

Two neurostructural subtypes: results of machine learning on brain images from 4,291 individuals with schizophrenia

Yuchao Jiang^{1,2}, Cheng Luo^{3,4,5}, Jijun Wang⁶, Lena Palaniyappan⁷, Xiao Chang^{1,2}, Shitong Xiang^{1,2}, Jie Zhang^{1,2}, Mingjun Duan³, Huan Huang³, Christian Gaser^{8,9,10}, Kiyotaka Nemoto¹¹, Kenichiro Miura¹², Ryota Hashimoto¹², Lars T. Westlye^{13,14,15}, Genevieve Richard^{13,14,15}, Sara Fernandez-Cabello^{13,14,15}, Nadine Parker¹⁴, Ole A. Andreassen¹⁴, Tilo Kircher¹⁶, Igor Nenadić¹⁶, Frederike Stein¹⁶, Florian Thomas-Odenthal¹⁶, Lea Teutenberg¹⁶, Paula Usemann¹⁶, Udo Dannlowski¹⁷, Tim Hahn¹⁷, Dominik Grotegerd¹⁷, Susanne Meinert¹⁷, Rebekka Lencer^{17,68,69}, Yingying Tang⁶, Tianhong Zhang⁶, Chunbo Li⁶, Weihua Yue^{18,19,20}, Yuyanan Zhang¹⁸, Xin Yu¹⁸, Enpeng Zhou¹⁸, Ching-Po Lin²¹, Shih-Jen Tsai²², Amanda L. Rodrigue²³, David Glahn²³, Godfrey Pearlson²⁴, John Blangero²⁵, Andriana Karuk^{26,27}, Edith Pomarol-Clotet^{26,27}, Raymond Salvador^{26,27}, Paola Fuentes-Claramonte^{26,27}, María Ángeles Garcia-León^{26,27}, Gianfranco Spalletta²⁸, Fabrizio Piras²⁸, Daniela Vecchio²⁸, Nerisa Banaj²⁸, Jingliang Cheng²⁹, Zhening Liu³⁰, Jie Yang³⁰, Ali Saffet Gonul³¹, Ozgul Uslu³², Birce Begum Burhanoglu³², Aslihan Uyar Demir³¹, Kelly Rootes-Murdy³³, Vince D. Calhoun³³, Kang Sim^{34,35,36}, Melissa Green³⁷, Yann Quidé³⁸, Young Chul Chung^{39,40,41}, Woo-Sung Kim^{39,41}, Scott R. Sponheim^{42,43,44}, Caroline Demro⁴³, Ian S. Ramsay⁴³, Felice Iasevoli⁴⁵, Andrea de Bartolomeis⁴⁵, Annarita Barone⁴⁵, Mariateresa Ciccarelli⁴⁵, Arturo Brunetti⁴⁶, Sirio Cocozza⁴⁶, Giuseppe Pontillo⁴⁶, Mario Tranfa⁴⁶, Min Tae M. Park^{47,48}, Matthias Kirschner^{49,50}, Foivos Georgiadis⁵⁰, Stefan Kaiser⁴⁹, Tamsyn E Van Rheenen^{51,52}, Susan L Rossell⁵², Matthew Hughes⁵², William Woods⁵², Sean P Carruthers⁵², Philip Sumner⁵², Elysha Ringin⁵³, Filip Spaniel⁵³, Antonin Skoch^{53,54}, David Tomecek^{53,55,56}, Philipp Homan^{57,58}, Stephanie Homan^{59,60}, Wolfgang Omlor⁵⁷, Giacomo Cecere⁵⁷, Dana D Nguyen⁶¹, Adrian Preda⁶², Sophia Thomopoulos⁶³, Neda Jahanshad⁶³, Long-Biao Cui⁶⁴, Dezhong Yao^{3,4,5}, Paul M. Thompson⁶³, Jessica A. Turner⁶⁵, Theo G.M. van Erp^{66,67}, Wei Cheng^{1,2,70,71,72}, ENIGMA Schizophrenia Consortium, ZIB Consortium, Jianfeng Feng^{1,2,72,73,74,75,76,*}

1 Institute of Science and Technology for Brain Inspired Intelligence, Fudan University, Shanghai, China.

2 Key Laboratory of Computational Neuroscience and Brain Inspired Intelligence (Fudan University), Ministry of Education, Shanghai, China.

3 The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, School of life Science and technology, University of Electronic Science and Technology of China, Chengdu, China.

4 High-Field Magnetic Resonance Brain Imaging Key Laboratory of Sichuan Province, Center for Information in Medicine, University of Electronic Science and Technology of China, Chengdu, China.

5 Research Unit of NeuroInformation (2019RU035), Chinese Academy of Medical Sciences, Chengdu, China.

6 Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China.

7 Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montréal, Canada.

8 Department of Psychiatry and Psychotherapy, Jena University Hospital, Jena, Germany.

9 Department of Neurology, Jena University Hospital, Jena, Germany.

10 German Center for Mental Health (DZPG), Site Jena-Magdeburg-Halle, Germany.

11 Department of Psychiatry, Division of Clinical Medicine, Institute of Medicine, University of Tsukuba, Tsukuba, 305-8575, Japan.

- 45 12 Department of Pathology of Mental Diseases, National Institute of Mental Health, National Center of
46 Neurology and Psychiatry, Kodaira, 187-8553, Japan.
- 47 13 Department of Psychology, University of Oslo, Oslo, Norway.
- 48 14 NORMENT Centre, Division of Mental Health and Addiction, Oslo University Hospital & Institute of
49 Clinical Medicine, University of Oslo, Oslo, Norway.
- 50 15 KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo and Oslo University
51 Hospital, Oslo, Norway.
- 52 16 Department of Psychiatry and Psychotherapy, Philipps Universität Marburg, Rudolf-Bultmann-Str. 8,
53 35039 Marburg, Germany.
- 54 17 Institute for Translational Psychiatry, University of Münster, Münster, Germany.
- 55 18 Peking University Sixth Hospital, Peking University Institute of Mental Health, NHC Key Laboratory of
56 Mental Health (Peking University), National Clinical Research Center for Mental Disorders (Peking
57 University Sixth Hospital), Beijing, PR China.
- 58 19 Chinese Institute for Brain Research, Beijing, PR China.
- 59 20 PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, PR China.
- 60 21 Institute of Neuroscience, National Yang Ming Chiao Tung University, Taipei, Taiwan.
- 61 22 Department of Psychiatry, Taipei Veterans General Hospital, Taipei, Taiwan.
- 62 23 Department of Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston MA, USA.
- 63 24 Olin Neuropsychiatry Research Center, Institute of Living, Hartford, CT, USA.
- 64 25 Department of Human Genetics and South Texas Diabetes and Obesity Institute, School of Medicine,
65 University of Texas of the Rio Grande Valley, Brownsville, TX, USA.
- 66 26 FIDMAG Germanes Hospitalàries Research Foundation, Barcelona 08035, Spain.
- 67 27 Centro de Investigación Biomédica en Red de Salud Mental, Instituto de Salud Carlos III, Spain.
- 68 28 Neuropsychiatry Laboratory, Department of Clinical Neuroscience and Neurorehabilitation, IRCCS
69 Santa Lucia Foundation, Rome, Italy.
- 70 29 Department of MRI, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China.
- 71 30 National Clinical Research Center for Mental Disorders, Department of Psychiatry, The Second
72 Xiangya Hospital of Central South University, Changsha, Hunan, PR China.
- 73 31 Ege University School of Medicine Department of Psychiatry, SoCAT Lab, Izmir, Turkey.
- 74 32 Ege University Institute of Health Sciences Department of Neuroscience, Izmir, Turkey.
- 75 33 Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS)
76 [Georgia State University, Georgia Institute of Technology, Emory University], Atlanta, GA, USA.
- 77 34 West Region, Institute of Mental Health, Singapore.
- 78 35 Yong Loo Lin School of Medicine, National University of Singapore, Singapore.
- 79 36 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.
- 80 37 School of Clinical Medicine, University of New South Wales, Sydney, Australia.
- 81 38 School of Psychology, University of New South Wales, Sydney, Australia.
- 82 39 Department of Psychiatry, Jeonbuk National University Hospital, Jeonju, Korea.
- 83 40 Department of Psychiatry, Jeonbuk National University, Medical School, Jeonju, Korea.
- 84 41 Research Institute of Clinical Medicine of Jeonbuk National University-Biomedical Research Institute
85 of Jeonbuk National University Hospital, Jeonju, Korea.
- 86 42 Minneapolis VA Medical Center, University of Minnesota, Minneapolis, MN, USA.
- 87 43 Department of Psychiatry and Behavioral Sciences, University of Minnesota, Minneapolis, MN, USA.
- 88 44 Department of Psychology, University of Minnesota, Minneapolis, MN, USA.

- 89 45 Section of Psychiatry - Department of Neuroscience - University "Federico II", Naples, Italy.
90 46 Department of Advanced Biomedical Sciences - University "Federico II", Naples, Italy.
91 47 Department of Psychiatry, Temerty Faculty of Medicine, University of Toronto, Toronto, Canada.
92 48 Centre for Addiction and Mental Health, Toronto, Canada.
93 49 Division of Adult Psychiatry, Department of Psychiatry, University Hospitals of Geneva, Switzerland.
94 50 Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric Hospital University of
95 Zurich, Switzerland.
96 51 Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne, Melbourne,
97 Australia.
98 52 Centre for Mental Health and Brain Sciences, School of Health Sciences, Swinburne University,
99 Melbourne, Australia.
100 53 National Institute of Mental Health, Klecany, Czech Republic.
101 54 MR Unit, Department of Diagnostic and Interventional Radiology, Institute for Clinical and
102 Experimental Medicine, Prague, Czech Republic.
103 55 Institute of Computer Science, Czech Academy of Sciences, Prague, Czech Republic.
104 56 Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic.
105 57 Psychiatric Hospital, University of Zurich, Zurich, Switzerland.
106 58 Neuroscience Center Zurich, University of Zurich & Swiss Federal Institute of Technology Zurich,
107 Zurich, Switzerland.
108 59 Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric University Hospital
109 Zurich, Switzerland.
110 60 Experimental Psychopathology and Psychotherapy, Department of Psychology, University of Zurich,
111 Switzerland.
112 61 Department of Pediatrics, University of California Irvine, Irvine, California, USA.
113 62 Department of Psychiatry and Human Behavior, University of California Irvine, Irvine, California,
114 USA.
115 63 Imaging Genetics Center, Stevens Neuroimaging & Informatics Institute, Keck School of Medicine,
116 University of Southern California, Los Angeles, CA, USA.
117 64 Department of Clinical Psychology, Fourth Military Medical University, Xi'an, PR China.
118 65 Psychiatry and Behavioral Health, Ohio State Wexner Medical Center, Columbus, OH, USA.
119 66 Clinical Translational Neuroscience Laboratory, Department of Psychiatry and Human Behavior,
120 University of California Irvine, Irvine Hall, room 109, Irvine, CA, 92697-3950, USA.
121 67 Center for the Neurobiology of Learning and Memory, University of California Irvine, 309 Qureshey
122 Research Lab, Irvine, CA, 92697, USA.
123 68 Department of Psychiatry and Psychotherapie and Center for Brain, Behavior and Metabolism,
124 Lübeck University, Lübeck, Germany.
125 69 Institute for Transnational Psychiatry and Otto Creutzfeldt Center for Behavioral and Cognitive
126 Neuroscience, University of Münster, Münster, Germany.
127 70 Shanghai Medical College and Zhongshan Hospital Immunotherapy Technology Transfer Center,
128 Shanghai, China.
129 71 Department of Neurology, Huashan Hospital, Fudan University, Shanghai, China.
130 72 Fudan ISTBI—ZJNU Algorithm Centre for Brain-Inspired Intelligence, Zhejiang Normal University,
131 Jinhua, China.
132 73 MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China.

- 133 74 Zhangjiang Fudan International Innovation Center, Shanghai, China.
134 75 School of Data Science, Fudan University, Shanghai, China.
135 76 Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK.
136 * Corresponding Authors to: Jianfeng Feng, jffeng@fudan.edu.cn

137 **Abstract**

138 Machine learning can be used to define subtypes of psychiatric conditions based on shared
139 clinical and biological foundations, presenting a crucial step toward establishing
140 biologically based subtypes of mental disorders. With the goal of identifying subtypes of
141 disease progression in schizophrenia, here we analyzed cross-sectional brain structural
142 magnetic resonance imaging (MRI) data from 4,291 individuals with schizophrenia (1,709
143 females, age=32.5 years±11.9) and 7,078 healthy controls (3,461 females, age=33.0
144 years±12.7) pooled across 41 international cohorts from the ENIGMA Schizophrenia
145 Working Group, non-ENIGMA cohorts and public datasets. Using a machine learning
146 approach known as Subtype and Stage Inference (SuStain), we implemented a brain
147 imaging-driven classification that identifies two distinct neurostructural subgroups by
148 mapping the spatial and temporal trajectory of gray matter (GM) loss in schizophrenia.
149 Subgroup 1 (n=2,622) was characterized by an early cortical-predominant loss (ECL) with
150 enlarged striatum, whereas subgroup 2 (n=1,600) displayed an early subcortical-
151 predominant loss (ESL) in the hippocampus, amygdala, thalamus, brain stem and striatum.
152 These reconstructed trajectories suggest that the GM volume reduction originates in the
153 Broca's area/adjacent fronto-insular cortex for ECL and in the hippocampus/adjacent
154 medial temporal structures for ESL. With longer disease duration, the ECL subtype
155 exhibited a gradual worsening of negative symptoms and depression/anxiety, and less of
156 a decline in positive symptoms. We confirmed the reproducibility of these imaging-based
157 subtypes across various sample sites, independent of macroeconomic and ethnic factors
158 that differed across these geographic locations, which include Europe, North America and
159 East Asia. These findings underscore the presence of distinct pathobiological foundations
160 underlying schizophrenia. This new imaging-based taxonomy holds the potential to identify
161 a more homogeneous sub-population of individuals with shared neurobiological attributes,
162 thereby suggesting the viability of redefining existing disorder constructs based on
163 biological factors.

164

165 **Keywords:** schizophrenia; structural MRI; artificial intelligence; subtype; ENIGMA; brain
166 gray matter

167 **1. Introduction**

168 A key goal of biological psychiatry is to define biological subtypes of major mental
169 disorders, based on objective measures derived from imaging and other biomarkers [1]. In
170 psychiatry, redefining subtypes of psychiatric disorders based on neural mechanisms
171 augmenting clinical behavioral criteria presents significant benefits. By using biological
172 characteristics to more efficiently identify biologically homogeneous clinical cohorts, clinical
173 trials could more effectively discern the biological effects of a given intervention. Artificial
174 intelligence (AI) methods such as machine learning can be applied to brain imaging [2] to
175 categorize individuals based on their profiles of brain metrics, and holds great potential for
176 revealing the underlying neurobiological mechanisms associated with disorder subtypes
177 [3].

178 Schizophrenia is one of the most severely disabling psychiatric disorders with a life-
179 time prevalence of 1%; it affects approximately 26 million people worldwide [4]. The
180 etiology of schizophrenia is still not fully understood. Current knowledge implicates multiple
181 neurobiological mechanisms and pathophysiologic processes [5, 6]. Furthermore, people
182 diagnosed with schizophrenia show a substantial heterogeneity in clinical symptoms [7],
183 disease progression [8], treatment response [9], and other biological markers [10, 11]. In
184 addition, currently available treatments are not aligned with specific pathophysiological
185 pathways/targets, which limits effectiveness of treatment selection [12]. Establishing a new
186 taxonomy by identifying distinct subtypes based on neurobiological data could help resolve
187 some of these heterogeneity-induced challenges.

188 Machine learning algorithms are increasingly used to subtype brain disorders [13-16].
189 Prior studies have primarily focused on grouping individuals into distinct categories without
190 considering disease progression [17, 18]. A major obstacle to identifying distinct patterns
191 of neuro-pathophysiological progression (referred to as progression subtypes) stems from
192 the lack of sufficient longitudinal data covering the lifespan of the disorder. Recently, a
193 novel data-driven machine learning approach known as Subtype and Stage Inference
194 (SuStaln) was introduced [19]. SuStaln uses a large number of cross-sectional
195 observations, derived from single time-point MRI scans, to identify clusters (subtypes) of
196 individuals with common trajectory of disease progression (i.e., the sequence of MRI
197 abnormalities across different brain regions) in brain disorders [20-22]. By applying
198 SuStaln to MRI data from individuals with schizophrenia, primarily collected from the
199 Chinese population, we found that the progression of gray matter loss in schizophrenia can
200 be better characterized through two distinct phenotypes: one characterized by a cortical-
201 predominant progression, originating in the Broca's area/fronto-insular cortex, and another
202 marked by a subcortical-predominant progression, starting in the hippocampus [22]. Such
203 brain-based taxonomies may reflect neurostructural subtypes with shared
204 pathophysiological foundations, with relevance for neurobiological classification [22].
205 However, the generalisability of the two neurostructural subtypes to diverse populations
206 outside of China, and external validation of the subgrouping is required before applying
207 this knowledge to stratify clinical trials.

208 The Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA,
209 <http://enigma.ini.usc.edu>) consortium is dedicated to conducting large-scale analyses by
210 pooling brain imaging data from research teams worldwide, using standardized image
211 processing protocols. Previously, ENIGMA published findings revealing thinner cerebral
212 cortex, smaller surface area, and altered subcortical volumes in schizophrenia compared
213 to controls [23, 24]. Here, we included structural MRI data obtained from 4,291 individuals
214 diagnosed with schizophrenia and 7,078 healthy controls from 41 international cohorts
215 from ENIGMA schizophrenia groups worldwide and other non-ENIGMA datasets
216 (**Supplementary Table 1**). The large sample size allowed us to conduct systematic and
217 comprehensive analyses to verify the reproducibility and generality of neurostructural
218 subtypes of schizophrenia across regions/locations and disease stages. This study's aims
219 were: (1) to validate the two neurostructural subtypes with distinct trajectories of neuro-
220 pathophysiological progression in schizophrenia, (2) to verify the reproducibility and
221 generality of the neurostructural subtypes, in subsamples across the world and across
222 disease stages, and (3) to characterize subtype-specific signatures in terms of
223 neuroanatomy and clinical symptomatic trajectory.

224 Together, these analyses aim to create a new, easily accessible (with a single
225 anatomical MRI), interpretable (based on 'progressive' pathology) and robustly
226 generalizable (across ethnic, sex and language differences) taxonomy of subtypes that
227 share common neurobiological mechanisms in schizophrenia. If proven effective, other
228 complex neuropsychiatric disorders with high heterogeneity [25, 26], such as major
229 depressive disorder, autism spectrum disorder, and obsessive-compulsive disorder, could
230 also benefit from such a subtyping paradigm. This has the potential to transition the field
231 of psychiatry from syndrome-based to both syndrome- and biology-based stratifications of
232 mental disorders.

233 2. Results

234 2.1 Two biotypes with distinct pathophysiological progression trajectories

235 Distinct patterns of spatiotemporal progression of pathophysiological progression
236 were identified using SuStaln, based on cross-sectional MRI data from 4,222 individuals
237 diagnosed with schizophrenia (1,683 females, mean age= 32.4 ± 11.9 years) and 7,038
238 healthy subjects (3,440 females, mean age= 33.0 ± 12.6 years) (**Table.1**). A 2-fold cross-
239 validation procedure resulted in an optimal number of $K=2$ clusters (subtypes) as
240 determined by the largest Dice coefficient (**Fig.1a**), indicating the best consistency of the
241 subtype labeling across all individuals between for a model in two independent
242 schizophrenia populations. **Fig.1b** shows that only 1.2% of people were moved from
243 subtype 1 to subtype 2, and 7.5% were moved from subtype 2 to subtype 1, indicating that
244 91.3% of individuals' subtype labels were consistent between the SuStaln classifications
245 from two non-overlapping data folds. These findings suggest the presence of two stable
246 schizophrenia biotypes with distinct 'trajectories' of pathophysiological progression (here
247 we put trajectory in quotes as the typical sequence of disease progression is reconstructed
248 from cross-sectional data).

249 Region of interest (ROI)-wise gray matter volume (GMV) z-scores, at each stage of
250 the 'trajectory' for each subtype, show the sequence of regional volume loss across the 17
251 brain regions for each 'trajectory' (**Fig.1c**). To visualize the spatiotemporal pattern of each
252 'trajectory', z-score whole brain images were mapped to a glass brain template (**Fig.1d**).
253 These maps show a progressive pattern of spatial expansion along with later 'temporal'
254 stages of pathological progression distinct for each 'trajectory'. Specifically, 'trajectory' 1
255 displayed an 'early cortical-predominant loss' biotype. It was characterized by an initial
256 reduction in Broca's area, followed by adjacent fronto-insular regions, then extending to
257 the rest of the neocortex, and finally to the subcortex (**Fig.1d**). Conversely, 'trajectory' 2
258 exhibited an 'early subcortical-predominant loss' biotype where volume loss began in the
259 hippocampus, spread to the amygdala and parahippocampus, and then extended to the
260 accumbens and caudate before affecting the cerebral cortex (**Fig.1d**). The two 'trajectories'
261 were highly consistent with our previous findings in a predominantly Chinese schizophrenia
262 cohort [22]. The phenotypic subtypes, based on the different pathophysiological
263 'trajectories', are thus replicated in a large cross-geography sample, confirming the
264 presence of two different neuropathological pathways with different anatomical origins in
265 schizophrenia [22].

266

267 2.2 Trajectories are repeated in first-episode and medication-naïve samples

268 The sample size of this study was large enough to allow further exploratory analyses
269 to identify pathophysiological progression trajectories in more homogeneous subsamples
270 of schizophrenia. Here, we re-estimated the SuStaln 'trajectories' based on a subsample
271 of data from individuals with first-episode schizophrenia with illness duration less than two
272 years ($N=1,122$; 513 females, mean age= 25.4 ± 8.6 years), and a subsample of
273 medication-naïve individuals with schizophrenia ($N=718$, 353 females, mean age= $23.7 \pm$
274 7.8 years) (**Supplementary Table 3**). In both subsamples, we replicated the two
275 'trajectories' with either the Broca's area or the hippocampus as the sites of origin

276 (Extended Data Fig.1), indicating that the two initiating regions - ranking ahead other
277 regional deficits - are the pathological effects of the disease itself, rather than medication-
278 induced effects. Broca's area and the hippocampus may therefore be candidate targets for
279 intervention in schizophrenia, as these two brain regions were affected early in the disease
280 process.

281

282 **2.3 Trajectories are reproducible for samples from different parts of the world**

283 To examine whether the 'trajectories' were reproducible for samples from different
284 parts of the world, we divided all samples into several sub-cohorts based on where the
285 samples were obtained (Extended Data Fig.5). Here, samples from China, Japan, South
286 Korea and Singapore were classified into the East Asian ancestry (EAS) cohort. Samples
287 from Europe, the United States, Canada and Australia were classified into the European
288 ancestry (EUR) cohorts (Supplementary Table 4). In addition, Chinese, Japanese,
289 European and North American cohorts were further classified by their site locations by their
290 site locations in terms of geographic distribution (Supplementary Table 4). Such a division
291 was based on the similar ethnic or environmental factors for each country, region, or
292 continent and the size of subsample, which need to be sufficient to conduct a reliable
293 inference of the SuStaln trajectory. We found that two 'trajectories' (the optimal number
294 was also $K=2$, which separately re-estimated in each cohort) - with Broca's area leading
295 and the hippocampus leading - were also repeated in EAS (Fig.2a) and EUR (Fig.2b)
296 cohorts. In addition, the spatiotemporal pattern of each 'trajectory' showed strong,
297 significant correlations between the EAS and EUR cohorts ('trajectory' 1, $r=0.948$, $p<0.001$;
298 'trajectory' 2, $r=0.842$, $p<0.001$; Spearman correlation test). This high level of similarity in
299 the trajectories was also observed between cohorts from other locations (Fig.2c). This
300 suggests that the two biotypes with distinct 'trajectories' of pathophysiological progression
301 in schizophrenia are robust, and their classification patterns are independent of macro-
302 environmental or ethnogenetic factors.

303

304 **2.4 Trajectories are associated with neurophysiological, pathological and** 305 **neuropsychological progressions in schizophrenia**

306 The SuStaln calculated the probability of each patient belonging to a specific
307 'trajectory' and further assigned them to a sub-stage within that trajectory. Individuals who
308 were assigned to the later stages of the 'trajectory' showed significant correlation with less
309 GMV of Broca's area (Fig.1e, $r=0.651$, $p<0.0001$) and hippocampus (Fig.1f, $r=0.615$,
310 $p<0.0001$). In addition, the later stages were correlated with longer disease duration
311 (Fig.1g, $r=0.105$, $p<0.0001$), worse negative symptoms (Fig.1h, $r=0.101$, $p<0.0001$) and
312 worse cognitive symptoms (Fig.1i, $r=0.080$, $p=0.004$). These results suggest that the
313 SuStaln 'trajectory' reflects the underlying neural progression in schizophrenia.

314

315 **2.5 Subtype-specific signatures in neuroanatomical pathology**

316 To characterize subtype-specific neuroanatomical signatures, we assessed regional
317 morphological measures using FreeSurfer in a subsample including 1,840 individuals with
318 schizophrenia and 1,780 healthy controls. A total of 330 regional morphological measures

319 in cortical thickness, cortical surface area, cortical volume, subcortical volume and
320 subregion segmentation were quantified (see [Methods](#)).

321 Regional morphological z-scores (i.e., normative deviations from healthy control group)
322 for each subtype were computed and compared ([Fig.3](#)). Morphological z-scores of all brain
323 regions and inter-subtype comparisons are provided in the [Supplementary Table S5](#).
324 Briefly, compared to healthy controls, average cortical volume/area reduction was only
325 observed in subtype 1 ([Extended Data Fig.2a-b](#)), though both subtype 1 and subtype 2
326 exhibited a moderate reduction in average cortical thickness ([Extended Data Fig.2c](#)).
327 Additionally, largest effects for cortical thickness/volume/area were located within the
328 superior frontal regions for subtype 1 and in the superior/medial temporal regions for
329 subtype2 ([Supplementary Table S5](#)). As for subcortical volume, larger effects for volumes
330 of hippocampus, amygdala, thalamus, accumbens and brain stem were observed in
331 subtype 2 compared to subtype 1 ([Extended Data Fig.2d-h](#)). The hippocampal/amygdala
332 subregions with the most significant reduction for subtype 2 were located in the molecular
333 layer and cortico-amygdaloid transition area ([Extended Data Fig.3-4](#)). Interestingly, we
334 observed that, compared to healthy controls, the striatum (i.e., caudate, putamen) was
335 larger among subtype 1 patients and smaller among subtype 2 patients ([Extended Data](#)
336 [Fig.2i-j](#)). The difference in the striatum between the two subtypes was also replicated in a
337 subsample of medication-naive individuals with schizophrenia ([Supplementary Table S6](#)).
338 The main findings of subtype-specific neuroanatomical signatures are described in [Table](#)
339 [2](#). Taken together, subtype 1 exhibited greater deficits in cortical morphology but
340 enlargement volume of striatum, whereas subtype 2 displayed more severe volume loss in
341 the subcortical regions including hippocampus, amygdala, thalamus, brain stem as well as
342 the striatum.

343

344 **2.6 Clinical characterization of subtypes**

345 A total of 2,622 (62.1%) individuals with schizophrenia were assigned to subtype 1
346 and the remaining 1,600 patients (37.9%) were assigned to subtype 2. The two subtypes
347 exhibit no significant difference in the age, sex, illness duration or PANSS scores ([Table](#)
348 [1](#)). To further characterize the psychotic symptomatic trajectory as disease progresses for
349 each subtype, we further defined three subgroups according to illness duration (early stage:
350 <2 years; middle stage: 2-10 years; late stage: >10 years). The results suggested distinct
351 trajectories of psychotic symptoms between the two subtypes ([Fig.4](#) and [Table 3](#)).
352 Specifically, lower positive symptom severity was observed in late stage patients compared
353 early stage patients in both subtypes (subtype 1, $F=37.4$, $p=1.60e-16$; subtype2, $F=41.9$,
354 $p=4.68e-18$); however, at the late stage, subtype 1 exhibited worse positive symptom
355 compared to subtype 2 ($t=2.8$, $p=0.005$). With the increase of the disease course, subtype
356 1 showed a gradual worsening of negative symptom ($F=4.6$, $p=9.98e-3$), whereas the
357 negative symptoms of subtype 2 remained stable across the three stages of the disease
358 course ($F=0.1$, $p=0.884$). Additionally, a gradual worsening of depression/anxiety was only
359 observed in subtype 1 ($F=5.9$, $p=2.86e-3$), which showed worse depression/anxiety at the
360 late stage, compared to subtype 2 ($t=2.1$, $p=0.036$).

361 **3. Discussion**

362 Our study, applying a machine learning algorithm to brain MRI data from over 4,000
363 individuals with schizophrenia, has revealed two distinct neurostructural subtypes based
364 on patterns of neuro-pathological progression. These subtypes are reproducible and
365 generalizable across different subsamples and illness stages, independent of
366 macroeconomic and ethnic factors that differed across collection locations. Specific
367 patterns of neuroanatomical pathology for each subtype were uncovered. Subtype 1 is
368 characterized by early cortical-predominant loss that first occurs in the Broca's area/fronto-
369 insular cortex, and shows adverse signatures in cortical morphology and an enlarged
370 striatum. In contrast, subtype 2 is marked by early subcortical-predominant loss that first
371 appears in the hippocampus, and displays significant volume loss in subcortical regions,
372 including the hippocampus, amygdala, thalamus, brain stem and striatum. Additionally, we
373 observed distinct trajectories of specific symptoms clusters in these two subtypes: as
374 disease progresses, subtype 1 exhibited a gradual worsening of negative and
375 depression/anxiety symptoms, and less of a decline in positive symptoms compared to
376 subtype 2.

377 Despite the growing body of evidence pointing to group-level gray matter volume
378 deficits in various brain regions - especially in frontal and temporal regions - as well as
379 altered subcortical volume in schizophrenia [27], substantial individual variations persist
380 within this population [11, 28]. These inter-individual differences in brain structure may stem
381 from two primary sources of variation. First, differences in underlying etiology and
382 pathogenesis could result in varying clinical characteristics (referred to as phenotypic
383 heterogeneity) [6, 29]. Second, relative differences among subjects in the stage of dynamic
384 progression (known as temporal heterogeneity) could further increase differences in the
385 clinical presentation [30, 31]. Such variations suggest that the pathological progression of
386 schizophrenia might not be attributed to a single unified pathophysiological process.
387 Indeed, our neurostructural subtypes uncovered two patterns of gray matter loss trajectory
388 through brain structural imaging. Several studies also reported dynamic patterns of
389 accelerated gray matter loss over time in individuals with schizophrenia [32, 33]. In addition,
390 staging of trajectory within subtype reflects the underlying neurophysiological, pathological
391 and neuropsychological progressions in schizophrenia. Furthermore, we demonstrated
392 that the phenotypic difference in the intrinsic neuro-pathophysiological trajectory was
393 reproducible across samples worldwide, independent of macroeconomic and ethnic factors
394 that differed across these sites.

395 The Broca's area/fronto-insular cortex and hippocampus are identified separately in
396 subtype 1 and subtype 2 as the first regions to show gray matter deficits. This is consistent
397 with our prior finding based on individuals with schizophrenia primarily collected from the
398 Chinese population [22]. Furthermore, the current study replicates the same two primary
399 regions in a medication-naïve and a first-episode cohort, suggesting that these
400 neuropathological changes are a reflection of the disease process, rather than medication
401 effects. Broca's area and the fronto-insular cortex have been extensively implicated in
402 schizophrenia [34], supporting Crow's linguistic primacy hypothesis [35] and a triple-
403 network model of the disorder [36]. Moreover, in individuals with psychosis, reductions in

404 the inferior frontal cortex preceding the initial psychotic episode have been reported [37,
405 38]. A prior study reported reduced dopamine release in the prefrontal cortex in patients
406 with schizophrenia [39]. In relation to hippocampal pathology, research has emphasized
407 the hippocampus as one of the initial regions to display volumetric loss in schizophrenia
408 [40, 41]. The hippocampus is thought to be involved in potential glutamatergic dysfunction
409 in schizophrenia [6]. Decreased levels of the NMDA co-agonist D-serine were linked to
410 neurobiological alterations similar to those seen in schizophrenia, including hippocampal
411 volume loss [42]. These findings offer evidence regarding the specific neuroanatomical
412 locations where gray matter loss is first observed in the schizophrenia subtypes. These
413 two potential origins could also offer a new viewpoint on the pathological 'spread' of the
414 disorder.

415 The new subtyping method employed exhibits high potential for distinguishing
416 neurostructural subtypes with shared pathophysiological foundations. Notably, subtype 1
417 displayed larger volume of the striatum, while subtype 2 demonstrated reduced volume.
418 The striatum plays a key role in the dopamine system, which contributes to psychotic
419 symptoms [43]. Nevertheless, studies of striatal pathology have reported inconsistent
420 differences between patients and controls [6]. The variability of the striatum is greater in
421 patients than in controls, which relates to overall structural morphometry [27], dopamine
422 D2 receptor and transporter levels [44]. This indicates that differences might exist within
423 subgroups of the disorder [6]. In addition, it is still uncertain whether the discrepancy in
424 striatum between cases and controls indicates a primary pathology or an effect of
425 antipsychotic treatment [6]. Interestingly, this study's subtype-specific striatal differences
426 were replicated in a subset of individuals who had not received antipsychotic treatment,
427 suggesting that striatal variability persists even in those without antipsychotic treatment. In
428 addition, a recent study reveals a more pronounced and widespread pattern of thinner
429 cortex in deficit schizophrenia, a clinically defined subtype with primary, enduring negative
430 symptoms, compared to non-deficit schizophrenia [45]. This also suggests the existence
431 of distinct subtypes distinguished by unique neuroimaging features. Taken together, our
432 neurostructural subtyping differentiated subgroups with unique pathological features,
433 thereby enhancing our understanding of the neurobiological mechanisms underlying
434 schizophrenia.

435 The two newly identified subtypes may have several potential therapeutic implications.
436 While the underlying mechanisms associated with a subtype-specific symptomatic
437 trajectory remain unclear, our research shows divergent long-term clinical outcomes
438 between the two neurostructural subtypes. As the disease advanced, for subtype 1, the
439 negative and depression/anxiety symptoms gradually worsened; for subtype 2 these
440 symptoms remained stable. In addition, subtype1 experienced worse positive symptoms
441 than subtype 2 at the late stage of disease (i.e., duration > 10 years). This is consistent
442 with a prior study that reported greater gray matter reduction in frontal regions in treatment-
443 resistant compared with treatment-responsive individuals with schizophrenia [46]. Another
444 intriguing aspect is that our prior research on treatment-resistant schizophrenia
445 demonstrated that electroconvulsive therapy (ECT) can substantially enhance the volume
446 of the hippocampus and insula; this is also associated with psychotic symptom alleviation
447 [47-49]. Notably, these two brain regions were also identified as the 'origins' of gray matter

448 loss separately in each subtype. This observation raises the possibility of exploring
449 neuromodulation interventions, such as transcranial magnetic stimulation (TMS), to target
450 these specific brain regions.

451 This study has several limitations. First, while the SuStaln algorithm estimates
452 pathophysiological trajectories from cross-sectional MRI data, it remains crucial to validate
453 these outcomes with longitudinal data to verify the brain changes with disease progression
454 over time. Second, the current study benefits from a large sample size, but the inclusion of
455 data from various sites could potentially be influenced by confounding factors, including
456 diverse cohorts, scanners, and locations. Harmonization methods have been employed to
457 alleviate disparities across MRI acquisition protocols. Nonetheless, it remains essential to
458 collect a sufficiently large sample from multi-centers under a standard imaging protocol
459 and experimental paradigm. Third, a substantial portion of individuals with schizophrenia
460 were likely to have received or currently use medications, and data from medication-
461 naïve/free individuals were only available for a subset of the datasets. One important
462 limitation is the assumption of progressive pathology in schizophrenia (discrete events of
463 tissue loss or continuous downward drift), when applying SuStaln. The few existing very
464 long-term imaging studies in schizophrenia support this stance [50] but selection bias
465 cannot be fully overcome in the recruitment process for neuroimaging studies. Routine
466 anatomical MRI for every person with psychosis seeking help, with periodic repeats, may
467 provide better view of the validity of progressive pathology in the future.

468 In summary, our study reveals two distinct neurostructural schizophrenia subtypes
469 based on patterns of pathological progression of gray matter loss. We extend the
470 reproducibility and generalisability of these brain imaging-based subtypes across illness
471 stages, medication treatments and different sample locations worldwide, independent of
472 macroeconomic and ethnic factors that differed across these sites. The identified subtypes
473 exhibit distinct signatures of neuroanatomical pathology and psychotic symptomatic
474 trajectories, highlighting the heterogeneity of the neurobiological changes associated with
475 disease progress. This new imaging-based taxonomy shows potential for identification of
476 homogeneous subsamples of individuals with shared neurobiological characteristics. This
477 may be a first crucial step in the transition from only syndrome-based to both syndrome-
478 and biology-based identification of mental disorder subtypes in the near future.

479 **4. Methods**

480 **4.1 Study samples**

481 This study analyzed cross-sectional T1-weighted structural MRI data from a total of
482 4,291 individuals diagnosed with schizophrenia (1,709 females, mean age=32.5±11.9
483 years) and 7,078 healthy controls (3,461 females, mean age=33.0±12.7 years). These
484 datasets came from 21 cohorts of ENIGMA schizophrenia working groups from various
485 countries around the world, 11 cohorts collected from Chinese hospitals over the last ~10
486 years, and 9 cohorts from publicly available datasets, i.e., HCP-EP [51], JP-SRPBS [52],
487 fBIRN [53], MCIC [54], NMorphCH [55], NUSDAST [56], DS000030 [57], DS000115 [58]
488 and DS004302 [59]. The datasets came from various countries around the world
489 (**Extended Data Fig.5**). Details of demographics, geographic location, clinical
490 characteristics, and inclusion/exclusion criteria for each cohort may be found in the
491 **Supplementary Information (Supplementary Table S1-2)**.

492 The severity of symptoms was evaluated by the Positive and Negative Syndrome
493 Scale (PANSS) [60], including a positive scale (total score of P1-P7), a negative scale (total
494 score of N1-N7), a general psychopathology scale (total score of G1-G16) and total score.
495 In addition, phenotypic characteristics were further quantified in three dimensions, such as
496 cognitive (total score of P2, N5, G5, G10, G11), depression/anxiety (total score of G1, G2,
497 G3, G6, G15) and excitement (total score of P4, P7, G44, G14) via a five-factor model of
498 schizophrenia [61].

499 All sites obtained approval from their local institutional review boards or ethics
500 committees, and written informed consent from all participants and/or their legal guardians.
501 The present study was carried out under the approve from the Medical Research Ethics
502 Committees of Fudan University (Number: FE222711).

503

504 **4.2 Image acquisition, processing and quality control**

505 T1-weighted structural brain MRI scans were acquired at each study site. We used a
506 standardized protocol for image processing using the ENIGMA Computational Anatomy
507 Toolbox (CAT12) across multiple cohorts (<https://neuro-jena.github.io/enigma-cat12/>).
508 These protocols enable region-based gray matter volume (GMV) measures for image data
509 based on the automated anatomical (AAL3) atlas [62]. Further details of image acquisition
510 parameters and quality control may be found in **Supplementary Information**
511 (**Supplementary Table S1-2**).

512

513 **4.3 Data harmonization**

514 The ROI-wise GMV measures were first adjusted by regressing out the effects of sex,
515 age, the square of age, site and total intracranial volume (TIV) using a regression model
516 [22]. Subsequently, a harmonization procedure was performed using the ComBat algorithm
517 for correcting multi-site data [63]. The adjusted values were transformed as z-scores (i.e.,
518 normative deviations) relative to the healthy control group. We multiplied these z-scores
519 by -1 so that the z-score increases as regional GMV decreases. Finally, we removed these
520 samples if they were marked as a statistical outlier (>5 standard deviations away from the

521 global mean). After the quality control, 11,260 individuals were included, of which 4,222
522 were schizophrenia patients (1,683 females, mean age= 32.4 ± 12.4 years) and 7,038
523 healthy subjects (3,440 females, mean age= 33.0 ± 12.4 years).

524

525 **4.4 Disease progress modelling**

526 To uncover diverse patterns of pathophysiological progression from cross-sectional
527 only MRI data and cluster individuals into groups (subtypes), we employed a novel
528 machine learning approach - Subtype and Stage Inference (SuStaln) [19]. The
529 methodology of SuStaln has been described in detail previously [19]. Here, we briefly
530 describe the main parameter choices specific to the current study. The SuStaln model
531 requires an $M \times N$ matrix as input. M represents the number of cases ($M=4,222$). N is the
532 number of biomarkers ($N=17$). 17 gray matter biomarkers that were previously used for
533 SuStaln modelling in schizophrenia [22]. Here, all of the AAL3 regions of whole brain were
534 separated and merged into 17 regions of interest (ROIs) [22], including frontal lobe,
535 temporal lobe, parietal lobe, occipital lobe, insula, cingulate, sensorimotor, Broca's area,
536 cerebellum, hippocampus, parahippocampus, amygdala, caudate, putamen, pallidum,
537 accumbens and thalamus ([Supplementary Table S6](#)). We then ran the SuStaln algorithm
538 with 25 start points and 100,000 Markov Chain Monte Carlo (MCMC) iterations [19] to
539 estimate the most likely sequence that describes spatiotemporal pattern of
540 pathophysiological progression (i.e., 'trajectory').

541 SuStaln can identify diverse trajectories of pathophysiological progression given a
542 subtype number K . We fitted the model for $K=2-6$ subtypes ('trajectories'), separately. The
543 optimal number of subtypes was determined according to the reproducibility of individual
544 subtyping via a two-fold cross-validation procedure, as described previously [22].
545 Specifically, all individuals were randomly split into two non-overlapping folds. For each
546 fold, we trained the SuStaln model. For each individual, the trained SuStaln model provides
547 a subtype label. We measured the consistency of the subtype labeling across all individuals
548 between two folds by using the Dice coefficient. This above procedure was repeated ten
549 times. The largest Dice coefficient was obtained for $K=2$ (see [Figure 1a](#)), indicating the
550 best consistency based on cross-validation. Finally, the two-cluster model of SuStaln was
551 fitted to the entire sample. The most probable sequence (i.e., the order of biomarkers) was
552 evaluated for each 'trajectory' via SuStaln. For each individual, SuStaln calculated the
553 probability (ranging from 0 to 1) of belonging to each 'trajectory', and assigned the
554 individual into a sub-stage of the maximum likelihood 'trajectory' through MCMC iterations.
555 We also estimated the SuStaln 'trajectories' based on a subsample from individuals with
556 first-episode schizophrenia whose illness duration was less than two years ($N=1,122$, 513
557 females, mean age= 25.4 ± 8.6 years), and a subsample of medication-naïve individuals
558 with schizophrenia ($N=718$, 353 females, mean age= 23.7 ± 7.8 years).

559

560 **4.5 Visualization of pathophysiological progression trajectory**

561 To visualize the spatiotemporal patterns of pathophysiological progression, we
562 calculated the mean z-score of regional GMV across individuals belonging to the same
563 substage of each SuStaln 'trajectory'. The images of ROI-wise GMV z-scores were

564 mapped into a glass brain template via visualization tools implemented in ENIGMA Toolbox
565 (<https://enigma-toolbox.readthedocs.io/en/latest/index.html>) and BrainNetViewer
566 (<https://www.nitrc.org/projects/bnv/>).

567 To examine whether the SuStaln stage (a continuous indicator of the 'temporal' stage
568 of SuStaln 'trajectory') is associated with pathological processes and clinical
569 characteristics in schizophrenia, we performed Spearman correlations between the
570 SuStaln stages and the degree of brain atrophy (i.e., regional GMV) in schizophrenia. We
571 also examined whether SuStaln stages were linked to disease duration, severity of
572 symptoms, and phenotypic characteristics.

573

574 **4.6 Neuroanatomical signatures using regional morphological measures**

575 To further characterize the neuroanatomical signatures associated with each subtype,
576 we conducted regional morphological analyses in a subsample including 1,840 individuals
577 with schizophrenia and 1,780 healthy controls. Brain morphological measures, such as
578 cortical thickness, cortical surface area, cortical volume and subcortical volume, were
579 quantified using FreeSurfer (version 7.3, <http://surfer.nmr.mgh.harvard.edu/>). A total of
580 68x3 regional measures for cortical thickness, cortical surface area and cortical volume
581 were extracted based on the DK atlas [64], along with 14 subcortical regions (bilaterally
582 nucleus accumbens, amygdala, caudate, hippocampus, pallidum, putamen and thalamus)
583 and 2 lateral ventricles. In addition, we performed an automated subregion segmentation
584 (<https://surfer.nmr.mgh.harvard.edu/fswiki/SubregionSegmentation>) for the hippocampal
585 substructures (n=38 subregions) [65], the nuclei of the amygdala (n=18) [66], the thalamic
586 nuclei (n=50) [67], and the brain stem structures (n=4) [68], yielding a total of 110
587 subregional volumetric measures.

588 Regional morphological measures for each individual with schizophrenia were
589 adjusted by regressing out the effects of sex, age, the square of age, TIV and site, and
590 then transformed to z-scores (i.e., normative deviations from healthy control group). The
591 mean regional morphological z-score across individuals belonging to each subtype was
592 calculated, and mapped to brain templates for visualization of neuroanatomical signature
593 deviation for each subtype relative to healthy population. To further manifest subtype-
594 specific signature in neuroanatomical pathology, we compared the regional morphological
595 z-scores between the two subtypes using two sample *t*-tests. Multiple comparisons were
596 corrected by family wise error (FWE) correction.

597

598 **4.7 Distinct symptom profiles between subtypes**

599 To characterize the psychotic symptomatic trajectory with disease duration increases
600 for each subtype, we further divided the individuals of each subtype into three subgroups
601 according to their illness durations (early stage: <2 years; middle stage: 2-10 years; late
602 stage: >10 years). We compared the difference of symptoms among the three stages of
603 disease in each subtype using ANOVA. Two sample *t*-tests were performed to compare
604 the inter-subtype differences separately between each of the stages.

605 Data availability

606 Data of NMorphCH, FBIRN and NUSDAST were obtained from the SchizConnect, a
607 publicly available website (http://www.schizconnect.org/documentation#by_project). The
608 NMorphCH dataset and NUSDAST dataset were download through a query interface at
609 the SchizConnect (<http://www.schizconnect.org/queries/new>). The FBIRN dataset was
610 download from <https://www.nitrc.org/projects/fbirn/>. The DS000115 dataset was download
611 from OpenfMRI database (<https://www.openfmri.org/>). The DS000030 dataset was
612 available at <https://legacy.openfmri.org/dataset/ds000030/>. The DS004302 dataset was
613 available at <https://openneuro.org/datasets/ds004302/versions/1.0.1>. The HCP-EP
614 dataset was available at [https://www.humanconnectome.org/study/human-connectome-
615 project-for-early-psychosis/](https://www.humanconnectome.org/study/human-connectome-project-for-early-psychosis/). The Japanese SRPBS Multi-disorder MRI Dataset was
616 available at <https://bicr-resource.atr.jp/srpbsopen/>. Requests for ENIGMA data can be
617 applied via the ENIGMA Schizophrenia Working Group
618 (<https://enigma.ini.usc.edu/ongoing/enigma-schizophrenia-working-group/>). Requests for
619 raw and analyzed data can be made to the corresponding author (J.Feng,
620 jffeng@fudan.edu.cn) and will be promptly reviewed by the Fudan University Ethics
621 Committee to verify whether the request is subject to any intellectual property or
622 confidentiality obligations.

623

624 Code availability

625 SuStaln algorithm is available on the UCL-POND GitHub ([https://github.com/ucl-
626 pond/](https://github.com/ucl-pond/)). T1-weighted images were processed using the Computational Anatomy Toolbox
627 for Standardized Processing of ENIGMA Data (<https://neuro-jena.github.io/enigma-cat12/>).
628 A protocol for the current data processing is available at
629 https://docs.google.com/document/d/1lb9v0v4j_OrgAKDh6_9fl3Hz2Wcfg46c/edit/.
630 FreeSurfer (version 7.3, <http://surfer.nmr.mgh.harvard.edu/>) was used to quantify various
631 morphological measures, such as cortical thickness, cortical surface area, cortical volume
632 and subcortical volume. The visualization of ROI-wise z-score images was conducted
633 using BrainNetViewer (<https://www.nitrc.org/projects/bnvl/>).

634

635 Acknowledgements

636 This work was supported by the grant from Science and Technology Innovation 2030-
637 Brain Science and Brain-Inspired Intelligence Project (No. 2022ZD0212800). This work
638 was supported by National Natural Science Foundation of China (No. 82202242,
639 82071997). This work was supported by the projects from China Postdoctoral Science
640 Foundation (No. BX2021078, 2021M700852). This work was supported by the Shanghai
641 Rising-Star Program (No. 21QA1408700) and the Shanghai Sailing Program
642 (22YF1402800) from Shanghai Science and Technology Committee. This work was
643 supported by National Key R&D Program of China (No. 2019YFA0709502), the grant from
644 Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01), ZJ Lab,

645 and Shanghai Center for Brain Science and Brain-Inspired Technology, and the grant from
646 the 111 Project (No. B18015). ASRB: Supported by NHMRC Project Grants (630471 &
647 1081603). ESO: Supported by Ministry of Health of the Czech Republic, grants nr. NU21-
648 08-00432 and NU20-04-00393. Supported by Ministry of Health, Czech Republic - DRO
649 2021 (“Institute for Clinical and Experimental Medicine - IKEM, IN: 00023001”). FOR2107:
650 Funded by the German Research Foundation (Deutsche Forschungsgemeinschaft DFG;
651 Forschungsgruppe/Research Unit FOR2107). Principal investigators (PIs) with respective
652 areas of responsibility in the FOR2107 consortium are: Work Package WP1,
653 FOR2107/MACS cohort and brainimaging: Tilo Kircher (speaker FOR2107; DFG grant
654 numbers KI588/14-1, KI588/14-2, KI588/20-1, KI588/22-1), Udo Dannlowski (co-speaker
655 FOR2107; DA 1151/5-1, DA 1151/5-2, DA1151/6-1), Axel Krug (KR 3822/5-1, KR 3822/7-
656 2), Igor Nenadic (NE2254/1-2, NE2254/3-1, NE2254/4-1, NE2254/2-1), Carsten Konrad
657 (KO 4291/3-1). IMH: Supported by research grants from the National Healthcare Group,
658 Singapore (SIG/05004; SIG/05028), and the Singapore Bioimaging Consortium (RP
659 C009/2006) research grants awarded to Kang Sim. JBUN: Supported by Korean Mental
660 Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea
661 (HL19C0015) and a grant of the Korea Health Technology R&D Project through the Korea
662 Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare,
663 Republic of Korea (HR18C0016). OLIN: Supported by MH106324. Osaka: Supported by
664 JSPS KAKENHI (Grant Number JP21K12153), ABiS (JSPS KAKENHI Grant Number
665 JP22H04926), Brain/MINDS & beyond studies (Grant Number JP18dm0307002) from the
666 AMED. PENS: Supported by department of Veterans Affairs CSR&D (I01CX000227) and
667 National Institutes of Health (U01MH108150) to Scott R. Sponheim. RomeSL: Supported
668 by Italian Ministry of Health, grant Ricerca Corrente RC 23, RF-2019-12370182. SoCAT:
669 Supported by THE SCIENTIFIC AND TECHNOLOGICAL RESEARCH COUNCIL OF
670 TÜRKİYE. SWIFT: Supported by a NARSAD grant from the Brain & Behavior Research
671 Foundation (28445) and by a Research Grant from the Novartis Foundation (20A058).
672 UCISZ: Supported by the National Institute of Mental Health of the National Institutes of
673 Health under award number R21MH097196. UNINA: Supported by
674 #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research
675 (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006) –
676 A Multiscale integrated approach to the study of the nervous system in health and disease
677 (DN. 1553 11.10.2022). COBRE: Supported by NIMH: 5R01MH094524-1;
678 3R01MH121246; 1P20RR021938. TOPSY: Supported by the Canada First Research
679 Excellence Fund, awarded to the Healthy Brains, Healthy Lives initiative at McGill
680 University (through a New Investigator Supplement) and the Monique H. Bourgeois Chair
681 in Developmental Disorders. He receives a salary award from the Fonds de recherche du
682 Québec - Santé (FRQS). OSLO: Supported by The Research Council of Norway (249795,
683 300767), the South-Eastern Norway Regional Health Authority (2014097, 2019101), KG
684 Jepsen Stiftelsen, the European Research Council under the European Union’s Horizon
685 2020 research and Innovation program (802998), and the European Union-funded Horizon
686 Europe project ‘environMENTAL’ (no. 101057429). HCP-EP: Research using Human
687 Connectome Project for Early Psychosis (HCP-EP) data reported in this publication was
688 supported by the National Institute of Mental Health of the National Institutes of Health

689 under Award Number U01MH109977. JP-SRPBS: Data used in the preparation of this
690 work were obtained from the DecNef Project Brain Data Repository ([https://bic-
691 resource.atr.jp/srpbsopen/](https://bic-resource.atr.jp/srpbsopen/)) gathered by a consortium as part of the Japanese Strategic
692 Research Program for the Promotion of Brain Science (SRPBS) supported by the
693 Japanese Advanced Research and Development Programs for Medical Innovation
694 (AMED). FBIRN: Data used for this study were downloaded from the FunctionBIRN Data
695 Repository, supported by grants to the Function BIRN (U24-RR021992) Testbed funded by
696 theNational Center for Research Resources at the National Institutes of Health, U.S.A.
697 MCIC: The imaging data and demographic information was collected and shared by
698 [University of Iowa, University of Minnesota, University of New Mexico, Massachusetts
699 General Hospital] the Mind Research Network supported by the Department of Energy
700 under Award Number DE-FG02-08ER64581. NMorphCH: Data collection and sharing was
701 funded by NIMH grant R01 MH056584. NUSDAST: Data collection and sharing was funded
702 by NIMH grant 1R01 MH084803. This work is supported by the Zhangjiang International
703 Brain Biobank (ZIB) Consortium.

704

705 **Competing Interests Statement**

706 Lena Palaniyappan reports personal fees from Janssen Canada, Otsuka Canada,
707 SPMM Course Limited UK and the Canadian Psychiatric Association; book royalties from
708 Oxford University Press; and investigator-initiated educational grants from Sunovion,
709 Janssen Canada and Otsuka Canada, outside the submitted work. Tilo Kircher received
710 unrestricted educational grants from Servier, Janssen, Recordati, Aristo, Otsuka,
711 neuraxpharm. Philipp Homan has received grants and honoraria from Novartis, Lundbeck,
712 Mepha, Janssen, Boehringer Ingelheim, Neurolite outside of this work. Ole A. Andreassen
713 is a consultant to Cortechs.ai and received speakers honorarium from Lundbeck, Janssen,
714 Sunovion. These interests played no role in the research reported here. Other authors
715 disclose no conflict of interest.

Table 1. Demographic and clinical characteristics in the primary sample including 4,222 schizophrenia patients and 7,038 healthy controls.

	HC(n=7,038)		SCZ(n=4,222)		SCZ subtype1(n=2,622)		SCZ subtype2(n=1,600)	
	n	mean(SD)	n	mean(SD)	n	mean(SD)	n	mean(SD)
Sex (Female/Male)	3440/3598	-	1683/2539	-	1044/1578	-	639/961	-
Age (years)	7038	33.0(12.6)	4222	32.4(11.9)	2622	32.4(11.8)	1600	32.4(12.0)
Illness duration (years)	-	-	2333	10.5(10.4)	1442	10.4(10.5)	891	10.5(10.4)
FES/Chronic/Unknown	-	-	1112/1623/1477	-	696/1002/924	-	426/621/553	-
PANSS Positive scale (P1-P7)	-	-	2651	17.2(6.8)	1622	17.3(3.9)	1029	17.0(6.7)
PANSS Negative scale (N1-N7)	-	-	2651	17.5(7.6)	1622	17.6(7.6)	1029	17.3(7.6)
PANSS General scale (G1-G16)	-	-	2651	34.8(11.6)	1622	35.2(11.6)	1029	34.3(11.6)
PANSS Total score	-	-	2651	69.5(22.4)	1622	70.0(22.4)	1029	68.6(22.5)
PANSS excitement dimension (P4, P7, G44, G14)	-	-	1322	8.2(3.5)	823	8.2(3.4)	499	8.2(3.5)
PANSS depression/anxiety dimension (G1, G2, G3, G6, G15)	-	-	1322	11.3(4.1)	823	11.4(4.1)	499	11.1(4.2)
PANSS cognitive dimension (P2, N5, G5, G10, G11)	-	-	1322	10.6(4.0)	823	10.5(4.0)	499	10.6(4.0)

Abbreviation: HC, healthy control; SCZ, schizophrenia; FES, first-episode schizophrenia; PANSS, Positive and Negative Syndrome Scale.

716

717

Table 2. Main findings of subtype-specific neuroanatomical signatures.

Morphometry measures	Subtype-specific neuroanatomical signatures
Cortical Thickness/Volume/Area	<p>a) Both subtype1 and subtype2 exhibit a moderate degree in the average cortical thickness reduction.</p> <p>b) Reduction of average cortical volume/area is only observed in the subtype1.</p> <p>c) The worst reduction of cortical thickness/volume/area is located within the superior frontal regions for the subtype1, but in the superior/medial temporal regions for the subtype2.</p>
Subcortical Volume	<p>a) Enlargement of lateral ventricle is found in both subtype1 and subtype2, but much larger in the subtype2.</p> <p>b) Worse loss volumes of hippocampus, amygdala, thalamus and accumbens are observed in the subtype2, compared to the subtype1.</p> <p>c) Volumes of striatum (i.e., caudate, putamen) are increased in the subtype1, but decreased in the subtype2, compared to the healthy population.</p>
Hippocampus segmentation	<p>a) Volume loss in hippocampal subregions is worse in the subtype2, compared to the subtype1.</p> <p>b) The most significant volume loss is in the molecular layer for the subtype2.</p>
Amygdala segmentation	<p>a) The subtype2 shows worse volume loss in amygdala subregions, compared to the subtype1.</p> <p>b) The most significant decrease in volume is in the cortico-amygdaloid transition area for both the subtypes.</p>
Thalamus segmentation	<p>a) The subtype2 shows worse volume loss in thalamus subregions, compared to the subtype1.</p>
Brain stem segmentation	<p>a) Volume loss of brain stem subregions is only observed in the subtype2.</p>

718

719

Table 3. Symptom scores for each subtype at different stages of disease duration.

Symptoms	Subtype 1			F test		Subtype 2			F test	
	Early	Middle	Late	F	P	Early	Middle	Late	F	P
	(n=579)	(n=362)	(n=400)			(n=371)	(n=216)	(n=282)		
PANSS Positive scale (P1-P7)	19.5(6.4)	16.0(6.7)	16.7(7.0)*	37.4	1.60E-16	19.6(6.4)	16.0(6.7)	15.2(6.2)*	41.9	4.68E-18
PANSS Negative scale (N1-N7)	16.8(7.3)	17.4(7.4)	18.3(7.7)	4.6	9.98E-03	17.3(7.4)	17.1(7.5)	17.4(7.5)	0.1	0.884
PANSS General scale (G1-G16)	37.6(10.0)	34.1(12.2)	35.3(13.3)*	10.6	2.80E-05	37.7(10.8)	33.7(12.4)	32.7(12.0)*	15.6	2.30E-07
PANSS Total score	73.9(19.7)	67.5(23.2)	70.2(25.0)*	9.3	9.40E-05	74.5(20.9)	66.7(23.4)	65.4(22.4)*	15.7	2.05E-07
PANSS excitement dimension (P4, P7, G44, G14)	8.8(3.4)	7.9(3.3)	8.2(3.4)	4.9	8.01E-03	8.7(3.3)	7.8(3.4)	7.8(3.7)	4.12	0.017
PANSS depression/anxiety dimension (G1, G2, G3, G6, G15)	11.2(3.7)	11.7(4.2)	12.5(4.9)*	5.9	2.86E-03	11.4(4.0)	10.8(4.4)	11.3(4.9)*	0.7	0.511
PANSS cognitive dimension (P2, N5, G5, G10, G11)	10.1(3.7)	10.8(4.0)	12.0(4.6)	13.5	1.74E-06	10.3(3.8)	10.9(3.9)	11.4(4.7)	2.8	0.061

* indicates significant difference between the subtype1 and subtype2 using two sample t test (p<0.05).

721 **Figures Legend**

722 **Figure 1. Two pathophysiological progression trajectories in schizophrenia.** (a) Dice coefficient
723 indicates that K=2 is the optimal number of subtypes with best consistency of the subtype labeling
724 between two independent schizophrenia populations using non-overlap 2-folds cross-validation
725 procedure. Data are presented as mean values +/- SD. (b) The proportion of individuals whose subtype
726 labels keep consistent by non-overlap cross-validation procedure. (c) Sequences of regional volume
727 loss across seventeen brain regions for each 'trajectory' via SuStaln are shown in y-axis. The heatmap
728 shows regional volume loss in which biomarker (y-axis) in a particular 'temporal' stage (T0-T16) in the
729 trajectory (x-axis). The Color bar represents the degree of gray matter volume (GMV) loss in
730 schizophrenia relative to healthy controls (i.e., z score). (d) Spatiotemporal pattern of pathophysiological
731 'trajectory'. The z-score images are mapped to a glass brain template for visualization. Spatiotemporal
732 pattern of gray matter loss displays a progressive pattern of spatial extension along with later 'temporal'
733 stages of pathological progression, that is distinct between trajectories. (e-f) Pathological stages of
734 SuStaln are correlated with reduced gray matter volume of Broca's area and hippocampus. (g-i)
735 Pathological stages of SuStaln are correlated with longer disease duration, worse negative symptoms
736 and worse cognitive symptoms.

737 **Figure 2. Trajectories are reproducibility for samples from different locations of the world.** Two
738 sets of 'trajectories' are separately derived from two non-overlapping location cohorts, that are (a) East
739 Asian ancestry (EAS) cohort, and (b) European ancestry (EUR) cohort. The Color bar represents the
740 degree of gray matter volume (GMV) loss in schizophrenia relative to healthy controls (i.e., z score). (c)
741 The similarity of the spatiotemporal pattern of each 'trajectory' between any two of cohorts is shown by
742 the heatmap. The color bar of the heatmap represents the similarity, which is quantified via the
743 Spearman correlation coefficient between the trajectories from two cohorts. A total of six location cohorts
744 are classified by where the sample locate at, including the EAS, EUR, China, Japan, Europe and North
745 America. The whole sample is labelled as a cross-ancestry cohort.

746 **Figure 3. Subtype-specific signatures in neuroanatomical pathology.** Regional Morphological z-
747 scores (i.e., normative deviations from healthy control group) for each subtype are mapped to a brain
748 template for visualization. Effect size of inter-subtype difference is quantified using Cohen's d.

749 **Figure 4. Symptomatic trajectories across three stages of disease duration.** Individuals of each
750 subtype are divided into three subgroups according to their illness durations (early stage: ≤ 2 years;
751 middle stage: 2-10 years; late stage: > 10 years). Data are presented as mean values +/- se. * $p < 0.05$.

752 **Extend Data Fig 1. Pathophysiological progression trajectories in first-episode population and**
753 **medication-naïve population.** Trajectories are repeated based on the subsample data from the first-
754 episode schizophrenia patients whose illness duration was less than two years (N=1,112, 513 females,
755 mean age= 25.4 ± 12.4 years), and another subsample data from medication-naïve patients with
756 schizophrenia (N=718, 353 females, mean age= 23.7 ± 12.1 years).

757 **Extend Data Fig 2. Comparisons of morphological z-score between the two subtypes.** A larger
758 positive z-score indicates a larger deviation of reduction relative to healthy control group. Two sample t
759 test is conducted to examine inter-subtype difference for the (a) averaged cortical volume ($t=9.36$,
760 $p < 10e-16$, Cohen's $d=0.446$); (b) averaged cortical area ($t=8.09$, $p < 10e-16$, Cohen's $d=0.386$); (c)

761 averaged cortical thickness ($t=1.29$, $p=0.198$, Cohen's $d=0.061$); (d) thalamus volume ($t=-4.28$, $p=1.97e-$
762 5 , Cohen's $d=-0.205$); (e) brain stem volume ($t=-9.79$, $p<10e-16$, Cohen's $d=-0.469$); (f) hippocampus
763 volume ($t=-9.25$, $p<10e-16$, Cohen's $d=-0.449$); (g) amygdala volume ($t=-7.83$, $p=8.44e-15$, Cohen's $d=-$
764 0.379); (h) accumbens volume ($t=-6.40$, $p=1.94e-10$, Cohen's $d=-0.305$); (i) caudate volume ($t=-9.82$,
765 $p<10e-16$, Cohen's $d=-0.468$); (j) putamen volume ($t=-8.14$, $p<10e-16$, Cohen's $d=-0.389$).

766 **Extend Data Fig 3. Hippocampus subregional morphological z-score for the two subtypes.** A

767 larger positive z-score indicates a larger deviation of reduction relative to healthy control group.

768 **Extend Data Fig 4. Amygdala subregional morphological z-score for the two subtypes.** A larger

769 positive z-score indicates a larger deviation of reduction relative to healthy control group.

770 **Extend Data Fig 5. Geographic map of included datasets.**

771 References

- 772 1. The, L., *ICD-11: a brave attempt at classifying a new world*. The Lancet, 2018.
773 **391**(10139): p. 2476.
- 774 2. Oren, O., B.J. Gersh, and D.L. Bhatt, *Artificial intelligence in medical*
775 *imaging: switching from radiographic pathological data to clinically*
776 *meaningful endpoints*. Lancet Digit Health, 2020. **2**(9): p. e486–e488.
- 777 3. Rajpurkar, P., et al., *AI in health and medicine*. Nat Med, 2022. **28**(1): p.
778 31–38.
- 779 4. Organization, W.H., *The global burden of disease: 2004 update*. 2008: World
780 Health Organization.
- 781 5. Howes, O.D. and E.C. Onwordi, *The synaptic hypothesis of schizophrenia version*
782 *III: a master mechanism*. Mol Psychiatry, 2023.
- 783 6. McCutcheon, R.A., J.H. Krystal, and O.D. Howes, *Dopamine and glutamate in*
784 *schizophrenia: biology, symptoms and treatment*. World Psychiatry, 2020. **19**(1):
785 p. 15–33.
- 786 7. Wolfers, T., et al., *Mapping the Heterogeneous Phenotype of Schizophrenia and*
787 *Bipolar Disorder Using Normative Models*. JAMA Psychiatry, 2018. **75**(11): p.
788 1146–1155.
- 789 8. Fusar-Poli, P., et al., *Heterogeneity of Psychosis Risk Within Individuals*
790 *at Clinical High Risk: A Meta-analytical Stratification*. JAMA Psychiatry,
791 2016. **73**(2): p. 113–20.
- 792 9. McCutcheon, R.A., et al., *The efficacy and heterogeneity of antipsychotic*
793 *response in schizophrenia: A meta-analysis*. Mol Psychiatry, 2021. **26**(4): p.
794 1310–1320.
- 795 10. Collado-Torres, L., et al., *Regional Heterogeneity in Gene Expression,*
796 *Regulation, and Coherence in the Frontal Cortex and Hippocampus across*
797 *Development and Schizophrenia*. Neuron, 2019. **103**(2): p. 203–216 e8.
- 798 11. Brugger, S.P. and O.D. Howes, *Heterogeneity and Homogeneity of Regional Brain*
799 *Structure in Schizophrenia: A Meta-analysis*. JAMA Psychiatry, 2017. **74**(11):
800 p. 1104–1111.
- 801 12. Braff, D.L., et al., *Lack of use in the literature from the last 20 years*
802 *supports dropping traditional schizophrenia subtypes from DSM-5 and ICD-11*.
803 Schizophr Bull, 2013. **39**(4): p. 751–3.
- 804 13. Wen, J., et al., *Multi-scale semi-supervised clustering of brain images:*
805 *Deriving disease subtypes*. Med Image Anal, 2022. **75**: p. 102304.
- 806 14. Lalouis, P.A., et al., *Heterogeneity and Classification of Recent Onset*
807 *Psychosis and Depression: A Multimodal Machine Learning Approach*. Schizophr
808 Bull, 2021. **47**(4): p. 1130–1140.
- 809 15. Chand, G.B., et al., *Two distinct neuroanatomical subtypes of schizophrenia*
810 *revealed using machine learning*. Brain, 2020. **143**(3): p. 1027–1038.
- 811 16. Yang, Z., et al., *A deep learning framework identifies dimensional*
812 *representations of Alzheimer’s Disease from brain structure*. Nat Commun, 2021.
813 **12**(1): p. 7065.

- 814 17. Dwyer, D.B., et al., *Brain Subtyping Enhances The Neuroanatomical*
815 *Discrimination of Schizophrenia*. Schizophr Bull, 2018. **44**(5): p. 1060-1069.
- 816 18. Luo, C., et al., *Subtypes of schizophrenia identified by multi-omic measures*
817 *associated with dysregulated immune function*. Mol Psychiatry, 2021. **26**(11):
818 p. 6926-6936.
- 819 19. Young, A.L., et al., *Uncovering the heterogeneity and temporal complexity of*
820 *neurodegenerative diseases with Subtype and Stage Inference*. Nat Commun, 2018.
821 **9**(1): p. 4273.
- 822 20. Vogel, J.W., et al., *Four distinct trajectories of tau deposition identified*
823 *in Alzheimer's disease*. Nat Med, 2021. **27**(5): p. 871-881.
- 824 21. Young, A.L., et al., *Characterizing the Clinical Features and Atrophy Patterns*
825 *of MAPT-Related Frontotemporal Dementia With Disease Progression Modeling*.
826 Neurology, 2021. **97**(9): p. e941-e952.
- 827 22. Jiang, Y., et al., *Neuroimaging biomarkers define neurophysiological subtypes*
828 *with distinct trajectories in schizophrenia*. Nature Mental Health, 2023. **1**(3):
829 p. 186-199.
- 830 23. van Erp, T.G.M., et al., *Cortical Brain Abnormalities in 4474 Individuals*
831 *With Schizophrenia and 5098 Control Subjects via the Enhancing Neuro Imaging*
832 *Genetics Through Meta Analysis (ENIGMA) Consortium*. Biol Psychiatry, 2018.
833 **84**(9): p. 644-654.
- 834 24. van Erp, T.G., et al., *Subcortical brain volume abnormalities in 2028*
835 *individuals with schizophrenia and 2540 healthy controls via the ENIGMA*
836 *consortium*. Mol Psychiatry, 2016. **21**(4): p. 585.
- 837 25. Okada, N., et al., *Subcortical volumetric alterations in four major*
838 *psychiatric disorders: a mega-analysis study of 5604 subjects and a volumetric*
839 *data-driven approach for classification*. Mol Psychiatry, 2023.
- 840 26. Koshiyama, D., et al., *White matter microstructural alterations across four*
841 *major psychiatric disorders: mega-analysis study in 2937 individuals*. Mol
842 Psychiatry, 2020. **25**(4): p. 883-895.
- 843 27. Howes, O.D., et al., *Neuroimaging in schizophrenia: an overview of findings*
844 *and their implications for synaptic changes*. Neuropsychopharmacology, 2023.
845 **48**(1): p. 151-167.
- 846 28. Alnaes, D., et al., *Brain Heterogeneity in Schizophrenia and Its Association*
847 *With Polygenic Risk*. JAMA Psychiatry, 2019. **76**(7): p. 739-748.
- 848 29. Howes, O.D. and S. Kapur, *A neurobiological hypothesis for the classification*
849 *of schizophrenia: type A (hyperdopaminergic) and type B (normodopaminergic)*.
850 Br J Psychiatry, 2014. **205**(1): p. 1-3.
- 851 30. Jiang, Y., et al., *Progressive Reduction in Gray Matter in Patients with*
852 *Schizophrenia Assessed with MR Imaging by Using Causal Network Analysis*.
853 Radiology, 2018. **287**(2): p. 729.
- 854 31. Kirschner, M., et al., *Orbitofrontal-Striatal Structural Alterations Linked*
855 *to Negative Symptoms at Different Stages of the Schizophrenia Spectrum*.
856 Schizophr Bull, 2021. **47**(3): p. 849-863.
- 857 32. Thompson, P.M., et al., *Mapping adolescent brain change reveals dynamic wave*

- 858 *of accelerated gray matter loss in very early-onset schizophrenia*. Proc Natl
859 Acad Sci U S A, 2001. **98**(20): p. 11650-5.
- 860 33. Thompson, P.M., et al., *Time-lapse mapping of cortical changes in*
861 *schizophrenia with different treatments*. Cereb Cortex, 2009. **19**(5): p. 1107-
862 23.
- 863 34. Fillman, S.G., et al., *Elevated peripheral cytokines characterize a subgroup*
864 *of people with schizophrenia displaying poor verbal fluency and reduced*
865 *Broca's area volume*. Mol Psychiatry, 2016. **21**(8): p. 1090-8.
- 866 35. Crow, T.J., *Is schizophrenia the price that Homo sapiens pays for language?*
867 Schizophr Res, 1997. **28**(2-3): p. 127-41.
- 868 36. Palaniyappan, L. and P.F. Liddle, *Does the salience network play a cardinal*
869 *role in psychosis? An emerging hypothesis of insular dysfunction*. J Psychiatry
870 Neurosci, 2012. **37**(1): p. 17-27.
- 871 37. Del Re, E.C., et al., *Baseline Cortical Thickness Reductions in Clinical High*
872 *Risk for Psychosis: Brain Regions Associated with Conversion to Psychosis*
873 *Versus Non-Conversion as Assessed at One-Year Follow-Up in the Shanghai-At-*
874 *Risk-for-Psychosis (SHARP) Study*. Schizophr Bull, 2021. **47**(2): p. 562-574.
- 875 38. Pantelis, C., et al., *Neuroanatomical abnormalities before and after onset*
876 *of psychosis: a cross-sectional and longitudinal MRI comparison*. Lancet, 2003.
877 **361**(9354): p. 281-8.
- 878 39. Slifstein, M., et al., *Deficits in prefrontal cortical and extrastriatal*
879 *dopamine release in schizophrenia: a positron emission tomographic functional*
880 *magnetic resonance imaging study*. JAMA Psychiatry, 2015. **72**(4): p. 316-24.
- 881 40. Steen, R.G., et al., *Brain volume in first-episode schizophrenia: systematic*
882 *review and meta-analysis of magnetic resonance imaging studies*. Br J
883 Psychiatry, 2006. **188**: p. 510-8.
- 884 41. van Erp, T.G., et al., *Subcortical brain volume abnormalities in 2028*
885 *individuals with schizophrenia and 2540 healthy controls via the ENIGMA*
886 *consortium*. Mol Psychiatry, 2016. **21**(4): p. 547-53.
- 887 42. Balu, D.T., et al., *Multiple risk pathways for schizophrenia converge in*
888 *serine racemase knockout mice, a mouse model of NMDA receptor hypofunction*.
889 Proc Natl Acad Sci U S A, 2013. **110**(26): p. E2400-9.
- 890 43. McCutcheon, R.A., T. Reis Marques, and O.D. Howes, *Schizophrenia—An Overview*.
891 JAMA Psychiatry, 2020. **77**(2): p. 201-210.
- 892 44. Brugger, S.P., et al., *Heterogeneity of Striatal Dopamine Function in*
893 *Schizophrenia: Meta-analysis of Variance*. Biol Psychiatry, 2020. **87**(3): p.
894 215-224.
- 895 45. Banaj, N., et al., *Cortical morphology in patients with the deficit and non-*
896 *deficit syndrome of schizophrenia: a worldwide meta- and mega-analyses*. Mol
897 Psychiatry, 2023.
- 898 46. Mouchlianitis, E., R. McCutcheon, and O.D. Howes, *Brain-imaging studies of*
899 *treatment-resistant schizophrenia: a systematic review*. Lancet Psychiatry,
900 2016. **3**(5): p. 451-63.
- 901 47. Jiang, Y., et al., *Structural and Functional MRI Brain Changes in Patients*

- 902 *with Schizophrenia Following Electroconvulsive Therapy: A Systematic Review.*
903 *Curr Neuropsychopharmacol*, 2022. **20**(6): p. 1241-1252.
- 904 48. Wang, J., et al., *ECT-induced brain plasticity correlates with positive*
905 *symptom improvement in schizophrenia by voxel-based morphometry analysis of*
906 *grey matter.* *Brain Stimul*, 2019. **12**(2): p. 319-328.
- 907 49. Jiang, Y., et al., *Insular changes induced by electroconvulsive therapy*
908 *response to symptom improvements in schizophrenia.* *Prog Neuropsychopharmacol*
909 *Biol Psychiatry*, 2019. **89**: p. 254-262.
- 910 50. Ho, B.C., et al., *Long-term antipsychotic treatment and brain volumes: a*
911 *longitudinal study of first-episode schizophrenia.* *Arch Gen Psychiatry*, 2011.
912 **68**(2): p. 128-37.
- 913 51. Lewandowski, K.E., et al., *Neuroprogression across the Early Course of*
914 *Psychosis.* *J Psychiatr Brain Sci*, 2020. **5**.
- 915 52. Tanaka, S.C., et al., *A multi-site, multi-disorder resting-state magnetic*
916 *resonance image database.* *Sci Data*, 2021. **8**(1): p. 227.
- 917 53. Keator, D.B., et al., *The Function Biomedical Informatics Research Network*
918 *Data Repository.* *Neuroimage*, 2016. **124**(Pt B): p. 1074-1079.
- 919 54. Gollub, R.L., et al., *The MCIC collection: a shared repository of multi-modal,*
920 *multi-site brain image data from a clinical investigation of schizophrenia.*
921 *Neuroinformatics*, 2013. **11**(3): p. 367-88.
- 922 55. Alpert, K., et al., *The Northwestern University Neuroimaging Data Archive*
923 *(NUNDA).* *Neuroimage*, 2016. **124**(Pt B): p. 1131-1136.
- 924 56. Kogan, A., et al., *Northwestern University schizophrenia data sharing for*
925 *SchizConnect: A longitudinal dataset for large-scale integration.* *Neuroimage*,
926 2016. **124**(Pt B): p. 1196-1201.
- 927 57. Poldrack, R.A., et al., *A phenome-wide examination of neural and cognitive*
928 *function.* *Sci Data*, 2016. **3**: p. 160110.
- 929 58. Repovs, G. and D.M. Barch, *Working memory related brain network connectivity*
930 *in individuals with schizophrenia and their siblings.* *Front Hum Neurosci*,
931 2012. **6**: p. 137.
- 932 59. Soler-Vidal, J., et al., *Brain correlates of speech perception in*
933 *schizophrenia patients with and without auditory hallucinations.* *PLOS ONE*,
934 2022. **17**(12): p. e0276975.
- 935 60. Kay, S.R., A. Fiszbein, and L.A. Opler, *The positive and negative syndrome*
936 *scale (PANSS) for schizophrenia.* *Schizophr Bull*, 1987. **13**(2): p. 261-76.
- 937 61. Lindenmayer, J.P., R. Bernstein-Hyman, and S. Grochowski, *Five-factor model*
938 *of schizophrenia. Initial validation.* *J Nerv Ment Dis*, 1994. **182**(11): p. 631-
939 8.
- 940 62. Rolls, E.T., et al., *Automated anatomical labelling atlas 3.* *Neuroimage*, 2020.
941 **206**: p. 116189.
- 942 63. Pomponio, R., et al., *Harmonization of large MRI datasets for the analysis*
943 *of brain imaging patterns throughout the lifespan.* *Neuroimage*, 2020. **208**: p.
944 116450.
- 945 64. Desikan, R.S., et al., *An automated labeling system for subdividing the human*

- 946 *cerebral cortex on MRI scans into gyral based regions of interest*. Neuroimage,
947 2006. **31**(3): p. 968–80.
- 948 65. Iglesias, J.E., et al., *A computational atlas of the hippocampal formation*
949 *using ex vivo, ultra-high resolution MRI: Application to adaptive segmentation*
950 *of in vivo MRI*. Neuroimage, 2015. **115**: p. 117–37.
- 951 66. Saygin, Z.M., et al., *High-resolution magnetic resonance imaging reveals*
952 *nuclei of the human amygdala: manual segmentation to automatic atlas*.
953 Neuroimage, 2017. **155**: p. 370–382.
- 954 67. Iglesias, J.E., et al., *A probabilistic atlas of the human thalamic nuclei*
955 *combining ex vivo MRI and histology*. Neuroimage, 2018. **183**: p. 314–326.
- 956 68. Iglesias, J.E., et al., *Bayesian segmentation of brainstem structures in MRI*.
957 Neuroimage, 2015. **113**: p. 184–95.
- 958

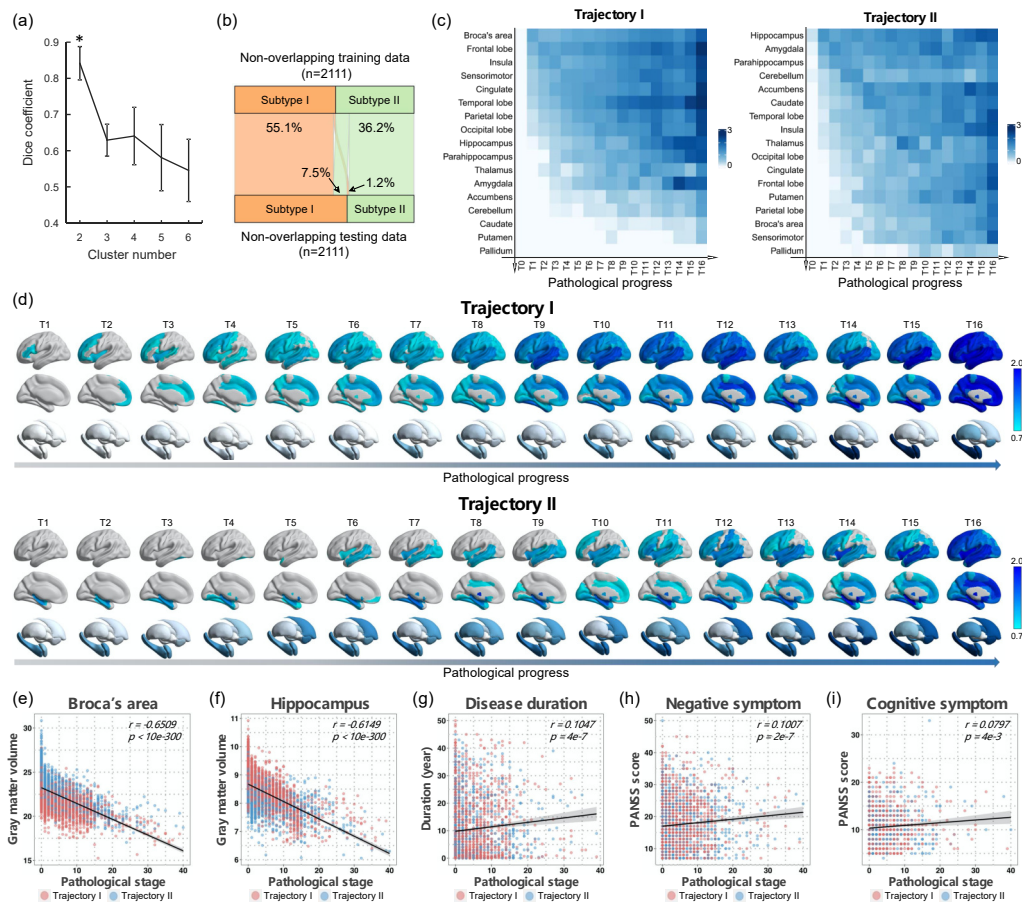
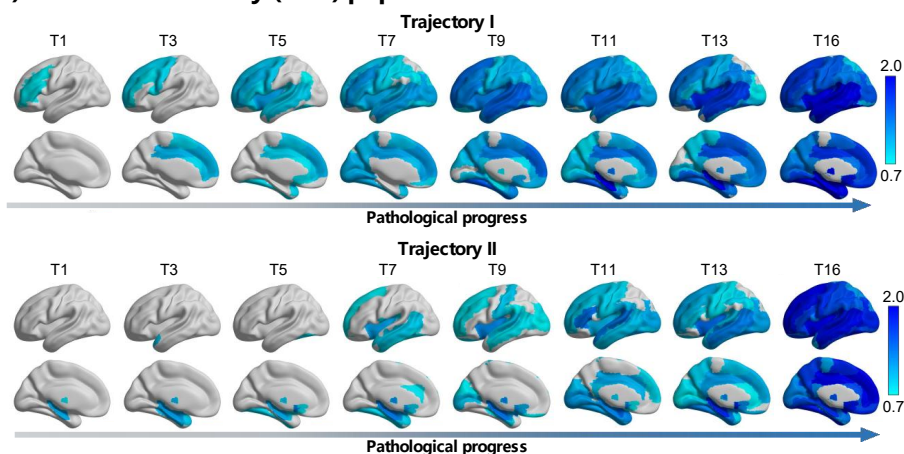
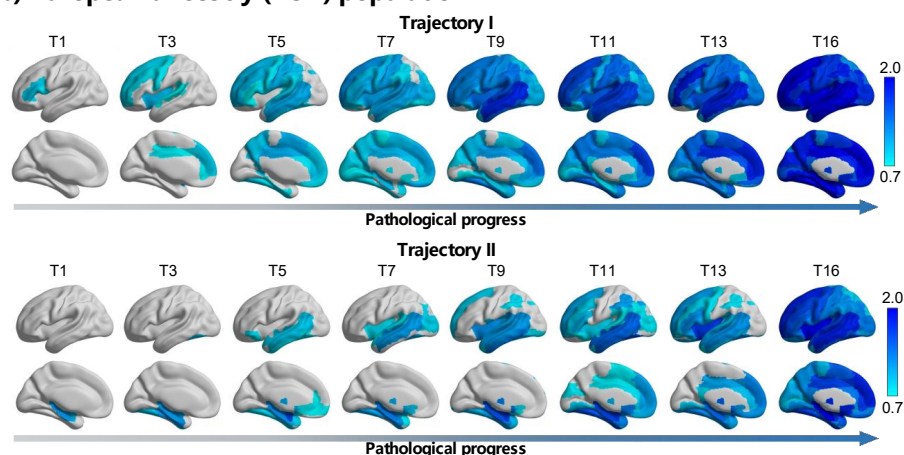


Figure 1. Two pathophysiological progression trajectories in schizophrenia. (a) Dice coefficient indicates that K=2 is the optimal number of subtypes with best consistency of the subtype labeling between two independent schizophrenia populations using non-overlap 2-folds cross-validation procedure. Data are presented as mean values +/- SD. (b) The proportion of individuals whose subtype labels keep consistent by non-overlap cross-validation procedure. (c) Sequences of regional volume loss across seventeen brain regions for each 'trajectory' via SuStaln are shown in y-axis. The heatmap shows regional volume loss in which biomarker (y-axis) in a particular 'temporal' stage (T0-T16) in the trajectory (x-axis). The Color bar represents the degree of gray matter volume (GMV) loss in schizophrenia relative to healthy controls (i.e., z score). (d) Spatiotemporal pattern of pathophysiological 'trajectory'. The z-score images are mapped to a glass brain template for visualization. Spatiotemporal pattern of gray matter loss displays a progressive pattern of spatial extension along with later 'temporal' stages of pathological progression, that is distinct between trajectories. (e-f) Pathological stages of SuStaln are correlated with reduced gray matter volume of Broca's area and hippocampus. (g-i) Pathological stages of SuStaln are correlated with longer disease duration, worse negative symptoms and worse cognitive symptoms.

(a) East Asian ancestry (EAS) population



(b) European ancestry (EUR) population



(c) Similarity of the trajectories among people from different parts of the world

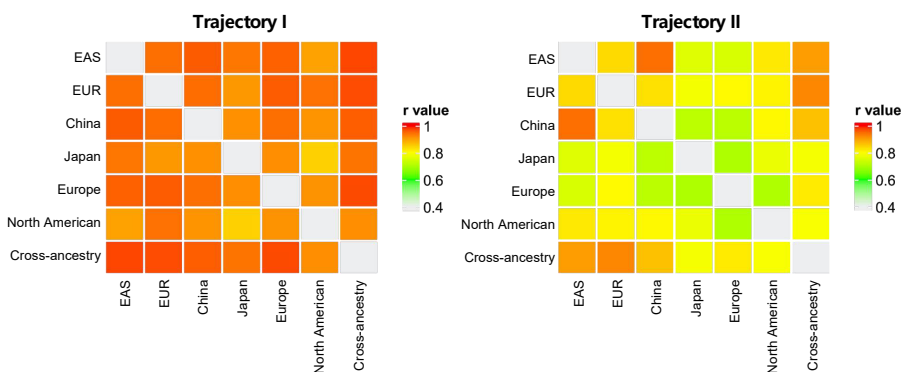
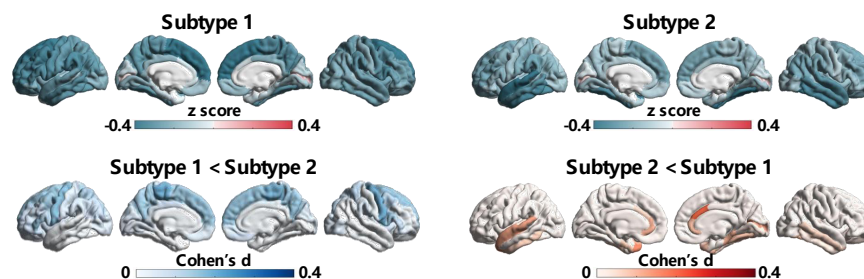


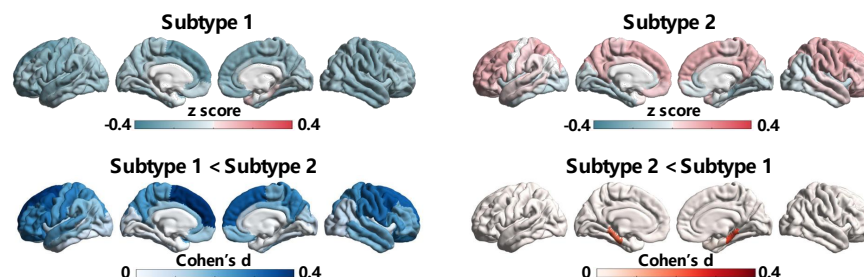
Figure 2. Trajectories are reproducibility for samples from different locations of the world. Two sets of ‘trajectories’ are separately derived from two non-overlapping location cohorts, that are (a) East Asian ancestry (EAS) cohort, and (b) European ancestry (EUR) cohort. The Color bar represents the degree of gray matter volume (GMV) loss in schizophrenia relative to healthy controls (i.e., z score). (c) The similarity of the

spatiotemporal pattern of each 'trajectory' between any two of cohorts is shown by the heatmap. The color bar of the heatmap represents the similarity, which is quantified via the Spearman correlation coefficient between the trajectories from two cohorts. A total of six location cohorts are classified by where the sample locate at, including the EAS, EUR, China, Japan, Europe and North American. The whole sample is labelled as a cross-ancestry cohort.

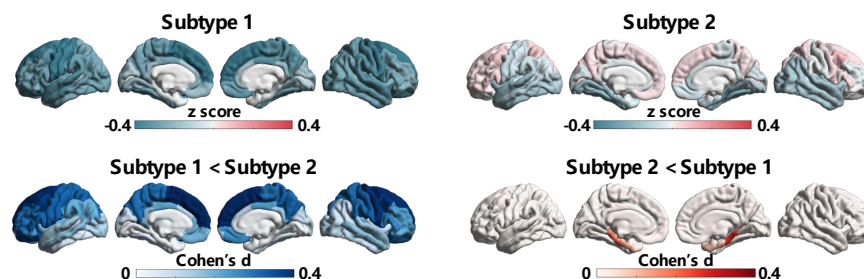
(a) Cortical Thickness



(b) Cortical Area



(c) Cortical Volume



(d) Subcortical Volume

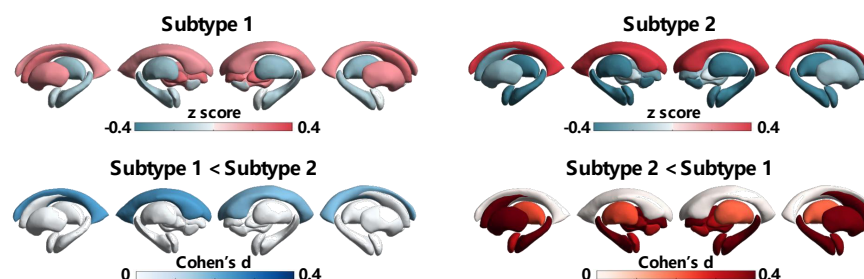


Figure 3. Subtype-specific signatures in neuroanatomical pathology. Regional Morphological z-scores (i.e., normative deviations from healthy control group) for each subtype are mapped to a brain template for visualization. Effect size of inter-subtype difference is quantified using Cohen's d.

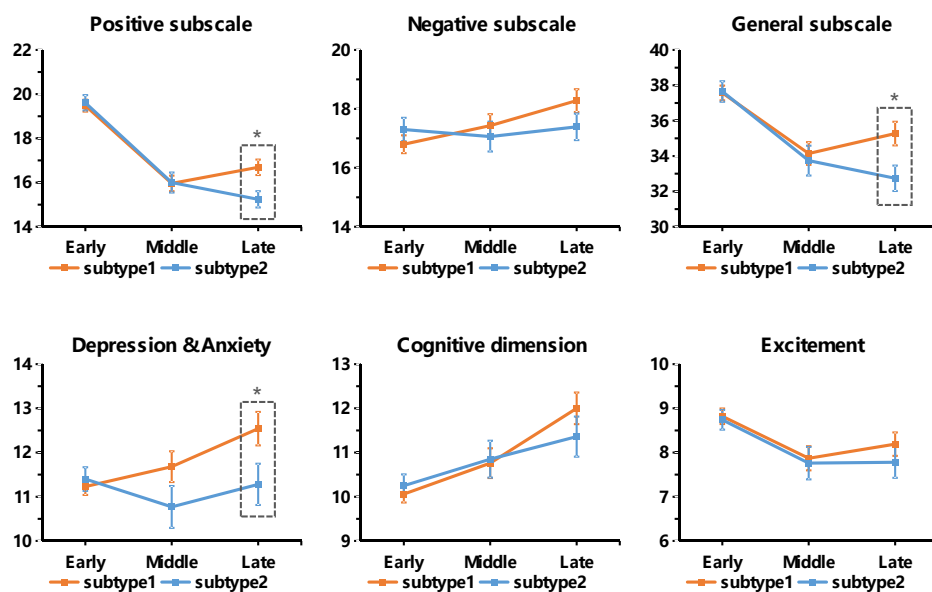
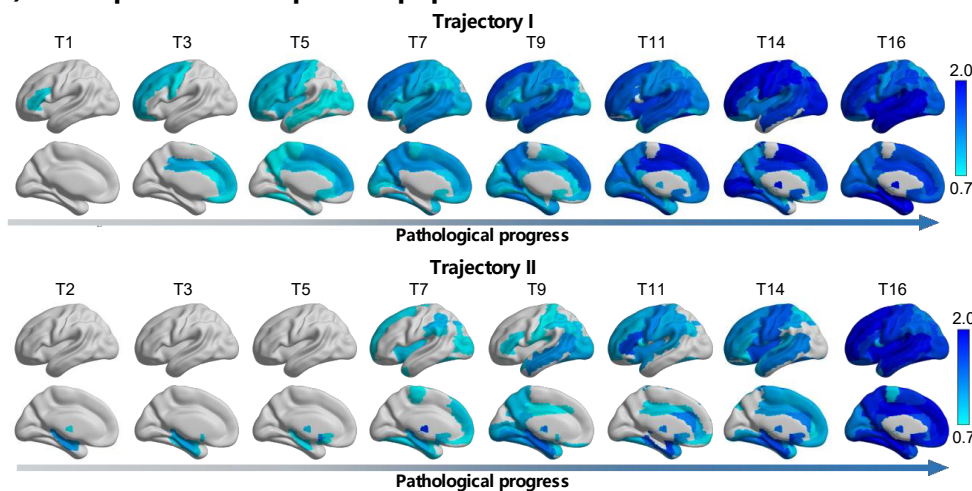


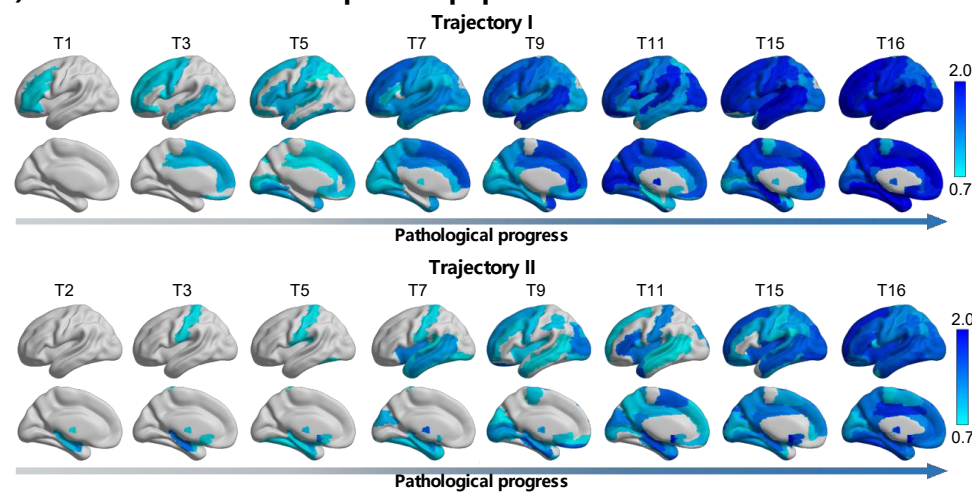
Figure 4. Symptomatic trajectories across three stages of disease duration.

Individuals of each subtype are divided into three subgroups according to their illness durations (early stage: ≤ 2 years; middle stage: 2-10 years; late stage: > 10 years). Data are presented as mean values \pm se. * $p < 0.05$.

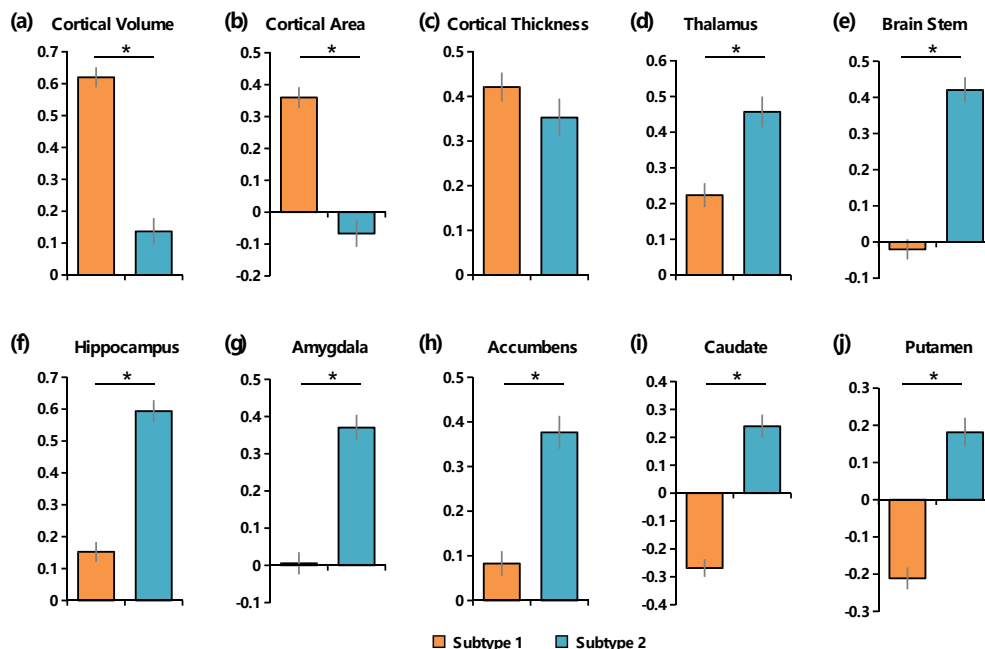
(a) First-episode schizophrenia population



(b) Medication-naïve schizophrenia population

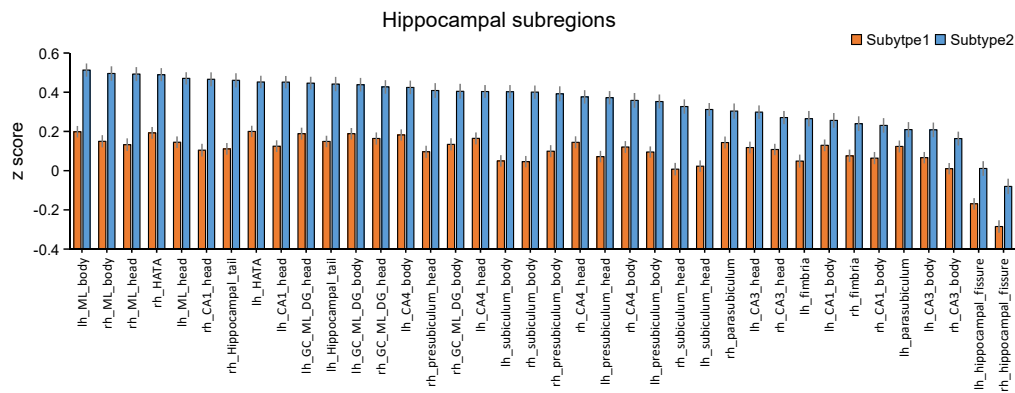


Extend Data Fig 1. Pathophysiological progression trajectories in first-episode population and medication-naïve population. Trajectories are repeated based on the subsample data from the first-episode schizophrenia patients whose illness duration was less than two years ($N=1,112$, 513 females, mean age= 25.4 ± 12.4 years), and another subsample data from medication-naïve patients with schizophrenia ($N=718$, 353 females, mean age= 23.7 ± 12.1 years).

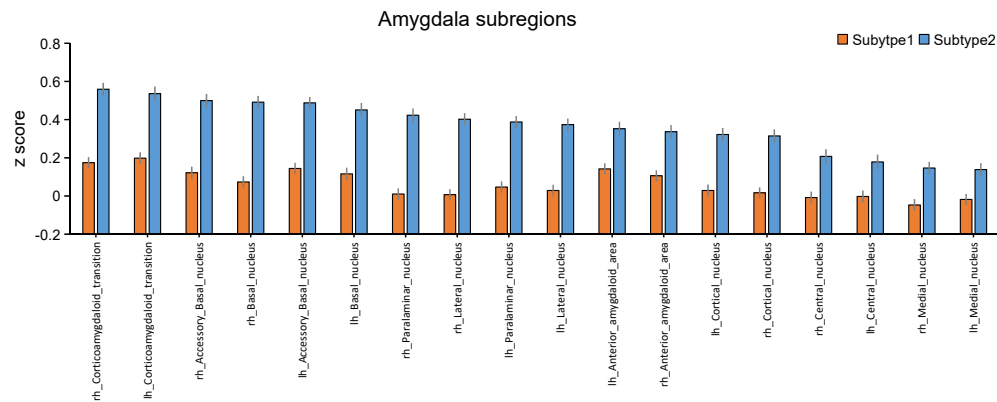


Extend Data Fig 2. Comparisons of morphological z-score between the two

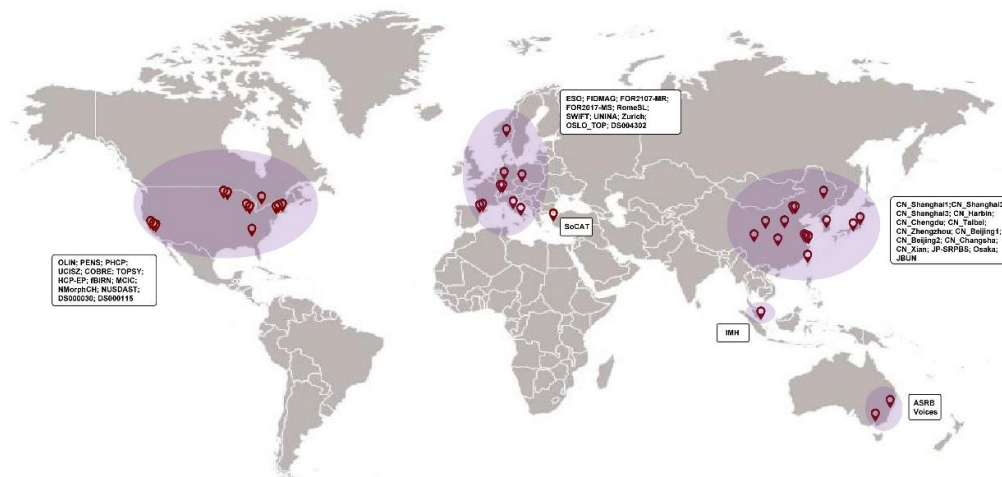
subtypes. A larger positive z-score indicates a larger deviation of reduction relative to healthy control group. Two sample t test is conducted to examine inter-subtype difference for the (a) averaged cortical volume ($t=9.36$, $p<10e-16$, Cohen's $d=0.446$); (b) averaged cortical area ($t=8.09$, $p<10e-16$, Cohen's $d=0.386$); (c) averaged cortical thickness ($t=1.29$, $p=0.198$, Cohen's $d=0.061$); (d) thalamus volume ($t=-4.28$, $p=1.97e-5$, Cohen's $d=-0.205$); (e) brain stem volume ($t=-9.79$, $p<10e-16$, Cohen's $d=-0.469$); (f) hippocampus volume ($t=-9.25$, $p<10e-16$, Cohen's $d=-0.449$); (g) amygdala volume ($t=-7.83$, $p=8.44e-15$, Cohen's $d=-0.379$); (h) accumbens volume ($t=-6.40$, $p=1.94e-10$, Cohen's $d=-0.305$); (i) caudate volume ($t=-9.82$, $p<10e-16$, Cohen's $d=-0.468$); (j) putamen volume ($t=-8.14$, $p<10e-16$, Cohen's $d=-0.389$).



Extend Data Fig 3. Hippocampus subregional morphological z-score for the two subtypes. A larger positive z-score indicates a larger deviation of reduction relative to healthy control group.



Extend Data Fig 4. Amygdala subregional morphological z-score for the two subtypes. A larger positive z-score indicates a larger deviation of reduction relative to healthy control group.



Extend Data Fig 5. Geographic map of included datasets.