University, Hiroshima 734-8551, Japan.
[6]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, 2300 Copenhagen, Denmark.[7]iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus 8000, Denmark.

*These authors contributed equally to this work

†Corresponding author. Email: per.q@biomed.au.dk

Abstract:

Background: Mental health encompasses emotional, psychological, and social dimensions, extending beyond the mere absence of illness. Shaped by a complex interplay of hereditary factors and life experiences, mental health can deteriorate into clinical conditions necessitating intervention. However, the ambiguity between pathological and non-pathological states, along with overlapping clinical profiles, challenges traditional diagnostic procedures, highlighting the need for a dimensional approach in stratified psychiatry.

Methods: We analyzed comprehensive phenotypic data from ~300 young Danish participants, including psychometric assessments, brain imaging, genetics, and circulatory OMICs markers. Using a novel psychometry-based archetyping approach, we employed soft-clustering analyses to stratify participants based on distinct cognitive, emotional, and behavioral patterns, while exploring their genetic and neurobiological underpinnings.

Results: Five psychometric archetypes were identified, representing a continuum of mental health traits. One archetype, characterized by high neuroticism, emotional dysregulation, and elevated stress and depression scores, was firmly associated with self-reported mental health diagnoses, psychiatric comorbidities, and family history of mental illness. Genetic predisposition to mental health conditions, reflected in polygenic scores (PGSs), accounted for up to 9% of the variance in archetypes, with significant contributions from neuroimaging-related PGSs. The overlaps between broader genetic profiles and archetypes further confirmed their biological foundations. Neuroimaging data linked the risk-associated archetype to both regional and global brain volumetric changes, while metabolomic analysis identified differentiating metabolites related to mood regulation and neuroinflammation.

Conclusions: This study demonstrates the feasibility of data-driven stratification of the general population into distinct risk groups defined by multimodal mental health signatures. This stratification offers a robust framework for understanding mental health variation and holds significant potential for advancing early screening and targeted intervention strategies in the young population.

Keywords: Mental health, Psychometric archetypes, Polygenic scores, Neuroimaging, Metabolomics, Precision psychiatry

INTRODUCTION

Mental health is a multifaceted and dynamic trait that extends beyond the absence of mental illness (MDx). It encompasses emotional, psychological, and social dimensions, affecting how individuals perceive, feel, and navigate life's challenges and rewards. Shaped by both genetic predispositions and life experiences[1], mental health can deteriorate into clinical conditions necessitating intervention. However, the line between pathological and non-pathological states is ambiguous, and overlaps in clinical profiles make diagnostic boundaries difficult to define. Suggestive of interconnected etiologies, identified risk factors are typically non-specifically associated with a range of MDx[2,3], and many patients go through a number of diagnostic categories over their life course. Collectively, there is compelling evidence[4,5], and an emerging consensus among clinicians[6], that MDx exist on a continuum that includes mental health traits descriptive of the general population[7,8] – thus challenging the validity of their current categorical diagnostic classification. A dimensional approach directed at the identification of prodromal-, intermediate-, and transdiagnostic mental health risk signatures is therefore paramount for the implementation of stratified and precision psychiatry[9,10].

Recent years have seen significant discoveries concerning the risk landscape of MDx. Genome-wide association studies (GWASs), in particular, have provided insight into the genetic architecture of both mental health traits and psychiatric conditions[11–15]. Propelled by a shift in ascertainment strategies toward more comprehensively phenotyped cohorts[16], the pleiotropic effects and broader phenotypic correlates of genetic risk factors are starting to emerge[17–20]. It has thus been established that genetic predisposition for MDx intersects with genetic variations associated with diverse behavioral and mental health traits in the general population[21,22], including personality[4], cognition[23], impulsivity[24], sleep patterns[25], and musicality[26]. Cumulative inherited risk burden in the form of polygenic scores (PGS) rarely explains more than 5% of the phenotypic variance in the context of mental health traits. However, capitalizing on the fact that MDx exists on a genetic spectrum with phenotypically overlapping traits, enrichment with data from related traits significantly increases both predictability for individual MDx, as well as prospective prediction of recurrence and psychiatric comorbidity[27,28]. However, as clinical manifestation is attributable to both heritable-, biological-, and lifelong exposure to external factors[11], genetic risk alone will always only provide limited information about the individual's current mental health status, clinical profile, and trajectory[29]. Offering a promising solution to this constraint, peripheral markers may encapsulate transient and persistent signatures of mental health status and predict mental health outcomes[30,31]. In this context, MDx blood state biomarkers have been identified that reflect dynamic changes in disease course[31,32], and non-invasive multimodal neuroimaging has identified subtle anatomical and functional brain changes even

at the prodromal stage of MDx[33]. Critically, a combination of PGS and other markers has proven to explain more variance of studied complex traits than PGS alone[34,35].

Critically, constituting a significant limitation in current psychiatric research, most insight has been gained from the most commonly used case-control study design[36]. By excluding the large portion of the population with mixed symptoms and mental health issues without explicit diagnoses, it thus fails to represent the full spectrum of mental health experiences[37]. In contrast, transdiagnostic approaches, that take into account the broader psychometric profiles of individuals and recognize the shared mechanisms across diagnostic categories, may be better positioned to elucidate overarching processes underlying psychopathology and propel the field towards more personalized diagnostics and therapeutics[38].

Here we utilize a psychometry-based archetyping approach in a representative and deep phenotyped sample to identify mental health-associated multimodal signatures in the young Danish population.

MATERIALS AND METHODS

Study population and procedures

The data were collected under the EU COST Action CA18106. The data is a segment of a larger dataset focused on the Neural Architecture of Consciousness, originating from Aarhus University, Denmark from which other studies with different aims have been conducted[39,40]. Parts of the method descriptions have been adapted from these previous publications. Participants were recruited from the participant pool of the Center of Functionally Integrative Neuroscience at Aarhus University and through local advertisements. Inclusion criteria were: a) Anatomically normal brain (no known abnormalities, brain damage, or brain surgery); b) Age between 18 and 50 years; c) physically healthy; d) Normal or corrected-to-normal vision; e) Normal hearing. Exclusion criteria: a) MR contraindications; b) Use of neuropharmacologicals or other medications that may affect neural states; c) Bodily build that does not allow for MR scan; d) Pregnancy; e) Skin diseases. Incentives were offered to participants. A total of 351 (210 females) participants were selected for this dataset (See **Figure S1** and **Table S1** for details). The study adhered to ethical standards, having received approval from the local ethics committee, De Videnskabsetiske Komitéer for Region Midtjylland in Denmark, and complied with all relevant guidelines and regulations.

Upon enrollment, participants completed the NEO-PI-3 personality test online. Additionally, they participated in an extensive online questionnaire session covering various psychological and behavioral domains, including awareness, mindfulness, perception, cognition, emotional regulation, anhedonia, depression, sleep patterns, impulsivity, and stress (see **Table S2** for the full list). Participants also provided basic demographic information, including employment status, educational level and field of study, civil status, and personal and family history of mental health issues.

Within approximately 1-2 weeks following the completion of the online sessions, participants underwent a 1-hour MRI session, which included three sequences: diffusion-weighted imaging (DWI), multi-parameter mapping (MPM), and resting-state functional MRI (fMRI). In a separate session, typically within 1-2 weeks of the MRI, an IQ test (WAIS-IV) was administered. DNA samples were collected using SK-1S Isohelix™ buccal swabs (YouDoBio, Rødovre, Denmark) from 304 participants, and they were invited to donate fasting-state blood within two weeks of the MRI session. DNA was extracted using the Isohelix™ Buccal-Prep Plus DNA Isolation Kit (YouDoBio, Rødovre, Denmark), and the DNA, along with serum and plasma samples (n = 198), was stored at -80°C until further processing.

Psychometry-based archetyping

Psychometry-based archetypes of the study participants were created through a soft clustering approach using the R package 'archetypes'[41], with the 'robustArchetypes'[42] function. To determine the optimal *k* number of archetypes the minimized residual sum of squares (RSS) was evaluated across varying numbers of archetypes using a scree plot analysis, which suggested that five archetypes provided the best fit (**Figure S2A**). This conclusion was further supported by the Calinski-Harabasz score, which was notably higher for the five-archetype model compared to other values of *k* (**Figure S2B**). Additionally, the Calinski-Harabasz score indicated that a membership cutoff of 0.5 or 0.6 for five archetypes was optimal for distinguishing these five archetypes (**Figure S2B**). In order to maximize *n*, we selected a threshold archetype score of 0.5, which still emphasized individuals with a single dominant archetype. All individuals who did not meet this requirement were assigned to a 'mixed archetype' group. A *post hoc* Silhouette analysis, conducted with a standard distance matrix using the'factoextra'[43] package in R, confirmed that the five extreme archetype groups were well-clustered, whereas the mixed etiology group did not form a homogeneous cluster (**Figure S3**).

Genotyping, relatedness pruning, and removal of ancestry outliers

Genotyping was performed at Statens Serum Institut (SSI, Copenhagen, Denmark) using the Global Screening Array v2 with a multi-disease drop-in (Illumina, San Diego, California) according to the manufacturer's instructions. Genotype calling was performed using GenTrain V3. Quality control (QC) was conducted at the marker level, retaining markers that met the following criteria: call rate ≥ 0.98, missing difference ≤ 0.02 between cases and controls, minor allele frequency (MAF) ≥ 0.005, and Hardy-Weinberg equilibrium (HWE) P-value $\geq 1x10^{-7}$ (See https://sites.google.com/a/broadinstitute.org/ricopili/preimputation-qc for further details).

We phased the full set of merged genotype samples using EAGLE (v2.4.1)[44] and imputed the phased data with the EUR population of the 1000g-phase-3-v5 (hg19) reference panel using Minimac4 (v1.0.0)[45]. Imputed SNPs with Rsq values <0.3 were excluded from further analysis.

Principal components were calculated using LDAK, and through visual inspection of the PCA plot, we identified and excluded 21 genetic outliers from the analysis (**Figure S4**). Following this exclusion, principal components were recalculated with the remaining individuals and used as covariates in subsequent analyses.

Polygenic scores, genetic prediction, and genetic archetypes

Polygenic scores (PGSs), were constructed from the largest publicly available summary statistics at the time of publication for each trait of interest using MegaPRS[46], which have shown good performance for

psychiatric traits compared to other advanced PGS methods[47]. Summary statistics files used for PGS construction (**Table S3)** were prepared and processed using the software program process_sumstats[48]. The MegaPRS software was run using standard settings recommended by the authors (https://dougspeed.com/megaprs/) using the BLD-LDAK heritability model and BayesR-SS to construct the prediction model. Model parameters were selected using pseudo cross-validation. The 1000 genomes from the non-Finnish European subpopulation were used as a reference to estimate SNP-SNP correlations. Only SNPs with a minor allele frequency ≥1% and an imputation score ≥0.9 (when available) were included for PGS construction.

Feature selection was performed to rank genetic traits according to their contribution to the archetypes. Specifically, Random Forest Regression with Recursive Feature Elimination (RFE) was used, targeting the score for each archetype. The dataset was split into training and test sets, and 5-fold cross-validation was employed to ensure the robustness of the model. This method iteratively removes the least important features and refits the model, thereby identifying the most significant PGSs. Testing for evidence of shared etiology between base and archetypes, R^2 was used to report the total phenotypic variance explained by the PGSs. A significance cutoff of $p < 0.05$ was applied with no correction for multiple testing. To further interpret the contribution of these important features to the model, we applied the Shapley additive explanations (SHAP)[49] approach, which provided insights into each feature's influence on the individual archetype scores.

Archetypes based on PGS derived from GWAS summary statistics listed in **Table S3**, were constructed as described for psychometry-based archetypes above, and the same approach was used for identifying the optimal k number of archetypes and threshold for membership. Scree plot analysis and the Calinski-Harabasz score both pointed to four archetypes as the optimal fit (**Figure S5A-B**), with the Calinski-Harabasz score indicating that a membership cutoff of 0.5 was optimal (**Figure S5B**). A post hoc Silhouette analysis showed that the archetype groups look well clustered and that the individuals in the mixed etiology group do not form a homogeneous cluster (**Figure S6**).

Magnetic resonance imaging (MRI)

Imaging was performed on a Siemens Magnetom Prisma-fit 3T MRI scanner. The procedure commenced with preliminary scouting scans. This was followed by two sequences of resting-state fMRI, lasting 12 and 6 minutes respectively. The session also included quantitative multi-parameter mapping (approximately 20 minutes) to facilitate the synthetic generation of T1-weighted images. Additionally, high-angular resolution diffusion imaging (HARDI) was conducted over a period of around 10 minutes, all within a single session lasting about one hour. For every participant, we acquired 1500 functional volumes, with a repetition time (TR) of 700 ms and an echo time (TE) of 30 ms. The parameters set were: a voxel size of 2.5 mm^3, a field

of view (FOV) of 200 mm, and a flip angle of 53°. The HARDI sequence incorporated multiple diffusion directions: 75 at b = 2500 s/mm^2, 60 at b = 1500 s/mm^2, 21 at b = 1200 s/mm^2, 30 at b = 1000 s/mm^2, 15 at b = 700 s/mm^2, and 10 at b = 5 s/mm^2. These different b-shells were acquired in a single series with a flip angle of 90°, a TR/TE of 2850/71 ms, a voxel size of 2 mm^3, a matrix size of 100 x 100, and 84 slices in total. The primary phase-encoding direction was from anterior to posterior (AP), with an additional acquisition in the opposite phase-encoding direction (PA) at b = xx s/mm^2 for EPI distortion correction.

To create synthetic T1-weighted images, high-resolution longitudinal relaxation rate (R1) and effective proton density (PD) maps were utilized, and obtained through the MPM sequence protocol[50,51]. Initially, these maps underwent thresholding to align with FreeSurfer's required units. The R1 map was transformed into a T1 map by inverting its values and applying a zero threshold, followed by a multiplication by 1000. Similarly, the PD map was zero-thresholded and scaled up by a factor of 100. These adjustments were carried out using FSL maths commands. The FreeSurfer's "mri_synthesize" command was then employed to generate a synthetic FLASH image, using the modified T1 (derived from the adjusted R1 map) and PD maps. Optional arguments were used to enhance the contrast between gray and white matter, with parameters set at 20, 30, and 2.5. In the final step, the synthetic T1-weighted image was reduced by a quarter to meet FreeSurfer's expected scale (see Keller et al.[52] for further details).

Metabolomics data generation and analysis

Serum samples (100 µL each) were randomly distributed across three 96-well plates (batches). Before sample preparation, a batch of serum was set aside as external control (EC) samples and stored at −80 °C. These EC samples, along with plate-specific pools of all serum samples within each batch, were analyzed to ensure quality control. Sample preparation involved extraction in 80% methanol, with samples incubated for 45 minutes and then centrifuged. The resulting supernatant was evaporated under nitrogen and reconstituted in 95% solvent A (99.8% water, 0.2% formic acid) and 5% solvent B (49.9% methanol, 49.9% acetonitrile, 0.2% formic acid). Mass spectrometry analysis was performed using a timsTOF Pro mass spectrometer coupled to a UHPLC Elute LC system, Bruker Daltonics (Billerica, MA, US). The analytical separation was performed on an Acquity HSS T3 (100 Å, 2.1 mm x 100 mm, 1.8 µm) column (Waters, Milford, MA, US). The analysis started with 99% solvent A for 1.5 min, thereafter a linear gradient to 95% solvent B for 8.5 min followed by an isocratic condition at 95% mobile phase B for 2.5 min before going back to 99% mobile phase A and equilibration for 2.4 min. Metabolomics preprocessing was done using the Ion Identity Network workflow in MZmine[53,54](version 3.3.9). Before statistical analysis, metabolite features present in less than 25% of the samples were removed and features present in fewer than 75% were

treated as binary variables (present or absent). This resulted in a final dataset with a total of 1076 metabolite features measured, among which 433 features were continuous and 643 were binary variables. Missing values for metabolite features with continuous measurements were further subjected to imputation using missForest[553] and subsequent batch correction was performed by centering and univariance scaling of each metabolite per batch. Annotation of metabolite features was performed using mass spectral molecular networking through the GNPS Platform, *in silico* annotation through Network Annotation Propagation, Sirius+CSI:FingerID, and deep neural networks in CANOPUS. Detailed descriptions of sample preparation, mass spectrometry analysis, preprocessing, annotation, and quality control procedures can be found in the **Supplementary Material** section.

For the analysis, annotated features of the dataset was divided into predictor variables (metabolites) and outcome variables (archetypes). The goal was to identify the most significant predictors for each archetype and assess their relevance using a two-step process: feature selection via Ridge Regression and statistical inference through Ordinary Least Squares (OLS) regression. A variance threshold filter was applied to retain only informative features, with features showing variance below 0.1 excluded as they were unlikely to contribute significantly to the predictive models. Subsequently, a correlation matrix was generated to assess multicollinearity among the remaining features. Features with an absolute Pearson's correlation coefficient above 0.9 were flagged as highly correlated. To reduce redundancy, only one feature from each correlated pair was retained, guided by the upper triangle of the correlation matrix. Ridge Regression was independently applied to each archetype (A1, A2, A3, A4, A5), chosen for its regularization properties that mitigate multicollinearity and reduce overfitting by shrinking the coefficients of less significant features. The top 20 features with the highest absolute coefficients from Ridge Regression were selected for further analysis. These top 20 features were then subjected to OLS regression, which provided unbiased estimates of the relationships between features and outcomes, including coefficient estimates, p-values, and confidence intervals for formal hypothesis testing. The null hypothesis for each feature was that its regression coefficient was zero, indicating no relationship with the outcome. A p-value threshold of 0.05 was used to determine statistical significance, with features below this threshold considered significantly related to the outcome.

Statistical analysis

Relevant statistical analyses and plotting were performed in the R Statistical Computing environment v4.3.1 (https://www.r-project.org/).

RESULTS

Deep psychometric profiling reveals five distinct archetypes

Psychometric data were collected from 351 research volunteers (~60% female) sampled from the young Danish population (mean age ~ 24.4 years; **Figure S1**). The cohort was predominantly composed of university and college students (~80%), with smaller proportions being full-time employed (7%) or unemployed (8.3%). Approximately 38% of participants reported being single (**Table S1)**. Noteworthy, ~13% of participants had a history of MDx, including ~6% reporting multiple MDx diagnoses, with depression (MDD), anxiety and phobia, attention deficit hyperactivity disorder (ADHD), and stress-related disorders being the most prevalent diagnostic entities (**Figure S7**) - thus roughly reflecting the Danish population average for this age group[56].

Recognizing the continuous nature of mental health traits, we employed a soft-clustering method, archetype analysis[57], on 258 individuals with complete psychometric data to stratify the sample based on their broadly characterized cognitive, emotional, and behavioral patterns (**Table S2**). By identifying extreme psychometric profiles on individual scales and positioning individuals within the phenotypic spectrum as convex combinations of these extremes, our analysis revealed five stable psychometric archetypes (A1-5) (**Figure 1A** and **Figure S2-3**). Each participant was thus assigned archetypal memberships based on their quantitative archetype scores. Most individuals (55%) were mainly affiliated with one dominating archetype (membership > 0.5), whereas 45% were located in the middle of the phenotype distribution with moderate contributions from two or more archetypes (mixed group) (**Figure 1B**). Although a gender bias was seen for the A2-4 archetypes, differences in gender and age did not appear to define the archetypes (**Figure S8**). By focusing on individuals at the extreme of the archetypal distribution (>0.5 score for any archetype), we delineated the defining psychometric features of each archetype. Differences in facets of personality were unsurprisingly a significant contributing factor (**Table S4**). Specifically, A1 (22 members) was characterized by high neuroticism and low extraversion and conscientiousness scores; A2 (24 members) by high openness score; A3 (26 members) by low agreeableness scores; A4 (39 members) by low openness scores; and A5 (30 members), being the antithesis of A1, with low neuroticism scores and high extraversion and conscientiousness scores (**Figure 1C** and **Table S4**). Variation in other psychometrics that were to some degree correlated with personality measures (**Figure S9** - e.g. perception, cognition, impulsivity, perceived stress, sleep patterns, mindfulness, emotional regulation, hedonic tone, and interoceptive awareness) further shaped the archetypes (**Figure 1C** and **Table S4**). In alignment with the marked differences in neuroticism score between archetypes, individuals in the A1 archetype scored significantly higher than the rest of the sample in the commonly applied depression screening tools[58], the CES-D depression scale (**Figure 1C-D** ; Mann-Whitney U test: $p_{A1} = 0.012$). This trend was also evident

across several other psychometric scales related to emotional regulation, impulse control, and cognitive function, which are often associated with mental health risks[59–61] (**Figure 1C**). Supporting the robustness of the association between archetypes and contributing psychometric measures, quantitative archetype scores from the entire cohort, and not only individuals with extreme scores, revealed similar associations (**Figure S10**). Consequently, in subsequent analyses, we used both archetype membership and quantitative archetype scores to enhance the statistical power for discovery.

Psychometric archetypes are associated with differences in the prevalence of mental health diagnoses

In order to evaluate the clinical relevance of the archetypes, we assessed the distribution of participants' self-reported diagnoses across the archetypes and their correlation with archetype scores. While having a dominant archetype did not significantly increase the likelihood of having a diagnosis compared to the mixed archetype group (**Tables S5**, Kruskal–Wallis test: $p = 0.052$), individual archetypes were associated with, respectively, increased (risk archetypes) or decreased (resilient archetypes) prevalence of MDx diagnoses (**Figure 1B**). In line with its characteristic high neuroticism and depression scores, the prevalence of MDx diagnoses was significantly higher in the A1 archetype compared to the average among participants, with nearly one-third reporting one or more MDx diagnoses (**Figure 1B** and **Table 1;** relative risk (RR) = 2.4, 95% CI 1.2-4.8, $p_{A1} = 0.012$). Notably, reported diagnoses were not restricted to specific diagnostic categories but broadly represented the most common neurodevelopmental, anxiety, mood, personality, stress, eating, and substance-related disorders (**Table 1**). In contrast, A5 was protective against MDx, with only one self-reported case in this group (**Figure 1B** and **Table 1**; RR = 0.25, 95% CI 0.04-1.78, $p_{A5} = 0.17$).

We then analyzed the correlation between individual archetype scores and the prevalence of MDx diagnoses, encompassing both single and comorbid conditions, as well as family history of MDx. The A1 score exhibited a strong positive correlation with both diagnosed MDx (**Figure 1E**; Pearson's correlation coefficient: $p_{A1} = 0.0085$) and psychiatric comorbidity (**Figure 1E**; Pearson's correlation coefficient: $p_{A1} = 0.0096$), thus highlighting the A1 score as a significant marker of mental health risk. The significant correlation between A1 scores and MDx diagnoses among first-degree relatives (**Figure 1E**; Pearson coefficient test, $p = 0.0118$) further suggests that individuals with a strong affiliation to the A1 archetype have an inherited predisposition to mental health problems. Whereas on average one quarter of individuals in the cohort reported having first-degree relatives with diagnosed MDx, the distribution was not even across archetypes (**Table 1**; Kruskal-Wallis test: $p = 0.024$), with the highest proportion (45%) seen in the A1 archetype.

While an elevated familial risk was observed for the A2 archetype (**Figure 1E** and **Table 1**; Pearson's correlation coefficient: $p_{A2} = 0.0128$), A2 archetype affiliation was not significantly associated with an increased risk of MDx diagnosis, nor did A2 scores show a significant correlation with MDx diagnoses (**Figure 1E** and **Table 1**). At the opposite end of the archetype spectrum, the A3 score was significantly negatively correlated with familial risk of MDx (**Figure 1E**; Pearson's correlation coefficient: $p_{A3} = 0.0458$), and a similar trend was seen for the A4 and A5 scores (**Figure 1E**), which both displayed significant negative correlations with psychiatric comorbidity (Pearson's correlation coefficient: $p_{A4} = 0.0225$ and $p_{A5} = 0.0176$). The A5 scores was further significantly negatively correlated with single MDx diagnoses (**Figure 1E**; Pearson's correlation coefficient: $p_{A5} = 0.0031$).

Heritable factors contribute significantly to the formation of psychometric archetypes

To assess whether archetypes represent distinct, biologically grounded entities influenced by inherited genetic variation, we aimed to quantify the genetic contribution to the archetypes. Given the limitations of our sample size, we opted for an alternative to direct heritability analysis by evaluating whether the variance in archetype scores could be explained by polygenic scores (PGS) derived from a wide range of mental health-related traits, including behavioral, cognitive, and neuroimaging traits (**Table S3**). Ranking of the PGSs according to their contribution to the archetypes revealed that different archetypes are driven by distinct sets of genetic influences, with few top-ranked PGS features shared between archetypes – and none shared between the A1 and A5 archetypes (**Table S6-10**). For the A1 archetype, top-ranked PGSs mostly related to neuroimaging traits (**Table S6)**, whereas the A5 archetype was influenced by PGSs relating to both neuroimaging, cognitive, behavioral and clinical traits. The latter included broad genetic risk disposition to psychopathology, measured in a cross disorder GWAS, as well as ADHD, neuroticism, extraversion (**Table S10**). For the A2-4 archetypes, top-ranked PGSs mainly included neuroimaging and cognitive traits, but also PGSs relating to sleep patterns, impulsivity, musicality (A2), risk tolerance (A4) and substance use (A2 and A3) (**Table S6-10**).

To interpret the contribution of these genetic factors to individual archetype, we assessed their influence on archetype scores using the Shapley additive explanations (SHAP) approach. This highlighted distinct patterns in how the same genetic risks contributed to different archetypes. For instance, genetic predisposition to the impulsive personality trait, perseverance, contributed positively to the A4 score (**Figure S11C**), but negatively to the A3 score (**Figure S11B**). Similarly, PGS for various traits relating to attention, displayed opposite direction of effect on the A1, A2 and A4 scores (**Figure 2A** and **Figure S10A** and **C**). The analysis further showed that PGSs derived from GWASs on major psychiatric disorders contributed with positive effect to the A1 score (autism spectrum disorder (ASD); **Figure 2A**), but with

negative effect on A5 score (Cross disorder and ADHD; **Figure 2B**) and A2 score (ADHD – adult subtype; **Figure S10A**). An exception to this trend, was PGSs relating to cannabis use disorder, which contributed with negative effect to both the A1 and A3 scores (**Figure 2A** and **Figure S10B**). For the neuroimaging traits, the direction of effect did in some instances differentiate between archetypes, i.e. PGS for cortical average thickness in the caudate anterior cingulate region contributed with a strong negative effect to A1 score (**Figure 2A**), but with a similar strong negative effect to the A2 score (**Figure S10A**). Generally, genetic predisposition to several brain imaging traits showed a strong impact on the model output for the A1 score. This included a negative impact of PGS relating to regional cortical surface area (e.g. entorhinal cortex) and accumbens volume, and positive impact for PGS for cortical thickness (e.g. insula) and surface area (pars orbitalis) (**Figure 2A**).

To determine how much of the variance in archetype categories and archetype scores could be explained by genetic contributions in the form of PGSs, we attempted archetype prediction to establish the coefficient of determination (R^2) for each trait. Notably, up to 9% of the variance in archetype scores could be explained by the PGS model based on life satisfaction, positive affect, neuroticism, and depressive symptoms, collectively captured in a well-being[62] PGS (**Table S11**). In contrast, when we categorized individuals into binary archetype groups, the strongest correlations emerged with brain imaging traits, particularly those related to cortical thickness, where the PGS explained up to 8% of the variance (**Table S11**). Discerning the specific genetic influences that shape the individual archetypes, we again found that PGSs for several brain imaging traits, particularly changes in cortical surface area, contributed significantly to the A1 archetype (**Figure 2C**). PGSs for well-being and the clinical phenotypes cross-disorder, anxiety, and ASD each additionally accounted for a significant portion of the variance in the A1 score (**Figure 2D**). For the A2-4 archetypes, PGSs associated with cognitive traits all explained a significant part of the variance (**Figure 2D**), supplemented with PGSs for personality traits like Openness (A2) and Extraversion (A3), risk tolerance and clinical traits such as BP and panic disorder (A2). PGSs related to EEG brain activity measures, further significantly influenced both the A2 and A4 scores (**Figure 2C-D**). For the resilient A5 archetype, PGSs showed significant correlations with a range of clinical traits, including suicide, ASD, cross-disorder, and ADHD (**Figure 2C-D**).

To determine whether the collective genetic predisposition to the included MDx-related traits forms a broader genetic signature associated with psychometry-based archetypes, we captured the variance across the complete set of PGSs in genetic-based archetypes. From this approach, four PGS archetypes emerged ($A1_{PGS}$-$A4_{PGS}$ - **Figure 3A** and **Figure S5-6**), mainly defined by differences in genetic predisposition for various neuroimaging-, cognitive-, and behavioral traits (**Figure 3B** and **Figure S12**). Notably, the defining neuroimaging-linked PGSs revealed contrasting genetic characteristics between the A1/A3 and the A2/A4

PGS-based archetypes in terms of genetic predisposition to volumes of subcortical structures (e.g., thalamus, accumbens, and amygdala). Particularly for the $A4_{PGS}$ archetype, PGS for suicide (**Figure 3B**), as well as clinical MDx traits such as cross disorder, anxiety, ADHD, alcohol use disorder, MDD, stress disorders, schizophrenia, borderline personality disorder and OCD all explained part of the variance in the archetype score (**Figure S12**). Accordingly, the $A4_{PGS}$ archetype displayed a significant overrepresentation of individuals with a self-reported MDx diagnosis (**Figure 3A** and **Table S12**; Mann-Whitney U test: $p_{A4PGS} = 0.00055$), and the $A4_{PGS}$ archetype score was significantly positively correlated with CES-D score and neuroticism – which, despite its markedly different genetic characteristics, was also the case for the $A2_{PGS}$ score (**Figure 3C**). In line with this finding, the overlap between the psychometry-based risk A1 archetype and the two PGS-based archetypes ($A2_{PGS}$ and $A4_{PGS}$) was significant (Fisher's Exact test: $p = 0.0044$), with approximately 80% of individuals affiliated to the psychometry-based a1 risk archetype belonging to these genetic archetypes. Conversely, more than 80% of individuals in the resilient A5 psychometric archetype were members of the $A1_{PGS}$ and $A3_{PGS}$ archetypes (**Figure 2D**).

Neuroarchitectural features define individual archetypes

Building on the suggested link between genetic predispositions for specific neuroimaging traits and psychometry-based archetypes, we leveraged the extensive imaging data from our samples to determine whether these genetic predispositions manifest as distinct neuroarchitectural differences across the identified archetypes. Focusing on brain volumetric and structural measures emphasized by the genetic data, we assessed the correlation between these MRI measures and PGSs derived from GWASs targeting the same neuroimaging traits. Except for measures of cortical thickness, all neuroimaging measures showed a positive correlation with the calculated PGSs (**Figure S13**), with statistically significant correlations observed in the context of whole brain volume (Pearson's correlation coefficient: $p = 0.00054$), cortical surface area in the cingulate isthmus (Pearson's correlation coefficient: $p = 0.000037$), and nucleus accumbens (Pearson's correlation coefficient: $p = 0.0075$) (**Figure 4A-C** and **Figure S13A-I**). Subsequently, we examined the association between neuroimaging measures and archetype scores. While no measure was discriminatory across archetypes (**Table S13**), in support of their biological underpinnings, some individual archetype scores were significantly associated with various neuroimaging traits (**Figure 4D** and **Table S14**). In particular, thinning of the insula and frontal pole regions of the cortex was significantly associated with the A3 scores (**Figure 4 D**, $p = 0.0013$ and $p = 0.0112$, respectively), whereas cortical surface area in the cingulate isthmus was significantly associated with A4 scores (**Figure 4 D**, $p = 0.0255$). Whereas not significant (**Figure S14**), in accordance with the negative correlation between PGS for enthorinal surface area and A1 scores, entorhinal cortical surface area was negatively correlated with

the A1 score (**Figure 4D**). Similarly matching the genetic prediction, cortical surface area in the pars orbitalis was smallest in members of the resilient A5 archetype and subcortical volumes generally smaller in the A1 archetype compared to, particularly, the A5 archetype (**Figure 4D**).

Archetypes are associated with distinct circulating metabolite signatures

To identify potential circulatory markers that could differentiate between the various archetypes, we performed an analysis of fasting-state blood metabolites across our sample. After thorough quality control (QC) procedures, we obtained quantitative data for a total of 433 compounds, with annotations available for 121 of these metabolites. Despite the relatively small proportion of samples with available blood profiles, we identified 11 annotated metabolites that were associated with one or more of the quantitative archetype scores (**Figure 5**). We found the highest number of discriminative metabolites for the A1 archetype score, which was significantly associated with high levels of 9-decenoylcarnitine, Betaine, Indole-3-lactic acid, and a phosphocholine metabolite (PC(O-18:0/2:0)) (**Figure 5** and **Table S14**). Notably, whereas not significant, all of these compounds were negatively correlated with the A5 archetype score. Similarly, the another phosphocholine metabolite (PC(0:0/18:1)) and Kynurenine were both significantly negatively correlated with the A5 scores, but positively correlated with the A1 score – thus collectively forming a blood signature that differentiates between the risk and protective archetypes (**Figure 5**).

DISCUSSION

Mental health challenges are diverse and widespread, influenced by cultural, socioeconomic, and environmental factors[63]. They affect our thoughts, feelings, and behaviors and often require intervention. A major issue in mental health care is the high rate of misdiagnosis, relabeling, and underdiagnosis. This often stems from categorical diagnostic practices that emphasize observable symptoms over underlying causes. Such an approach can be inadequate because it overlooks the clinical variability, overlapping symptoms, and evolving nature of mental health conditions. Moreover, the persistent stigma surrounding mental illness discourages many from seeking help, contributing to the underdiagnosis and inadequate treatment of mental health issues. Consequently, many individuals continue to suffer in silence, with their conditions remaining unnoticed, misunderstood, and untreated—further exacerbating their symptoms and hindering recovery.

To address these challenges, it is essential to affirm the biological basis of mental illness and leverage this understanding to develop more comprehensive and personalized approaches to mental health care. This study introduces a psychometry-based archetyping approach to stratify the young population according to cognitive, emotional, and behavioral profiles related to mental health. Unlike traditional psychiatric research, which often focuses on extreme cases using case-control study designs[64], our method aims to capture a broader and more representative spectrum of mental health issues. Although mental health conditions affect all age groups, they are particularly prevalent among young adults (18-30 years old), with up to 25% experiencing a mental health disorder annually[65]. By targeting a demographic disproportionately affected by mental health disorders and sampled near the mean age of onset for the most common conditions, our approach seeks to pair these psychometric archetypes with biological markers - offering a promising framework for biology-informed risk-stratification.

Psychometric Archetypes and Mental Health Risk

Our identification of five distinct psychometric archetypes underscores the continuous and multifaceted nature of mental health traits. These archetypes capture a broad spectrum of cognitive, emotional, and behavioral patterns, each differentially associated with mental health outcomes. The A1 archetype, characterized by high neuroticism and low extraversion, emerged as a significant risk archetype, with an increased prevalence of mental health diagnoses and psychiatric comorbidities. This finding is consistent with existing literature that links neuroticism to mental health risks, particularly anxiety and mood disorders[66]. However, the implications of these archetypes extend beyond personality traits alone, as evident by the contribution of psychometric measures with only limited correlation to neuroticism score, i.e. anhedonia (SHAPS), impulsivity (BIS), and emotional processing (TAS). Additionally, cognitive styles

associated with the A1 archetype, such as negative thinking patterns (CES-D) and heightened emotional reactivity (PSS), may further contribute to the development and persistence of mental health issues. Thus, the A1 archetype represents a convergence of psychometric factors that collectively increase the risk for MDx. Conversely, the A5 archetype, defined by low neuroticism and high extraversion, appeared to be protective against MDx, with a notably low prevalence of self-reported diagnoses. This protective effect may be attributed not only to personality traits but also to enhanced awareness (MAIA), mental imagery (VVIQ), creativity (MSI), sleep patterns (ESS), and positive coping strategies (MMQ, CFQ, and DERS) providing a generally more capable and optimistic outlook on life. The delineation of these psychometric archetypes provides valuable insights into the diverse ways in which cognitive, emotional, and behavioral patterns interact to influence mental health outcomes.

Archetypes have a biological foundation

By leveraging PGSs derived from a broad spectrum of mental health traits, we demonstrate that these genetic profiles account for a substantial proportion of the variance in archetype scores. This finding not only supports the biological foundation of these psychometric profiles but also highlights their etiological connections with clinical traits. Particularly, our analysis revealed that PGSs related to mental health traits, such as life satisfaction, neuroticism, depressive symptoms, and clinical conditions like ASD, bipolar disorder, and cross-disorder, significantly contribute to shaping psychometric archetypes. Notably, up to 9% of the variance in archetype scores could be explained by PGSs for well-being traits, illustrating a robust genetic influence from these psychometric dimensions. Additionally, PGSs for brain imaging traits, such as cortical thickness, also contributed significantly, with up to 8% of variance explained. Highlighting its etiological relatedness to clinical mental health conditions, the risk associated A1 score was significantly associated with PGSs for neurodevelopmental disorders like ASD and general psychopathology captured through cross-disorder PGS scores. Notably, while ASD and suicide PGS both explained a significant proportion of the variance in the A1 and A5 scores, individually, their direction of effect was opposite in the two archetypes.

An intriguing aspect of our study is the identification of composite genetic profiles (PGS-based archetypes that integrate genetic risk across hundreds of mental health and neuroimaging traits) that display a significant overlap with psychometry-based archetypes. This overlap underscores the genetic consistency of these archetypes and supports their biological relevance. For instance, the A1 psychometric risk archetype showed a notable overlap with the risk-associated PGS-based archetypes, specifically $A4_{PGS}$ and $A2_{PGS}$. The $A4_{PGS}$ and $A2_{PGS}$ archetypes, however, exhibit distinct genetic characteristics. The $A4_{PGS}$ archetype is characterized by low PGSs for sleep measures, educational attainment, and suicide, indicating

a genetic profile divergent from the $A2_{PGS}$, which is associated with high PGSs for these same variables. This distinction suggests the presence of two etiologically distinct sub-groups within the A1 risk archetype.

Despite their differences, A2PGS and A4PGS share significant correlations with subcortical MRI-estimated volumes. Both PGS-based archetypes are also linked to higher depression scores and associated psychometric measures, though they diverge in aspects such as impulsivity, awareness, and personality traits. This overlap in neuroimaging traits but divergence in other psychometric characteristics point to nuanced genetic influences that differentiate these sub-groups within the broader A1 risk archetype.

By leveraging PGSs derived from genetic studies on neuroimaging traits, we show that the genetic predispositions for specific neuroimaging traits are reflected in distinct neuroarchitectural features associated with each archetype. In line with the strong negative impact of PGS derived from GWAS on cingulate isthmus cortical surface area on the A4 score, MRI-derived measure of this area was significantly negatively correlated with the A4 score. Similarly, the reduced entorhinal surface area liked to the A1 score, was predicted by the PGS for this neuroimaging measure. This is consistent with clinical MRI studies, which have found volume of entorhinal cortex to be reduced in patients with depression[67] and schizophrenia[68] – although studies have been conflicting[69]. However, while cortical thickness in the frontal pole and insula area was significantly negatively correlated with the A3 score, neither of these were predicted by the PGS for these measures. Albeit not statistically significant, the smaller whole brain volume observed in the A1 risk archetype is consistent with previous findings linking reduced brain volume to higher mental health risks[70]. These neuroarchitectural differences not only reinforce the biological validity of the identified archetypes but also highlight the potential for integrating neuroimaging data into personalized mental health assessments.

Circulating Metabolite Signatures and Mental Health

Our study identified distinct circulating metabolite signatures associated with specific archetypes, offering potential biomarkers to differentiate between varying mental health profiles. The A1 archetype, linked to higher neuroticism and increased mental health risks, showed significant correlations with several metabolites, including Betaine, 9-decenoylcarnitine, Indole-3-lactic acid, and the phopshocholine metabolite, O-18:0/2:0, among which, several are neuroactive with potential links to mental health and biological processes underlying psychopathology. Betaine, for instance, is a known osmolyte involved in methylation processes crucial for homocysteine metabolism and it has a documented role as neuromodulator in the nematode nervous system[71]. Proper methylation is essential for maintaining mental health, and disruptions in this pathway have been linked to various psychiatric disorders, including

depression and anxiety in preclinical models[72–74]. Additionally, and imbalanced homocysteine metabolism has been linked to both depression and schizophrenia[75,76]. 9-Decenoylcarnitine, an acylcarnitine involved in mitochondrial energy production, plays a role in neuroprotection and cholinergic neurotransmission[77]. Altered levels of 9-decenoylcarnitine have been observed in schizophrenia, underlining its potential as a biomarker for psychopathology[78]. Indole-3-lactic acid, a tryptophan metabolite produced by gut microbiota, has been reported to play a role neuronal developmental processes[79], and alterations in indole levels have been linked to severity of depression and anxiety scores in patients with clinical depression[80]. Interestingly, another compound acting in the tryptophan pathway, Kynurenine, similarly known for its role in neuroinflammation and neurotoxicity, was negatively correlated with the A5 score. Current evidence suggests that a functional imbalance in the synthesis of Tryptophan metabolites causes the appearance of pathophysiologic mechanisms that leads to various neuropsychiatric diseases[81]. Finally, opposing levels of phosphocholine metabolites (O-18:0/2:0 and 0:0/18:1) in the A1 and A5 archetypes suggest that these archetypes are defined by alterations in phospholipid metabolism. Such alterations are linked to neuroinflammation, cell membrane integrity and signaling in the brain, and accumulating evidence suggests a broad implications of phospholipids in the etiopathologies of MDx[82].

Perspectives

This study leverages a psychometry-based archetyping approach in a deeply phenotyped young Danish population to explore the complex interplay between genetic predispositions, neuroarchitectural features, and circulating metabolite profiles associated with mental health. By identifying distinct psychometric archetypes and linking them to heritable factors, neuroimaging traits, and blood metabolites, we provide a comprehensive characterization of mental health variation within the general population and highlight the potential of combining layers of data to advance precision psychiatry. While this study offers valuable insights, it is important to acknowledge its limitations. The sample size, although deeply phenotyped, may limit the generalizability of our findings. Future studies with larger, more diverse populations are needed to validate our results and explore the potential for applying this archetyping approach in clinical settings. Additionally, the cross-sectional nature of the data limits our ability to draw conclusions about causality, and longitudinal studies are necessary to understand the dynamic interactions between genetic predispositions, neuroarchitectural changes, and mental health trajectories.

References:

1. Wray, N. R. *et al.* Research Review: Polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry* **55**, 1068–1087 (2014).

2. Lee, S. H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–94 (2013).

3. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet (London, England)* **381**, 1371–1379 (2013).

4. Lo, M.-T. *et al.* Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat. Genet.* **49**, 152–156 (2017).

5. Selzam, S., Coleman, J. R. I., Caspi, A., Moffitt, T. E. & Plomin, R. A polygenic p factor for major psychiatric disorders. *Transl. Psychiatry* **8**, 205 (2018).

6. Watson, D. & Naragon-Gainey, K. Personality, Emotions, and the Emotional Disorders. *Clin. Psychol. Sci.* **2**, 422–442 (2014).

7. A Widiger, T. Personality and psychopathology. *World Psychiatry* **10**, 103–6 (2011).

8. Rosenström, T. *et al.* Joint factorial structure of psychopathology and personality. *Psychol. Med.* 1–10 (2018) doi:10.1017/S0033291718002982.

9. Arns, M., van Dijk, H., Luykx, J. J., van Wingen, G. & Olbrich, S. Stratified psychiatry: Tomorrow's precision psychiatry? *Eur. Neuropsychopharmacol.* **55**, 14–19 (2022).

10. Caspi, A. & Moffitt, T. E. All for One and One for All: Mental Disorders in One Dimension. *Am. J. Psychiatry* **175**, 831–844 (2018).

11. Uher, R. & Zwicker, A. Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness. *World Psychiatry* **16**, 121–129 (2017).

12. Assary, E., Vincent, J. P., Keers, R. & Pluess, M. Gene-environment interaction and psychiatric disorders: Review and future directions. *Semin. Cell Dev. Biol.* **77**, 133–143 (2018).

13. Uher, R. Gene-environment interactions in severe mental illness. *Front. Psychiatry* **5**, 48 (2014).

14. Zacharias, H. U. *et al.* A metabolome-wide association study in the general population reveals decreased levels of serum laurylcarnitine in people with depression. *Mol. Psychiatry* **26**, 7372–7383 (2021).

15. Kurochkin, I. *et al.* Metabolome signature of autism in the human prefrontal cortex. *Commun. Biol.* **2**, 234 (2019).

16. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med.* **12**, e1001779 (2015).

17. Wright, J. T. & Herzberg, M. C. Science for the Next Century: Deep Phenotyping. *J. Dent. Res.* **100**, 785–789 (2021).

18. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).

19. Sewell, M. D. E. *et al.* Associations between major psychiatric disorder polygenic risk scores and blood-based markers in UK biobank. *Brain. Behav. Immun.* **97**, 32–41 (2021).

20. Shen, X. *et al.* A phenome-wide association and Mendelian Randomisation study of polygenic risk for depression in UK Biobank. *Nat. Commun.* **11**, 2301 (2020).

21. Hindley, G. *et al.* Charting the Landscape of Genetic Overlap Between Mental Disorders and Related Traits Beyond Genetic Correlation. *Am. J. Psychiatry* **179**, 833–843 (2022).

22. Abdellaoui, A. & Verweij, K. J. H. Dissecting polygenic signals from genome-wide association studies on human behaviour. *Nat. Hum. Behav.* **5**, 686–694 (2021).

23. Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).

24. Wang, Y., Sibaii, F., Lee, K., J. Gill, M. & L. Hatch, J. The pleiotropic architecture of human impulsivity across biological scales. *medRxiv* **1**, 1–13 (2021).

25. Austin-Zimmerman, I. *et al.* Genome-wide association studies and cross-population meta-analyses investigating short and long sleep duration. *Nat. Commun.* **14**, 6059 (2023).

26. Wesseldijk, L. W., Lu, Y., Karlsson, R., Ullén, F. & Mosing, M. A. A comprehensive investigation into the genetic relationship between music engagement and mental health. *Transl. Psychiatry* **13**, 15 (2023).

27. Karlsson Linnér, R. *et al.* Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nat. Neurosci.* **24**, 1367–1376 (2021).

28. Als, T. D. *et al.* Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. *Nat. Med.* **29**, 1832–1844 (2023).

29. Dubovsky, S. L. The Limitations of Genetic Testing in Psychiatry. *Psychother. Psychosom.* **85**,

129–135 (2016).

30. Soares, S., Rocha, V., Kelly-Irving, M., Stringhini, S. & Fraga, S. Adverse Childhood Events and Health Biomarkers: A Systematic Review. *Front. Public Heal.* **9**, (2021).

31. García-Gutiérrez, M. S. *et al.* Biomarkers in Psychiatry: Concept, Definition, Types and Relevance to the Clinical Reality. *Front. Psychiatry* **11**, (2020).

32. Lenski, M. *et al.* Metabolomic alteration induced by psychotropic drugs: Short-term metabolite profile as a predictor of weight gain evolution. *Clin. Transl. Sci.* **14**, 2544–2555 (2021).

33. Arbabshirani, M. R., Plis, S., Sui, J. & Calhoun, V. D. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage* **145**, 137–165 (2017).

34. Mccartney, D. L. *et al.* Epigenetic prediction of complex traits and death. doi:10.1186/s13059-018-1514-1.

35. Odintsova, V. V. *et al.* Predicting Complex Traits and Exposures From Polygenic Scores and Blood and Buccal DNA Methylation Profiles. *Front. Psychiatry* **12**, (2021).

36. Rutjes, A. W., Reitsma, J. B., Vandenbroucke, J. P., Glas, A. S. & Bossuyt, P. M. Case–Control and Two-Gate Designs in Diagnostic Accuracy Studies. *Clin. Chem.* **51**, 1335–1341 (2005).

37. Andreassen, O. A., Hindley, G. F. L. L., Frei, O. & Smeland, O. B. New insights from the last decade of research in psychiatric genetics: discoveries, challenges and clinical implications. *World Psychiatry* **22**, 4–24 (2023).

38. Dalgleish, T., Black, M., Johnston, D. & Bevan, A. Transdiagnostic approaches to mental health problems: Current status and future directions. *J. Consult. Clin. Psychol.* **88**, 179–195 (2020).

39. Kvamme, T. L. *et al.* Vividness of Visual Imagery Supported by Intrinsic Structural-Functional Brain Network Dynamics. *bioRxiv* 2024.03.02.582470 (2024) doi:10.1101/2024.03.02.582470.

40. Tchemerinsky Konieczny, D., Wieck Fjaeldstad, A. & Sandberg, K. Test-retest reliability and validity of the Importance of Olfaction Questionnaire in Denmark. *PLoS One* **19**, e0269211 (2024).

41. Eugster, M. J. A. & Leisch, F. From Spider-Man to Hero - Archetypal Analysis in R. *J. Stat. Softw.* **30**, (2009).

42. Eugster, M. J. A. & Leisch, F. Weighted and robust archetypal analysis. *Comput. Stat. Data Anal.* **55**, 1215–1225 (2011).

43. Kassambara, A. & Mundt, F. Extract and Visualize the Results of Multivariate Data Analyses [R

package factoextra version 1.0.7]. in (2020).

44. Loh, P.-R., Palamara, P. F. & Price, A. L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat. Genet.* **48**, 811–816 (2016).

45. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).

46. Zhang, Q., Privé, F., Vilhjálmsson, B. & Speed, D. Improved genetic prediction of complex traits from individual-level data or summary statistics. *Nat. Commun.* **12**, 4192 (2021).

47. Ni, G. *et al.* A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts. *Biol. Psychiatry* **90**, 611–620 (2021).

48. Pallesen, J. process_sumstats. at https://github.com/ymer/process_sumstats?fbclid=IwAR0qjxK3ppXjUVPed3IIMNoMCQJvRhnuJ1fH64TD3GYB63ZZksWkkLc5Huc (2024).

49. Lundberg, S. M. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. in *Advances in Neural Information Processing Systems* (eds. Guyon, I. et al.) vol. 30 (Curran Associates, Inc., 2017).

50. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical Surface-Based Analysis. *Neuroimage* **9**, 179–194 (1999).

51. Fischl, B., Sereno, M. I. & Dale, A. M. Cortical Surface-Based Analysis. *Neuroimage* **9**, 195–207 (1999).

52. Lumaca, M., Keller, P. & Pando-naude, V. Frontoparietal network topology as a neuromarker of music perceptual abilities. 1–29 (1963) doi:10.21203/rs.3.rs-3930575/v1.

53. Schmid, R. *et al.* Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. *Nat. Commun.* **12**, 3832 (2021).

54. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).

55. Stekhoven, D. J. & Bühlmann, P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112–118 (2012).

56. Dalsgaard, S. *et al.* Incidence Rates and Cumulative Incidences of the Full Spectrum of Diagnosed

Mental Disorders in Childhood and Adolescence. *JAMA Psychiatry* **77**, 155 (2020).

57. archetypes. http://www.jstatsoft.org/v30/i08/.

58. Vilagut, G., Forero, C. G., Barbaglia, G. & Alonso, J. Screening for Depression in the General Population with the Center for Epidemiologic Studies Depression (CES-D): A Systematic Review with Meta-Analysis. *PLoS One* **11**, e0155431 (2016).

59. Aldao, A., Gee, D. G., De Los Reyes, A. & Seager, I. Emotion regulation as a transdiagnostic factor in the development of internalizing and externalizing psychopathology: Current and future directions. *Dev. Psychopathol.* **28**, 927–946 (2016).

60. Preece, D. A. *et al.* Alexithymia profiles and depression, anxiety, and stress. *J. Affect. Disord.* **357**, 116–125 (2024).

61. Crisp, Z. C. & Grant, J. E. Impulsivity across psychiatric disorders in young adults. *Compr. Psychiatry* **130**, 152449 (2024).

62. Baselmans, B. M. L. *et al.* Multivariate genome-wide analyses of the well-being spectrum. *Nat. Genet.* **51**, 445–451 (2019).

63. McGrath, J. J. *et al.* Age of onset and cumulative risk of mental disorders: a cross-national analysis of population surveys from 29 countries. *The lancet. Psychiatry* **10**, 668–681 (2023).

64. Insel, T. R. & Cuthbert, B. N. Medicine. Brain disorders? Precisely. *Science* **348**, 499–500 (2015).

65. Sommer, M. *Mental Health Among Youth in Denmark*. https://nordicwelfare.org/en/publikationer/mental-health-among-youth-in-denmark/ (2017).

66. Zinbarg, R. E. *et al.* Testing a Hierarchical Model of Neuroticism and Its Cognitive Facets: Latent Structure and Prospective Prediction of First Onsets of Anxiety and Unipolar Mood Disorders During 3 Years in Late Adolescence. *Clin. Psychol. Sci.* **4**, 805–824 (2016).

67. Furtado, C. P., Maller, J. J. & Fitzgerald, P. B. A magnetic resonance imaging study of the entorhinal cortex in treatment-resistant depression. *Psychiatry Res. Neuroimaging* **163**, 133–142 (2008).

68. Schultz, C. C. *et al.* Psychopathological correlates of the entorhinal cortical shape in schizophrenia. *Eur. Arch. Psychiatry Clin. Neurosci.* **260**, 351–358 (2010).

69. van Erp, T. G. M. *et al.* Cortical Brain Abnormalities in 4474 Individuals With Schizophrenia and 5098 Control Subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium. *Biol. Psychiatry* **84**, 644–654 (2018).

70. Vissink, C. E. *et al.* Structural Brain Volumes of Individuals at Clinical High Risk for Psychosis: A Meta-analysis. *Biol. Psychiatry Glob. Open Sci.* **2**, 147–152 (2022).

71. Hardege, I. *et al.* Neuronally produced betaine acts via a ligand-gated ion channel to control behavioral states. *Proc. Natl. Acad. Sci.* **119**, (2022).

72. Yang, Z.-J. *et al.* Betaine alleviates cognitive impairment induced by homocysteine through attenuating NLRP3-mediated microglial pyroptosis in an m6A-YTHDF2-dependent manner. *Redox Biol.* **69**, 103026 (2024).

73. Liang, Y. *et al.* Betaine eliminates CFA-induced depressive-like behaviour in mice may be through inhibition of microglia and astrocyte activation and polarization. *Brain Res. Bull.* **206**, 110863 (2024).

74. Chen, S.-T. *et al.* Betaine prevents and reverses the behavioral deficits and synaptic dysfunction induced by repeated ketamine exposure in mice. *Biomed. Pharmacother.* **144**, 112369 (2021).

75. Moustafa, A. A., Hewedi, D. H., Eissa, A. M., Frydecka, D. & Misiak, B. Homocysteine levels in schizophrenia and affective disordersâ€"focus on cognition. *Front. Behav. Neurosci.* **8**, (2014).

76. Bhatt, M., Di Iacovo, A., Romanazzi, T., Roseti, C. & Bossi, E. Betaine—The dark knight of the brain. *Basic Clin. Pharmacol. Toxicol.* **133**, 485–495 (2023).

77. Jones, L. L., McDonald, D. A. & Borum, P. R. Acylcarnitines: Role in brain. *Prog. Lipid Res.* **49**, 61–75 (2010).

78. Cao, B. *et al.* Metabolic profiling for water-soluble metabolites in patients with schizophrenia and healthy controls in a Chinese population: A case-control study. *World J. Biol. Psychiatry* **21**, 357–367 (2020).

79. Wong, C. B., Tanaka, A., Kuhara, T. & Xiao, J. Potential Effects of Indole-3-Lactic Acid, a Metabolite of Human Bifidobacteria, on NGF-Induced Neurite Outgrowth in PC12 Cells. *Microorganisms* **8**, 398 (2020).

80. Brydges, C. R. *et al.* Indoxyl sulfate, a gut microbiome-derived uremic toxin, is associated with psychic anxiety and its functional magnetic resonance imaging-based neurologic signature. *Sci. Rep.* **11**, 21011 (2021).

81. Davidson, M., Rashidi, N., Nurgali, K. & Apostolopoulos, V. The Role of Tryptophan Metabolites in Neuropsychiatric Disorders. *Int. J. Mol. Sci.* **23**, (2022).

82. Schneider, M. *et al.* Lipids in psychiatric disorders and preventive medicine. *Neurosci. Biobehav.*

Rev. **76**, 336–362 (2017).

Acknowledgments:

This article is based upon work from COST Action CA18106 (The Neural Architecture of Consciousness), supported by COST (European Cooperation in Science and Technology). The study was supported by grants from The Lundbeck Foundation (ADB and PQ), The Jascha Foundation (PQ), Direktør Kurt Bønnelycke og Hustrufru Grethe Bønnelyckes Fond (PQ), and Oda and Hans Svenningsens Fond (PQ). We extend our special thanks to Alexander Fjældsted, Francesca Fardo, Inga Griskova-Bulanova, Martin Dresler, and Henrique Fernandes for their significant input in developing the questionnaires utilized in this research. We thank Claude Bajada for providing MRI analysis scripts/processed data available. We also wish to acknowledge Sierra Lyn Wittrup Olsen for her meticulous assistance with DNA extraction.

Author contributions:

Conceptualization: PQ

Methodology: KS, FO, ME, JBG, NM, ABC, PQ, JDG, BZ, TLK, SD, JB, MGM, MY

Investigation: NM, ABC, TK,, SD, JD, FO

Visualization: NM, ABC, PQ

Funding acquisition: KS, ADB, PQ

Project administration: PQ

Supervision: KS, PQ

Writing – original draft: NM, ABC and PQ

Writing – review & editing: All authors

Figure 1

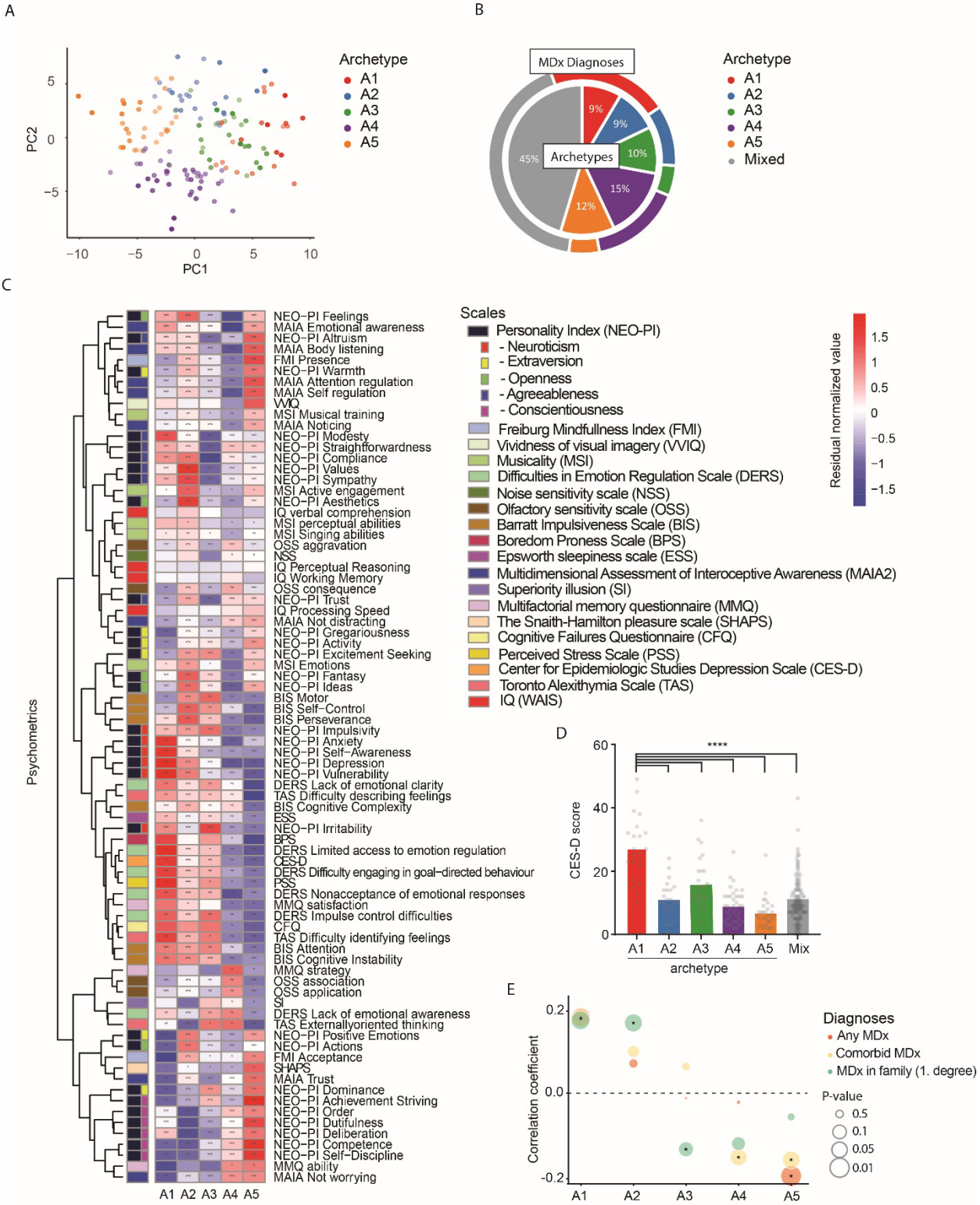

Figure 1 | A) Representation of the five archetypes (A1-5) projected in dimensions following principal component analysis. Each dot represents an individual, with color codes representing the each of the five archetypes. Strength of colors represents the level of archetype affiliation. Only individual a dominating archetype affiliation is shown (membership > 0.5). **B)** Pie chart showing

the percentage of individuals with a dominating affiliation to one of the five archetypes, as well as the relative distribution of self-reported MDx diagnoses among the archetypes. **C)** Heatmap showing the psychometric characteristics of each of the five archetypes. All variables were rank-normally transformed, and test of significance of differences between individuals in a given archetype and all other individuals indicated (Mann-Whitney U test). **D)** Plot of CES-D scores for each of the five archetypes and the mix group. **E)** Correlations between archetype scores and respectively any self-reported MDx diagnoses, more than one diagnosis (comorbid MDx), as well as MDx diagnoses among first degree relatives. Values statistically different from zero are marked as $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.

Figure 2

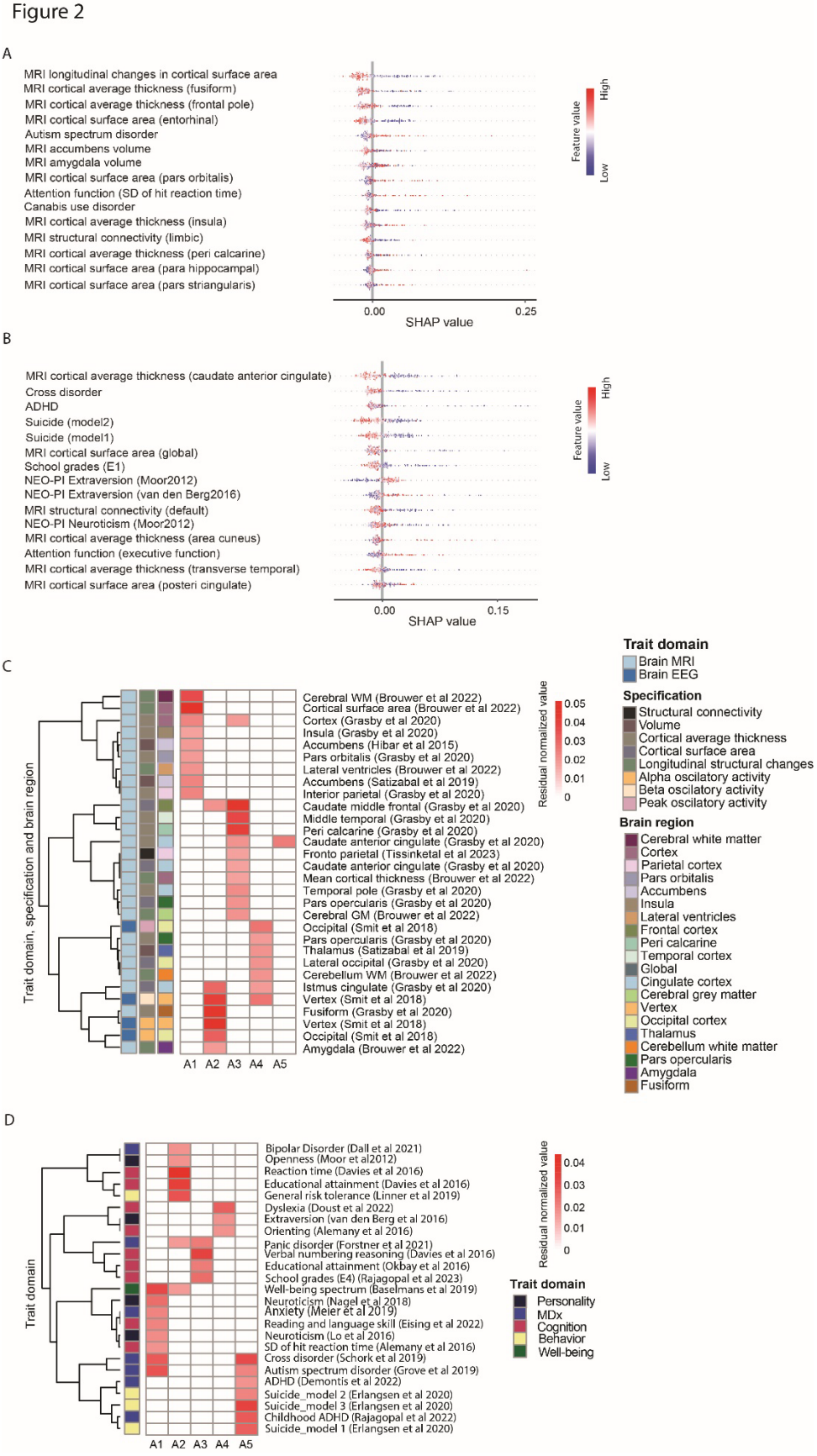

Figure 2 | Shapley additive explanations (SHAP) analysis of impact on model of top-tanked PGSs **A)** A1 scores **B)** A5 scores. Y-axis indicates the feature names in order of importance from top to bottom, whereas X-axis represents the SHAP value, which indicates the degree of change in log odds. The color of each point on the graph represents the value of the corresponding feature, with red indicating high values and blue indicating low values. **C)** Heatmap showing coefficient of determination (R^2) for PGSs with nominally significant prediction ($p > 0.05$) for the A1 score, and **D)** the A5 score.

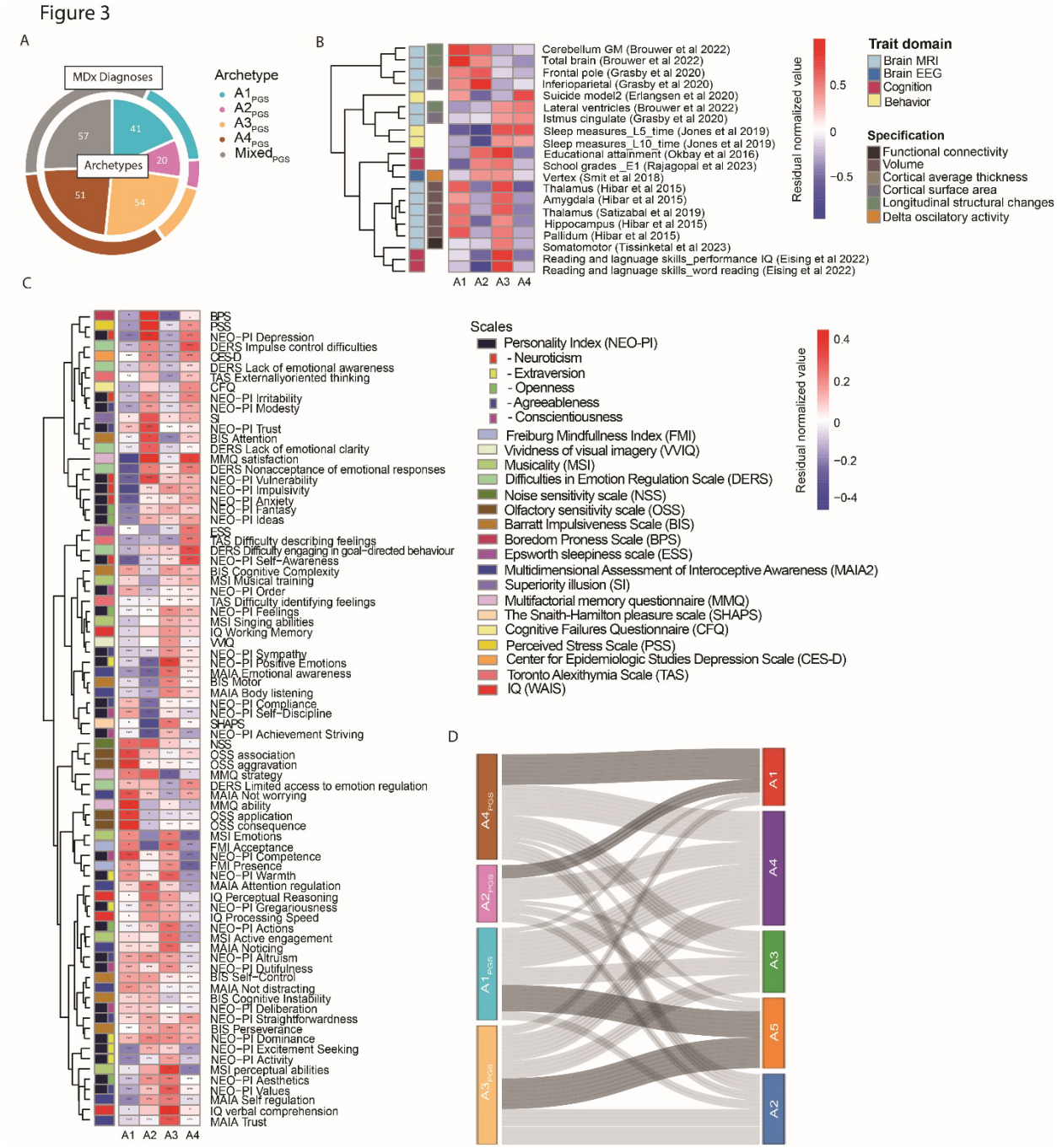

Figure 3 | A) Pie chart showing the percentage of individuals with a dominating affiliation to one of the four PGS-based archetypes, as well as the relative distribution of self-reported MDx diagnoses among these archetypes. **B)** Heatmap showing the neuroimaging-linked PGSs that define each of the four PGS-based archetypes. All variables were rank-normally transformed, and test of significance of differences between individuals in a given archetype and all other individuals indicated (Mann-Whitney U test). **C)** Heatmap showing the behavior-, cognition, and MDx-linked PGSs that define each of the four PGS-based archetypes. All variables were rank-normally transformed, and test of significance of differences between individuals in a given archetype and all other individuals indicated (Mann-Whitney U test). **D)** Sankey plot showing the relative overlap of individuals between the five psychometry-based archetypes and the four PGS-based archetypes. Overlap between risk associated as well as resilience-associated psychometry and PGS-based archetypes are highlighted in dark grey.

Figure 4

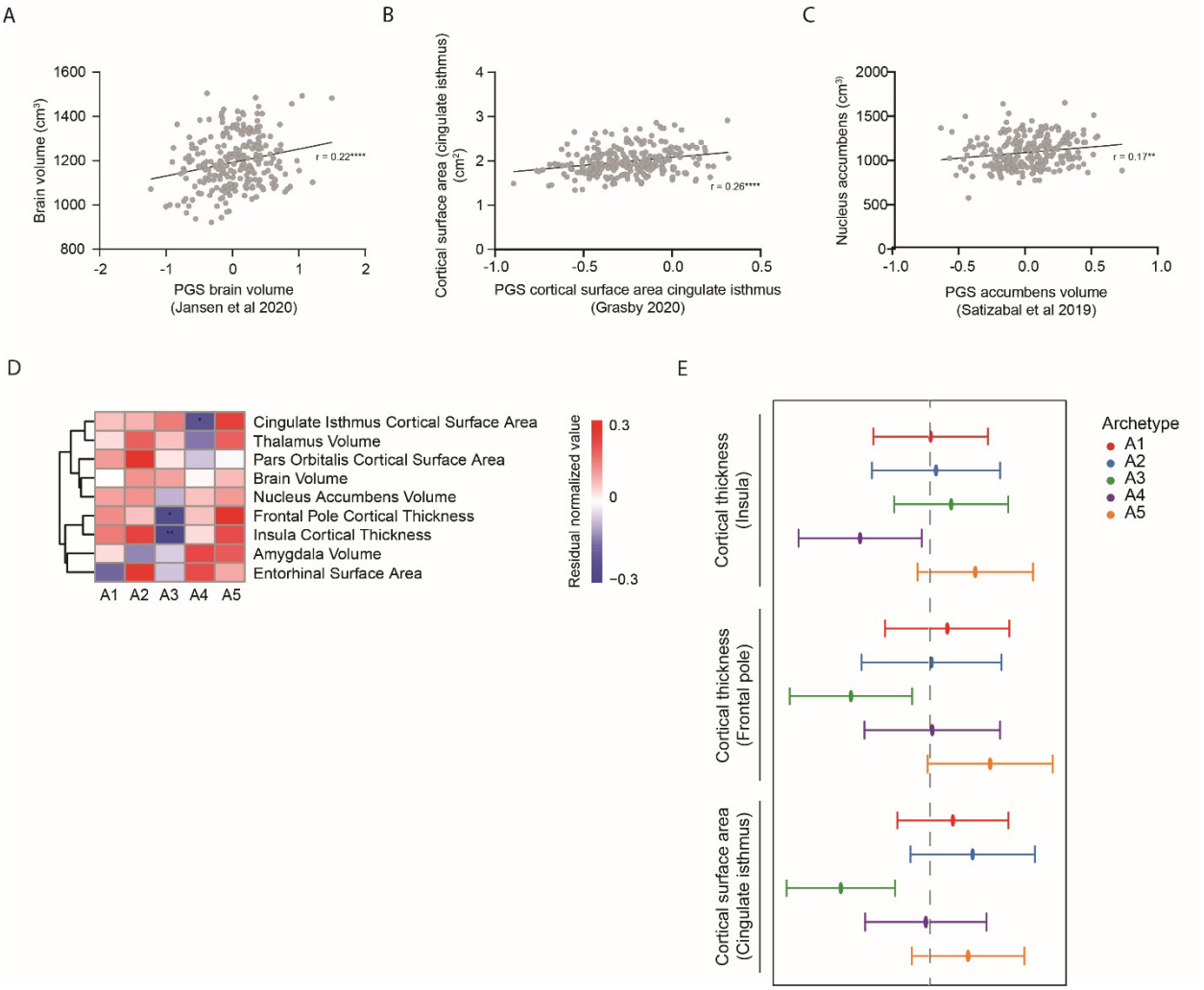

Figure 4 | Correlation between neuroimaging traits measured by MRI and PGSs calculated for the same traits. Each dot represent an individual. **A)** brain volume, **B)** cortical surface area (cingulate isthmus) **C)** nucleus accumbens. **D)** Heatmap showing rank norally transformed values for each trait and for each archetype (A1-5). Statistically significant results from linear regression are marked by $*p < 0.05$, $**p < 0.01$ **E)** Forest plot showing MRI measures that differ significantly between archetypes.

Figure 5

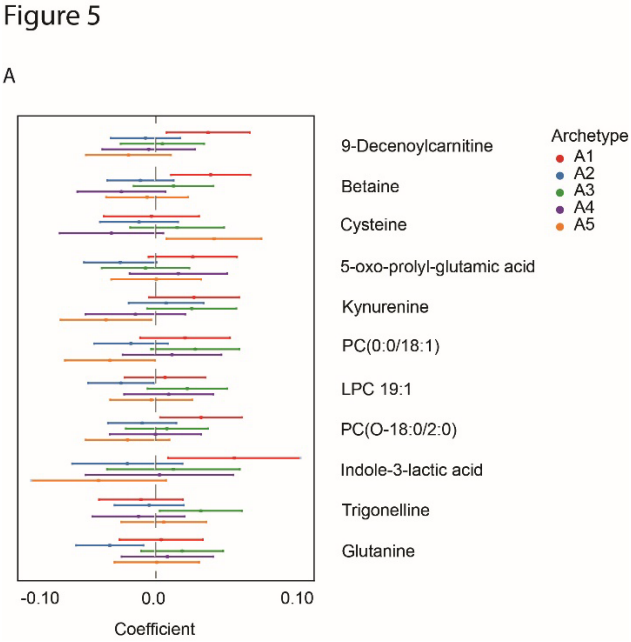

Figure 5 | Forest plot showing the association between archetype and annotated blood metabolites.

Table 1 | Distribution of self-reported diagnoses across archetypes

	Archetypes					
	A1	**A2**	**A3**	**A4**	**A5**	**Mixed**
# of Individuals	22	24	26	39	30	117
with any MDx diagnosis	7 (32%)	5 (21%)	3 (12%)	6 (15%)	1 (3%)	12 (10%)
with comorbid MDx diagnoses	3 (14%)	3 (13%)	2 (8%)	3 (8%)	0 (0%)	4 (3%)
with first-degree relatives with MDx diagnosis	10 (45%)	10 (42%)	4 (15%)	6 (15%)	8 (27%)	27 (23%)
Diagnostic entities:						
ADHD	2	2	2	0	0	2
Anxiety or phobia	2	1	0	3	1	3
Depression	4	2	2	5	0	8
Substance use disorder	1	0	1	0	0	1
Personality disorder	1	0	0	1	0	0
Self-harm	0	1	1	0	0	0
Eating disorder	2	1	0	0	0	0
Stress related disorder	3	1	2	0	0	3