Digital Twin Brain simulation and manipulation of a functional brain network underlying mental illness.

Authors:

Yunman Xia[1, 2†], Songjun Peng[3†], Juergen Dukart[4, 5], Chao Xie[6], Shitong Xiang[7, 8], Spase Petkoski[9], Zilin Li[10], Joerg Hipp[11], Suresh Muthukumaraswamy[12], Anna Forsyth[12], Tianye Jia[2, 7, 8], Nilakshi Vaidya[1], Tristram, Lett[1], Liyi Qian[13], Xiao Chang[1, 2, 7, 8], Yuxiang Dai[1, 2], Tobias Banaschewski[14], Gareth J. Barker[15], Arun L. W. Bokde[16], Rüdiger Brühl[17], Sylvane Desrivières[18], Herta Flor[19, 20], Penny Gowland[21], Antoine Grigis[22], Andreas Heinz[23, 24], Hervé Lemaître[22, 25], Frauke Nees[26], Dimitri Papadopoulos Orfanos[20], Tomáš Paus[27, 28], Luise Poustka[29], Michael N. Smolka[24], Sarah Hohmann[14], Henrik Walter[30], Robert Whelan[31], Paul Wirsching[30], Zuo Zhang[18, 32], Lauren Robinson[18, 33], Jeanne Winterer[30, 34], Yuning Zhang[35], Hedi Kebir[1], Ulrike Schmidt[36, 37], Julia Sinclair[38], Yuchen Liu[7], Jiexiang Wang[7], Fei Dai[7], Longbin Zeng[7], Yubo Hou[39], Huarui Wang[7], Leijun Ye[7], Chunhe Li[7, 40], Qibao Zheng[7], Andre Marquand[41], Changsong Zhou[3], Viktor Jirsa[9], Jianfeng Feng[7, 8, 40, 42-44], Wenlian Lu[7, 8, 40, 45], Gunter Schumann[1, 2, 46-48*], IMAGEN Consortium, STRATIFY Consortium, DTB Consortium, and environMENTAL Consortium.

Affiliations:

1. Centre for Population Neuroscience and Stratified Medicine (PONS), Department of Psychiatry and Neuroscience, Charite-Universitaetsmedizin Berlin, Germany.

2. PONS, Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China.

3. Department of Physics, Hong Kong Baptist University, Hong Kong SAR, China

4. Institute of Neurosciences and Medicine, Brain & Behaviour (INM-7), Research Centre Juelich, Juelich, Germany.

5. Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany.

6. Department of Psychological and Cognitive Sciences, Tsinghua University, Beijing, China.

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

8. Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Fudan University, Ministry of Education, Shanghai, China.

9. Institut de Neurosciences des Systèmes (INS) UMR1106, Aix Marseille Université, Marseille, France.

10. Department of Psychiatry, University of Oxford, Oxford, United Kingdom.

11. Roche Pharma Research and Early Development, Neuroscience and Rare Diseases, Roche Innovation Center Basel, F. Hoffmann–La Roche Ltd., Basel, Switzerland.

12. School of Pharmacy, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand.

13. Department of Neurosurgery, Huashan Hospital, Fudan University, Shanghai, China.

14. Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany.

15. Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom.

16. Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland.

17. Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany.

18. Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom.

19. Institute of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany.

20. Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany.

21. Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, United Kingdom.

22. NeuroSpin, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France.

23. Department of Psychiatry and Psychotherapy, University of Tübingen, Germany; German Center for Mental Health (DZPG), Site Tübingen, Germany.

24. Department of Psychiatry and Psychotherapy, Technische Universität Dresden, Dresden, Germany.

25. Institut des Maladies Neurodégénératives, UMR 5293, CNRS, CEA, Université de Bordeaux, 33076 Bordeaux, France.

26. Institute of Medical Psychology, Ludwig-Maximilians-Universität (LMU) in Munich, Munich, Germany.

27. Department of Psychiatry, Faculty of Medicine and Centre Hospitalier Universitaire Sainte-Justine, University of Montreal, Montreal, Quebec, Canada.

28. Departments of Psychiatry and Psychology, University of Toronto, Toronto, Ontario, Canada.

29. Department of Child and Adolescent Psychiatry, Center for Psychosocial Medicine, University Hospital Heidelberg, Heidelberg, Germany.

30. Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany.

31. School of Psychology and Global Brain Health Institute, Trinity College Dublin, Ireland.

32. School of Psychology, Institute for Mental Health, University of Birmingham, Birmingham, United Kingdom.

33. Oxford Institute of Clinical Psychology Training and Research, University of Oxford, Oxford, United Kingdom.

34. Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany.

35. Psychology Department, University of Southampton, Southampton, United Kingdom.

36. Department of Psychological Medicine, Section for Eating Disorders, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom.

37. South London and Maudsley NHS Foundation Trust, London, United Kingdom.

38. Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom.

39. School of Mathematical Sciences, Fudan University, Shanghai, China.

40. Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China.

41. Donders Institute, Radboud University, Nijmegen, Netherlands.

42. School of Data Science, Fudan University, Shanghai, China.

43. Department of Computer Science, University of Warwick, Coventry, United Kingdom.

44. Zhangjiang Fudan International Innovation Centre, Shanghai, China.

45. Center for Applied Mathematics, Fudan University, Shanghai, China

46. PONS Centre, National Center for Neurologic Disorders, Huashan Hospital, Fudan University, Shanghai, China.

47. German Centre for Mental Health (DZPG), Berlin-Brandenburg site, Germany.

48. Department of Psychiatry, Cambridge University, United Kingdom.

*Corresponding author:

Gunter Schumann

Email: gunter.schumann1961@gmail.com and gunter.schumann@charite.de

† These authors contributed equally to this work.

Abstract

Linking synaptic-level perturbations to distributed brain-network dynamics remains a central challenge for understanding and treating mental illness. Although recent whole-brain models can reproduce individual brain activity patterns, they largely function as descriptive simulators rather than mechanistic, intervention-capable systems. Here we present an intervention-capable digital twin of the human brain, integrating individual neuroanatomy and task-evoked dynamics within a neuronal-scale framework. Individualised digital twin brains recapitulate a participant-specific compact cortico-subcortical network phenotype that captures transdiagnostic psychopathology across population and clinical cohorts. In silico modulation of excitatory and inhibitory synaptic conductance produces bidirectional, heterogeneous network responses across individuals. Population-scale simulations stratify individuals and predict longitudinal symptom trajectories from DTB-derived response profiles. Independent pharmacological functional MRI data further validate the predicted baseline-dependent network responses in vivo. Together, these findings establish digital brain models as experimental platforms for mechanistic perturbation, behavioural prediction and stratification, providing a foundation for precision neuroscience and psychiatry.

Main

As human neuroimaging has reached individual-level precision[1,2], a central challenge has emerged: how to transform rich multimodal measurements of brain structure and function into mechanistic models that can be systematically perturbed to generate causal and predictive insights into highly prevalent mental illness[3,4]. Digital twin models of the human brain offer a powerful framework for addressing this challenge by integrating neuroanatomy, physiology, and large-scale functional dynamics within a unified computational system[5–8].

Recent advances have demonstrated that Digital Twin Brain (DTB) framework can approximate large-scale neuronal dynamics using the biologically constrained spiking neural networks grounded in individual anatomy and empirical BOLD signals[9,10], establishing the feasibility of personalised whole-brain simulation. However, existing digital twin brains primarily function as descriptive simulators rather than enabling systematic mechanistic perturbation and prediction.

This limitation is especially salient in psychiatry. Mental disorders are characterised by distributed network alterations[11] and heterogeneous treatment response[12–15]. Neuroimaging studies have identified reproducible functional connectivity abnormalities across disorders, often implicating common circuits involved in reward processing, cognitive control, and salience attribution[15–22]. Yet without the ability to perform controlled perturbations, these findings remain difficult to translate into mechanistic insight or predictive models of intervention effects[13,14]. At the same time, most psychiatric treatments act on specific molecular targets - such as glutamatergic or GABAergic neurotransmission - whose effects propagate across multiple spatial scales, from synapses to distributed functional brain networks[23–27]. Bridging these levels of description remains a critical gap.

A promising strategy is to focus on behaviourally grounded, transdiagnostic brain networks that provide low-dimensional interfaces between cognition, symptoms, and neurobiology[18–20,28]. Converging evidence suggests that shared dimensions of psychopathology are underpinned by common neural circuit disruptions that cut across traditional diagnostic boundaries[17–20,28–30]. The neuropsychopathology (NP) factor, derived from task-based functional connectivity, captures shared continuous variance across internalising and externalising symptoms in the population cohort[28]. As such, it provides a

principled target for testing whether digital twin brains can move beyond simulation toward mechanistic manipulation of behaviourally relevant circuits.

In parallel, a large body of neurochemical evidence implicates disruptions in cortical excitation–inhibition (E/I) balance in the pathophysiology of mood and addiction disorders. Magnetic resonance spectroscopy and postmortem studies report reductions in both glutamatergic and GABAergic signalling across cortex[31–36], while pharmacological agents targeting these systems - including ketamine and benzodiazepines - can induce rapid changes in symptoms and functional connectivity[23,26,37–41]. Despite this, the relationship between neurotransmitter-level perturbations and large-scale functional network reconfiguration remains poorly understood. Empirical pharmaco-fMRI can quantify individual-level effects of neurochemical interventions, but offers limited leverage to mechanistically disentangle or systematically manipulate the underlying synaptic processes. Existing whole-brain models, in turn, have not yet been configured to support controlled, neurotransmitter-level perturbations to task-state network reconfiguration at the individual level[8,42–44].

Here we build on the DTB framework[9,10] to address this unmet need. We tested whether a neuronal-scale digital twin can reproduce a behaviourally relevant transdiagnostic brain network (NP factor) and mechanistically manipulate it through abstract synaptic gain control parameters inspired by E/I mechanisms. We further examined whether such in silico manipulations yield predictions that generalise across tasks and populations, align with empirical pharmacological effects, and forecast longitudinal symptom trajectories. Together, this work marks a step change in digital brain modelling from passive emulation of neural activity to intervention-capable, mechanistically closed-loop systems that integrate simulation, perturbation, and prediction.

A compact cortico–subcortical network phenotype of psychopathology.

We first delineated the NP factor, a shared functional connectivity phenotype that captures psychopathology across symptom domains and cohorts. As in our previous work[28], conducted in 14 year-old adolescents of the IMAGEN study, we applied connectome-based predictive modeling to the task-based functional connectomes derived from the monetary incentive delay and stop-signal task in 19 year old young adults of IMAGEN follow-up 2 (n=1,050, age 18.4±0.7 years, Table S1). Four of six task conditions (stop success and failure in SST; reward anticipation and positive feedback in MID) significantly predicted a range of externalising and internalising symptoms at age 19 (Fig.1a; Table S2-3). Across these predictive task conditions, we identified two transdiagnostic functional connectivity (FC) profiles: one positive profile of 34 edges where stronger connectivity was associated with higher symptom burden across domains, and one negative profile of 29 edges where weaker connectivity was associated with higher burden. These edges were concentrated within and between default mode, frontoparietal, somatomotor–frontal, limbic, and cerebellum systems (Fig.1b). To quantify these functional connectivity profiles, we computed positive and negative NP scores for each individual by summing the strength of edges in each profile.

To test whether the NP scores generalise to clinical populations, we re-computed them from task-state FC in an independent case–control cohort (STRATIFY, n=434, age 21.8±2.0 years). Only the negative NP score showed robust case-control differences, with lower scores in patients than in healthy controls (Fig.1c; Fig.S1). Thus, the negative NP profile defines a shared network phenotype of transdiagnostic psychopathology.

To characterise the neurobiological context of this negative NP circuit, we examined whether negative NP-related regions exhibit a distinctive neurochemical signature in

196 empirical PET receptor maps. Using published PET-derived receptor density maps[45], we
197 found that NP-related regions showed higher receptor-map values for glutamatergic,
198 serotonergic, cholinergic, and cannabinoid systems than NP-unrelated regions (Fig.S2, Table
199 S4, all $P_{corrected} < 0.05$). These receptor systems are primarily involved in regulating cortical
200 excitatory drive and neuromodulatory control of network dynamics, and are known to
201 influence E/I balance at the circuit level through modulation of glutamatergic
202 transmission[46,47], GABA interneuron recruitment[48–50], recurrent coupling[51], and gain control[52]
203 mechanisms.

204 To enable computationally tractable digital twin brain (DTB) modeling, we restricted the
205 negative NP profile to cortico-subcortical edges, excluding the cerebellum and brainstem.
206 These regions exhibit distinct cytoarchitecture and neuronal dynamics[53] and are not currently
207 parameterized in our DTB framework (see Methods for details). Finally, we identified a
208 compact set of 12 task-state functional connections linking regions spanning prefrontal
209 control, default-mode, sensorimotor, and limbic systems (Fig.1d; Table S5). The summed
210 strength of these 12 edges remained significantly reduced in patients relative to controls
211 (Fig.1e; Fig.S1), and served as the final NP factor and target for subsequent DTB simulations
212 and in-silico E/I modulations.

213 **Individual digital twin brains reveal a bidirectionally E/I-tunable NP factor.**

214 We next asked whether this NP factor can be reconstructed and mechanistically manipulated
215 in individualised, biologically grounded models. To this end, we applied the previously
216 established DTB framework[9,10] to generate participant-specific task-state models (Methods;
217 Fig.2a; Fig.S3). Following a calibration experiment across neuronal resolutions (Fig.S4; Table
218 S8; more details in Methods and Supplementary Results), we implemented DTBs at three

complementary scales: billion-neuron voxel-wise models to establish biological fidelity and simulation feasibility, 100-million-neuron models for controlled perturbational exploration, and computationally tractable regional models for population-level statistical inference. Across all scales, models were constructed using a common pipeline. The individual voxel-wise grey and white matter anatomy were first transformed into a multi-population neuronal network. The activity of each neuron in this network was modelled using the leaky integrate-and-fire (LIF) model[54], equipped with excitatory (AMPA-mediated) and inhibitory (GABA-A-mediated) synapses. We then estimated hyperparameters by assimilating empirical BOLD signals in task-engaged regions (Methods; Table S6) to mimic external current inputs driving the model from rest to task-states[9,10]. Finally, the voxel-wise neuronal activity (mean firing rate of neuronal population) was converted into BOLD signals using the Balloon-Windkessel model[55].

We first selected four participants as illustrative exemplars according to the behavioural symptoms and negative NP scores, including participants with major depressive disorder (MDD), alcohol use disorder (AUD), and healthy controls (HC01 and HC02) (demographics information in Table S7, selection criteria in Methods), to demonstrate the feasibility and mechanistic properties of neuronal-scale digital twin simulations. These examples are presented for illustration only; all statistical inferences are based on group-level analyses that account for site and sex effects (see Supplementary Results). Using a 1-billion-neuron scale (Fig.S4; Table S8), their individualised DTBs accurately reproduced task-evoked dynamics in MID and SST. Simulations achieved high voxel-wise correlations with empirical BOLD in assimilated regions (mean r=0.9, in-sample correlations). Across the whole brain, mean correlations of four individuals (mean r=0.30-0.44) were clearly above a reference simulation obtained by a single random parameter shuffle (mean r=0) (Fig.2b-c; Fig.S5-6;

243 Table S9). These DTBs also produced whole-brain functional connectivity patterns that
244 closely matched empirical connectomes in both tasks (Fig.2d; Table S9).

245 Based on the simulated baseline NP configuration, we next examined whether NP circuit
246 dynamics could be systematically manipulated within the DTB framework. In recurrent
247 cortical networks, E/I balance emerges from the interaction of inhibitory interneuron
248 recruitment[56,57], recurrent coupling[51,57], and synaptic gain[58], and is governed by circuit
249 dynamics rather than receptor density alone. Consistent with this systems-level perspective,
250 we treated both AMPA- and GABA-A-mediated synaptic conductance as global dynamical
251 control parameters, enabling controlled perturbation of NP network activity within
252 individualised digital twins. Notably, publicly available PET datasets did not show
253 preferential expression of GABA receptor systems in NP-related regions, reinforcing the
254 interpretation that inhibitory modulation in the model reflects circuit-level dynamical control
255 rather than regionally localized receptor enrichment.

256 Accordingly, we varied the synaptic conductance parameters for AMPA and GABA-A
257 receptors in the 100-million-neuron DTBs of the same illustrative participants. Despite the
258 reduced resolution, DTBs preserved high fidelity in assimilated regions and acceptable
259 whole-brain correspondence (Table S10). Moreover, 100-million-neuron DTBs retained
260 sufficient subject-specific NP-factor phenotypes such that each individual's simulated NP
261 factor best matched their own empirical value, with significantly lower mean-squared error
262 for within-subject matches than any cross-subject pairing (permutation tests across 1,575
263 simulated–empirical pairings, Supplementary Results; Fig.S7), therefore supporting its use for
264 controlled perturbational experiments. Using these established DTB models, we performed
265 systematic perturbations by upregulating synaptic conductance for AMPA (0.0020-0.0052)
266 and GABA-A receptors (0.0015-0.0040) across a predefined parameter grid, guided by

evidence for reduced glutamatergic and GABAergic function in depression[24,25] and addiction[27,34]. Parameter ranges were constrained by biologically plausible neuronal firing rates[59], synchronization[60], and simulated BOLD activity. During GABA-A manipulations, AMPA conductance was fixed at the value that maximized NP factor enhancement in prior sweeps, allowing controlled assessment of E/I balance and avoiding excessive inhibition (see more details in Methods).

Virtual modulation of AMPA- and GABA-A-mediated synaptic conductance produced bidirectional, heterogeneous changes in NP connectivity across illustrative participants. Specifically, in two patient brain models, both excitatory and inhibitory upregulation increased NP-factor strength, and in illustrative healthy control models, the same perturbations yielded divergent responses, with modest increases in one individual and clear decreases in the other (Fig.2e-f; Tables S11-12). These examples demonstrate that identical synaptic perturbations can lead to bidirectional changes in NP-factor strength across individuals.

To test whether this bidirectional response generalises across cognitive contexts, we repeated the same E/I perturbations in an alternative task (the Emotional Face Task) in the same participants. Comparable bidirectional modulation effects were observed (Supplementary Results; Fig.S8; Table S13), indicating that the E/I-sensitive response properties of the NP circuit are not task-specific, but generalise across distinct cognitive contexts.

Population-scale in silico perturbations attenuate NP abnormalities and stratify individuals.

The single-subject analyses suggest that in silico E/I perturbation can shift NP network toward the control distribution observed in STRATIFY (reflected by increased NP-factor strength), although not uniformly across individuals. To further explore whether these effects generalise

290 across the cohort and relate to behavioural differences, we performed population-scale
291 simulations, which allow us to link individual susceptibility to excitatory and inhibitory
292 synaptic perturbations with behavioural symptoms. To enable large-cohort simulations, we
293 developed regional DTBs that convert regional MRI-derived structural features into spiking
294 neural networks comprising 3 million neurons (identified from calibration experiment, Fig.S4,
295 Table S14), with neuron counts allocated proportionally to regional grey matter volume to
296 preserve anatomical scaling while maintaining computational tractability (Methods). While
297 voxel-wise DTBs establish mechanistic feasibility at high-fidelity resolution, these regional
298 DTBs are used to test whether these principles scale to population-level inference.

299 We constructed individualised DTBs for 72 patients with MDD, 59 with AUD and 69 healthy
300 controls from STRATIFY, as well as 90 individuals with high externalising and internalising
301 symptoms from IMAGEN (Table S15). These participants from STRATIFY were included
302 because they possessed complete, high-quality multimodal imaging (T1-weighted, DTI, fMRI)
303 and behavioural data. These DTB models reproduced individual task-evoked BOLD signals
304 with moderate-to-high accuracy during in-sample fitting phase (Fig.3a), generated NP factors
305 that closely matched empirical values at the group level (Fig.S9), and mirrored case–control
306 group differences observed in vivo (Fig.3b; Fig.S9). As an additional validation beyond NP
307 circuit, simulated whole-brain FC also predicted task-related behavioural performances and,
308 on average, outperformed empirical FC (Supplementary Results; Fig.S10; Table S16),
309 suggesting that DTBs distill behaviourally relevant variance that is only partially captured by
310 conventional functional connectivity.

311 Motivated by these converging validations, we then applied the same AMPA and GABA-A
312 neuromodulation schemes used in the 100-million-neuron DTBs to all 290 participants. At
313 the group level, both excitatory and inhibitory upregulation shifted NP factors in patients

314 toward values observed in controls, in other words, after either AMPA or GABA-A
315 modulation, previously significant differences between patients, high-symptom individuals
316 and healthy controls were no longer significant (Fig.3c-d). Within each clinical group, NP
317 factors increased on average following both AMPA and GABA-A manipulations (Fig.3e; Fig.
318 S11).

319 Despite clear group-level increases, individual responses were heterogeneous and often
320 bidirectional (Fig.3f). Following AMPA modulation, both patients and controls showed
321 increased NP factor on group average, but the proportion of "increasers" (ΔNP > 0) was
322 substantially higher among patients than controls (Table S17). Baseline simulated NP factor
323 was strongly and inversely correlated with AMPA-induced change (Fig.3g), such that
324 individuals with lower baseline connectivity, typically patients, exhibited the largest gains,
325 whereas those with high baseline NP showed modest increases or decreases. GABA-A
326 modulation produced a similar, though weaker, negative association (Fig.3h). Importantly,
327 these baseline–ΔNP couplings significantly exceeded a null distribution generated by
328 permuting ΔNP values across individuals (10,000 permutations, two-sided $P<0.0001$; see
329 Methods), indicating that the relationship reflects structured baseline-dependent dynamics
330 rather than statistical regression to the mean.

331 To facilitate clinical interpretation of this response heterogeneity, we descriptively stratified
332 individuals based on the direction of their joint responses to AMPA and GABA-A
333 perturbations. Participants whose NP factors increased under both manipulations were
334 classified as "increased responders", whereas those with at least one decrease were classified
335 as "decreased responders". This stratification was used for visualization and group-level
336 comparison only; the underlying baseline–ΔNP relationships were treated as continuous
337 throughout statistical analyses. Increased responders were markedly more prevalent among

338 patients than among high-symptom individuals or healthy controls (Fig.3i) and did not differ
339 systematically by site or sex (Tables S18–19; Supplementary Results). Clinically, the
340 increased responders, characterised by lower simulated baseline NP factors, showed greater
341 symptom burden, including more panic attacks and negative affect, and reduced social
342 support, compared with the decreased responders after controlling for sex, site, and head
343 motion (Fig.3j; Table S20).

344 Together, these findings indicate that DTB-derived virtual responses offer an individualised,
345 mechanistically interpretable stratification, whereby different responses in NP circuit map
346 onto meaningful differences in clinical and behavioural profiles. A critical question, however,
347 is whether these virtual baseline-dependent, bidirectional response structures also emerge
348 under real pharmacological perturbations in vivo.

349 **Pharmacological E/I perturbation in vivo mirrors digital twin predictions of NP**
350 **modulations.**

351 To address this, we analysed pharmacological fMRI data from 27 healthy male participants
352 (27.3±6.2 years), who performed the MID task following ketamine (0.25 mg/kg), midazolam
353 (0.03 mg/kg), or placebo administration in a randomized, single-blinded, three-way
354 crossover design (Fig.4a). Ketamine, an NMDA receptor antagonist, reduces inhibitory
355 interneuron activity and indirectly enhances AMPA-mediated excitatory transmission[23,61],
356 whereas midazolam, a GABA-A receptor positive allosteric modulator, increases inhibitory
357 synaptic conductance and dampens network excitability[62,63]. Because pharmacological data
358 were only available for the MID task-fMRI, we focused on a MID-specific NP connectivity,
359 defined as the summed strength of the six MID edges that contributed to the NP circuit.
360 Placebo served as the empirical analogue of the DTB baseline simulation, whereas ketamine

361 and midazolam provided in vivo perturbations with predominant effects on excitatory-
362 leaning and inhibitory neurotransmission, respectively. We tested whether these
363 perturbations reproduce the baseline-dependent response structures predicted by AMPA and
364 GABA-A modulations in silico.

365 At the group level, neither ketamine nor midazolam induced a significant change in summed
366 NP-related MID connectivity relative to placebo (Fig.S12), suggesting that conventional
367 group-mean analyses would miss potential relationship between brain network configuration
368 and pharmacological response. However, individual-level analyses revealed a bidirectional,
369 baseline-dependent pattern closely resembling DTB simulations. Unsupervised k-means
370 clustering of combined ketamine- and midazolam-induced FC changes identified two
371 subgroups (Fig.S12). In Group 1 (n=19), both ketamine and midazolam significantly
372 increased NP-related connectivity after adjusting for head motion effects (Fig.4b-c), whereas
373 in Group 2 (n=8), both drugs significantly decreased it (Fig. 4d-e). Baseline NP-related
374 connectivity under placebo was significantly lower in the increased-response group than in
375 the decreased-response group (Fig.4f), and baseline FC was inversely related to drug-induced
376 change across individuals (Fig.4g-h). This inverse baseline–ΔFC relationship significantly
377 exceeded a permutation-derived null distribution generated by shuffling drug-induced
378 changes across individuals (10,000 permutations; midazolam: two-sided P=0.0056,
379 ketamine: two-sided P=0.069, showing a trend in the same direction; see Methods),
380 indicating that the bidirectional response is unlikely to reflect regression to the mean alone.
381 Thus, pharmacological E/I perturbation in vivo reproduces the bidirectional, baseline-
382 dependent modulation of NP circuit predicted in silico by DTBs.

383 To match the pharmacological dataset (MID only), we restricted virtual modulation to MID-
384 specific NP connectivity to place the model and pharmacological data in the same task

385 context. Importantly, a bidirectional, baseline-dependent response toward virtual
386 modulation of MID-specific NP connectivity was preserved (Supplementary Results; Fig.
387 S13), indicating that the core dynamic property of the NP circuit is robust to this task
388 restriction.

389 We then asked whether a quantitative mapping learned entirely in silico could generalise to
390 predict individual pharmacological responses. In DTBs, we first estimated a linear mapping
391 between simulated baseline MID FC and virtual modulation-induced changes in NP-related
392 connectivity. Applying this same mapping to empirical placebo MID FC, we obtained
393 subject-specific predictions of ketamine- and midazolam-induced FC changes. Predicted
394 and observed drug-induced changes closely matched at the individual level (Fig.4i-j) and
395 stratified participants with opposite response directions (AUC of ketamine=0.69; AUC of
396 midazolam=0.83; Fig.4k), further supporting the translational relevance of DTB-derived
397 virtual modulation.

398 Together, these findings show that DTBs not only recapitulate individual-level, baseline-
399 dependent response structure observed under pharmacological perturbation, but also
400 provide a mechanistic template for forecasting individual drug-induced circuit shifts from
401 baseline network configuration.

402 **DTB-derived perturbational responses predicts future changes in symptoms.**

403 We next asked whether DTB-derived perturbational responses carry prospective information
404 about longer-term symptom trajectories. Focusing on internalising symptoms (given the
405 limited prevalence of externalising symptoms at age 23), we first learn a linear mapping from
406 empirical NP connectivity at age 19 to concurrent internalising symptom scores (Fig.4l). We
407 then applied this mapping to simulated NP connectivity before and after virtual E/I

408 modulation to obtain subject-specific predicted symptom changes, termed the "behavioural
409 restoration" index.

410 We quantified observed symptom change over four years (Δbehaviour, ages 19–23) and
411 tested whether the DTB-derived "behavioural restoration" index predicted these longitudinal
412 changes. Across individuals, baseline symptoms and the restoration index for AMPA
413 modulation together explained ~23% of the variance in the symptom changes (Fig.4m). The
414 restoration index alone was significantly correlated with empirical symptoms changes
415 (Fig.S14), indicating that individuals whose NP circuit was more "normalisable" in silico
416 were less likely to deteriorate and more likely to remit. Importantly, the restoration index
417 contributed unique predictive value beyond baseline symptoms, significantly increasing
418 explained variance by ~5% relative to a symptom-only model (Fig.S14). However, the
419 behavioural restoration index for GABA-A modulation was not correlated with empirical
420 symptom changes (Fig. S14).

421 These findings suggest that DTB-derived perturbational response metrics capture not only
422 cross-sectional heterogeneity but also clinically relevant dynamical properties that forecast
423 longitudinal trajectories. Taken together with the pharmacological validation above, these
424 results position DTB models as a promising tool for predicting and stratifying individualised
425 responses to neurotransmitter-targeted interventions, with potential implications for the
426 development and personalization of future psychiatric treatments.

427 **Discussion**

428 We simulated a transdiagnostic, behaviourally relevant task-evoked functional network
429 phenotype (i.e., NP factor) within individualised neuronal-scale digital twin brain models.
430 By systematically perturbing AMPA- and GABA-A–mediated synaptic gain, we quantified

individual-specific, bidirectional network responses and linked these virtual response profiles to both pharmacological fMRI effects and longitudinal symptom trajectories. These results demonstrate that digital brain models can move beyond descriptive simulation of neural activity to perform controlled in silico perturbations for mechanistic interrogation and individualised counterfactual prediction. By introducing biologically interpretable synaptic parameters, our framework allows precise manipulation of excitation–inhibition balance at the neuronal scale while preserving each individual's anatomical and task-evoked constraints, establishing a new class of perturbational and predictive digital brain models for computational neuroscience and psychiatry.

The NP factor represents a compact functional connectivity pattern that relates to a broad spectrum of internalising and externalising symptoms across heterogeneous populations. It is characterised by reduced functional integration among limbic, default-mode, frontoparietal control, and cingulo-opercular networks, which jointly support reward processing, cognitive control, and attention allocation[64–67]. Disruptions within and between these networks have been reported across mood, anxiety, and substance use disorders, as well as in population-based samples spanning the full range of symptom severity[18–20,68]. In this sense, the NP factor reflects a shared circuit phenotype rather than a disorder-specific marker. Importantly, we treat NP as a continuous circuit dimension rather than a diagnostic classifier. This makes it suitable as a perturbational readout: it is sufficiently low-dimensional to allow explicit predictions about network reconfiguration, while remaining anchored in identifiable large-scale functional systems.

A central contribution of the NP-DTB framework is that it renders a descriptive task-evoked connectivity phenotype experimentally manipulable. Conventional neuroimaging analyses typically relate connectivity measures to behavioural symptoms, but they do not permit

controlled intervention at the circuit level[69,70]. By contrast, DTBs establish an explicit link between large-scale functional network organization and synaptic-scale parameters, enabling in silico perturbation of defined excitatory and inhibitory mechanisms. In addition, DTBs reproduce individualised, task-evoked functional connectivity patterns that are challenging for the conventional whole-brain models, which often rely on coarse-grained regional parcellations and focus primarily on resting-state dynamics[71,72]. This task-state emphasis is critical, as many clinically relevant circuit abnormalities are context dependent[20,73], and network configurations expressed during reward processing or inhibitory control cannot be inferred from resting-state activity alone[74]. By assimilating task-state signals and applying perturbations at the synaptic level, DTBs generate individual-specific predictions of how functional circuits respond to mechanistically interpretable excitatory or inhibitory modulation.

Using this framework, we observed a pronounced baseline-dependent heterogeneity in NP connectivity response to virtual E/I perturbations. Individuals with lower baseline NP factor scores tend to show larger increases following perturbation, whereas those with higher baseline NP show modest increases or even decreases. Interestingly, both excitatory-leaning (AMPA) and inhibitory-leaning (GABA-A) perturbations often shifted NP connectivity in the same direction within a given individual model. This convergence does not contradict excitation–inhibition balance; rather, it reflects the nonlinear, regime-dependent dynamics of large-scale recurrent networks, in which multiple control parameters can move the system toward a balanced operating regime depending on baseline state[75–77], consistent with prior empirical and computational modeling studies[57,78–81]. Notably, the bidirectional convergence of excitatory- and inhibitory-leaning perturbations at the network level does not imply

478 equivalence of their molecular mechanisms, but instead reflects compensatory and
479 nonlinear circuit dynamics in which E/I balance acts as a key control variable[57,79,80].

480 Two independent lines of evidence further support the biological plausibility of the causal
481 links between microscale neurotransmitter systems and large-scale functional networks
482 inferred by the DTB models. First, PET-derived receptor maps show that regions contributing
483 to the NP circuit are enriched in receptor systems involved in regulating excitation–inhibition
484 balance[46,49,51,52], providing a neurochemical context for why E/I perturbations plausibly
485 modulate this circuit, although these maps are group-level and cannot establish individual
486 causality. Second, pharmacological fMRI data provide an independent empirical
487 perturbation reference for examining how excitatory and inhibitory neurotransmitter
488 modulation affects NP-related connectivity. Although ketamine and midazolam did not
489 produce robust group-mean effects on NP-related connectivity, individual-level analyses
490 revealed bidirectional, baseline-dependent responses that closely mirror DTB predictions.
491 The convergence between in silico perturbations and pharmacological modulation in vivo
492 suggests that heterogeneity in NP network responses may reflect differences in baseline
493 circuit state, a property also evident in pharmacological responses, rather than an artifact of
494 the digital twin models.

495 Therefore, a key future implication of the DTB framework is that an individual's baseline
496 functional network configuration carries information about the direction and magnitude of
497 network change under perturbation. This is supported by our analyses showing a significant
498 inverse relationship between baseline NP connectivity and modulation-induced ΔNP in both
499 AMPA- and GABA-A–mediated virtual perturbations, as well as in pharmacological fMRI
500 data (Fig. 3g–h; Fig. 4g-k). Depending on their initial circuit state, the same intervention may
501 increase network integration in some individuals while decreasing it in others. In this work,

we illustrate the translational potential of this approach in two ways. First, we learned a mapping from baseline NP connectivity to virtual perturbation responses in silico and applied it out-of-sample to forecast drug-induced circuit changes in vivo. Second, by linking virtual NP network changes to behaviour, we define a "behavioural restoration" index that quantifies an individual's brain circuit susceptibility to intervention. This index predicts differences in future symptom trajectories beyond what baseline symptom severity alone can explain, providing a measurable, testable readout of circuit "normalisability" based directly on the model's simulated perturbation responses. Beyond prediction, the DTB framework also supports mechanistic stratification based on baseline brain network configuration rather than symptom severity alone. Individuals with similar symptom burden but distinct NP configurations can follow divergent trajectories over time, offering a mechanistic explanation for why symptom-based stratification often fails to anticipate treatment response. However, we should note that the excitatory and inhibitory perturbations implemented in the DTB represent controlled shifts in synaptic gain that move the system across distinct network operating regimes, rather than modeling reality drug actions. Thus, this predictive framework is intended for risk stratification and mechanistic insight, rather than deterministic individual-level prognosis.

Notably, model-empirical similarity increased with neuronal resolution in voxel-wise simulations, consistent with scaling properties of the previous studies[9,10]. Regional DTBs showed higher similarity in aggregated BOLD dynamics but lower correspondence in functional connectivity relative to voxel-wise models, likely reflecting reduced spatial granularity for capturing individual-level network structure. Nevertheless, regional models remained computationally tractable and retained sufficient fidelity to predict task performance and empirical group-level differences, supporting their use for population-scale

perturbational analyses. Rather than serving as exact replicas of biological brains, DTBs should be viewed as experimentally controllable models that enable causal hypothesis testing in systems otherwise inaccessible to direct manipulation.

Several limitations should be noted. DTB inferences depend on simplifying modelling assumptions, e.g., fixed excitatory-inhibitory neuron ratios, homogeneous synaptic architecture, and simplified neuronal dynamics. Accordingly, AMPA- and GABA-A–labelled parameters should be interpreted as phenomenological gain parameters rather than direct mapping to drug mechanisms. The present DTB framework does not incorporate the full range of neurotransmitter systems, nor cerebellar microcircuit dynamics, and pharmacological validation was limited to a single task and cohort. Although PET receptor maps provide biological plausibility, they do not capture individual neurochemical variability. Despite these constraints, the convergent evidence suggests that a compact task-evoked circuit phenotype can be linked to explicit synaptic control parameters in individualised models, yielding baseline-dependent, and bidirectional predictions that align with pharmacological perturbations and carry prognostic information. Together, these findings motivate a shift from descriptive biomarkers toward perturbational models that explain and predict individual differences in psychiatric trajectories.

References

1. Dubois, J. & Adolphs, R. Building a Science of Individual Differences from fMRI. *Trends Cogn. Sci.* **20**, 425–443 (2016).

2. Bassett, D. S. & Sporns, O. Network neuroscience. *Nat. Neurosci.* **20**, 353–364 (2017).

3. Sejnowski, T. J., Churchland, P. S. & Movshon, J. A. Putting big data to good use in neuroscience. *Nat. Neurosci.* **17**, 1440–1441 (2014).

4. Huys, Q. J. M., Maia, T. V. & Frank, M. J. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* **19**, 404–413 (2016).

5. Lee, J. H., Liu, Q. & Dadgar-Kiani, E. Solving brain circuit function and dysfunction with computational modeling and optogenetic fMRI. *Science* **378**, 493–499 (2022).

6. Jirsa, V. K., Sporns, O., Breakspear, M., Deco, G. & McIntosh, A. R. Towards The Virtual Brain: network modeling of the intact and the damaged brain. *Arch. Ital. Biol.* **148**, 189–205 (2010).

7. Deco, G. & Kringelbach, M. L. Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. *Neuron* **84**, 892–905 (2014).

8. Shine, J. M. *et al.* Computational models link cellular mechanisms of neuromodulation to large-scale neural dynamics. *Nat. Neurosci.* **24**, 765–776 (2021).

9. Lu, W. *et al.* Imitating and exploring human brain's resting and task-performing states via resembling brain computing: scaling and architecture. *Natl. Sci. Rev.* nwae080 (2024) doi:10.1093/nsr/nwae080.

565 10. Lu, W. *et al*. Simulation and assimilation of the digital human brain. *Nat. Comput.*
566 *Sci.* **4**, 890–898 (2024).

567 11. Braun, U. *et al*. From Maps to Multi-dimensional Network Mechanisms of Mental
568 Disorders. *Neuron* **97**, 14–31 (2018).

569 12. Global, regional, and national burden of 12 mental disorders in 204 countries and
570 territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019.
571 *Lancet Psychiatry* **9**, 137–150 (2022).

572 13. Patel, V. *et al*. The Lancet Commission on global mental health and sustainable
573 development. *The Lancet* **392**, 1553–1598 (2018).

574 14. McIntyre, R. S. *et al*. Treatment-resistant depression: definition, prevalence, detection,
575 management, and investigational interventions. *World Psychiatry Off. J. World Psychiatr.*
576 *Assoc. WPA* **22**, 394–412 (2023).

577 15. Kotov, R. *et al*. The Hierarchical Taxonomy of Psychopathology (HiTOP): A
578 dimensional alternative to traditional nosologies. *J. Abnorm. Psychol.* **126**, 454–477 (2017).

579 16. Insel, T. R. The NIMH Research Domain Criteria (RDoC) Project: precision medicine
580 for psychiatry. *Am. J. Psychiatry* **171**, 395–397 (2014).

581 17. Caspi, A. *et al*. The p Factor: One General Psychopathology Factor in the Structure
582 of Psychiatric Disorders? *Clin. Psychol. Sci. J. Assoc. Psychol. Sci.* **2**, 119–137 (2014).

583 18. Vanes, L. D. & Dolan, R. J. Transdiagnostic neuroimaging markers of psychiatric risk:
584 A narrative review. *NeuroImage Clin.* **30**, 102634 (2021).

585 19. Goodkind, M. *et al*. Identification of a Common Neurobiological Substrate for Mental
586 Illness. *JAMA Psychiatry* **72**, 305–315 (2015).

587 20. McTeague, L. M. *et al*. Identification of Common Neural Circuit Disruptions in
588 Cognitive Control Across Psychiatric Disorders. *Am. J. Psychiatry* **174**, 676–685 (2017).

589 21. Zilverstand, A., Huang, A. S., Alia-Klein, N. & Goldstein, R. Z. Neuroimaging
590 Impaired Response Inhibition and Salience Attribution in Human Drug Addiction: A
591 Systematic Review. *Neuron* **98**, 886–903 (2018).

592 22. Russo, S. J. & Nestler, E. J. The brain reward circuitry in mood disorders. *Nat. Rev.*
593 *Neurosci.* **14**, 609–625 (2013).

594 23. Zanos, P. & Gould, T. D. Mechanisms of Ketamine Action as an Antidepressant. *Mol.*
595 *Psychiatry* **23**, 801 (2018).

596 24. Duman, R. S., Sanacora, G. & Krystal, J. H. Altered connectivity in depression: GABA
597 and glutamate neurotransmitter deficits and reversal by novel treatments. *Neuron* **102**, 75–
598 90 (2019).

599 25. Hu, Y.-T., Tan, Z.-L., Hirjak, D. & Northoff, G. Brain-wide changes in excitation-
600 inhibition balance of major depressive disorder: a systematic review of topographic patterns
601 of GABA- and glutamatergic alterations. *Mol. Psychiatry* **28**, (2023).

602 26. Ren, Z. *et al*. Bidirectional Homeostatic Regulation of a Depression-Related Brain
603 State by Gamma-Aminobutyric Acidergic Deficits and Ketamine Treatment. *Biol. Psychiatry*
604 **80**, 457–468 (2016).

605 27. Biria, M. *et al*. Cortical glutamate and GABA are related to compulsive behaviour in
606 individuals with obsessive compulsive disorder and healthy controls. *Nat. Commun.* **14**,
607 3324 (2023).

608 28. Xie, C. *et al*. A shared neural basis underlying psychiatric comorbidity. *Nat. Med.* **29**,
609 1232–1242 (2023).

610 29. McTeague, L. M., Goodkind, M. S. & Etkin, A. Transdiagnostic impairment of
611 cognitive control in mental illness. *J. Psychiatr. Res.* **83**, 37–46 (2016).

612 30. Husain, M. & Roiser, J. P. Neuroscience of apathy and anhedonia: a transdiagnostic
613 approach. *Nat. Rev. Neurosci.* **19**, 470–484 (2018).

614 31. Ritter, C. *et al*. Evaluation of Prefrontal γ-Aminobutyric Acid and Glutamate Levels
615 in Individuals With Major Depressive Disorder Using Proton Magnetic Resonance
616 Spectroscopy. *JAMA Psychiatry* **79**, 1209–1216 (2022).

617 32. Marinkovic, K., Alderson Myers, A. B., Arienzo, D., Sereno, M. I. & Mason, G. F.
618 Cortical GABA levels are reduced in young adult binge drinkers: Association with recent
619 alcohol consumption and sex. *NeuroImage Clin.* **35**, 103091 (2022).

620 33. Prévot, T. & Sibille, E. Altered GABA-mediated information processing and cognitive
621 dysfunctions in depression and other brain disorders. *Mol. Psychiatry* **26**, 151–167 (2021).

622 34. Zhou, H. *et al*. Glutamate concentration of medial prefrontal cortex is inversely
623 associated with addictive behaviors: a translational study. *Transl. Psychiatry* **14**, 433 (2024).

624 35. Godfrey, K. E. M., Gardner, A. C., Kwon, S., Chea, W. & Muthukumaraswamy, S. D.
625 Differences in excitatory and inhibitory neurotransmitter levels between depressed patients

626 and healthy controls: A systematic review and meta-analysis. *J. Psychiatr. Res.* **105**, 33–44
627 (2018).

628 36. Sanacora, G., Treccani, G. & Popoli, M. Towards a glutamate hypothesis of
629 depression: An emerging frontier of neuropsychopharmacology for mood disorders.
630 *Neuropharmacology* **62**, 63–77 (2012).

631 37. Grabski, M. *et al.* Adjunctive Ketamine With Relapse Prevention–Based
632 Psychological Therapy in the Treatment of Alcohol Use Disorder. *Am. J. Psychiatry* **179**,
633 152–162 (2022).

634 38. Dubovsky, S. L. & Marshall, D. Benzodiazepines Remain Important Therapeutic
635 Options in Psychiatric Practice. *Psychother. Psychosom.* **91**, 307–334 (2022).

636 39. Forsyth, A. *et al.* Comparison of local spectral modulation, and temporal correlation,
637 of simultaneously recorded EEG/fMRI signals during ketamine and midazolam sedation.
638 *Psychopharmacology (Berl.)* **235**, 3479–3493 (2018).

639 40. Abdallah, C. G. *et al.* Ketamine Treatment and Global Brain Connectivity in Major
640 Depression. *Neuropsychopharmacology* **42**, 1210–1219 (2017).

641 41. Kraguljac, N. V. *et al.* Ketamine modulates hippocampal neurochemistry and
642 functional connectivity: a combined magnetic resonance spectroscopy and resting-state
643 fMRI study in healthy volunteers. *Mol. Psychiatry* **22**, 562–569 (2017).

644 42. Breakspear, M. Dynamic models of large-scale brain activity. *Nat. Neurosci.* **20**,
645 340–352 (2017).

646 43. Wang, H. *et al*. Virtual brain twins: from basic neuroscience to clinical use. *Natl. Sci.*
647 *Rev*. **11**, (2024).

648 44. Griffiths, J. D., Bastiaens, S. P. & Kaboodvand, N. Whole-Brain Modelling: Past,
649 Present, and Future. in *Computational Modelling of the Brain: Modelling Approaches to Cells,*
650 *Circuits and Networks* (eds. Giugliano, M., Negrello, M. & Linaro, D.) 313–355 (Springer
651 International Publishing, Cham, 2022). doi:10.1007/978-3-030-89439-9_13.

652 45. Hansen, J. Y. *et al*. Mapping neurotransmitter systems to the structural and functional
653 organization of the human neocortex. *Nat. Neurosci.* **25**, 1569–1581 (2022).

654 46. Brunel, N. Dynamics of sparsely connected networks of excitatory and inhibitory
655 spiking neurons. *J. Comput. Neurosci.* **8**, 183–208 (2000).

656 47. Swanson, C. J. *et al*. Metabotropic glutamate receptors as novel targets for anxiety
657 and stress disorders. *Nat. Rev. Drug Discov*. **4**, 131–144 (2005).

658 48. Puig, M. V., Watakabe, A., Ushimaru, M., Yamamori, T. & Kawaguchi, Y. Serotonin
659 Modulates Fast-Spiking Interneuron and Synchronous Activity in the Rat Prefrontal Cortex
660 through 5-HT1A and 5-HT2A Receptors. *J. Neurosci.* **30**, 2211–2222 (2010).

661 49. Celada, P., Puig, M. V. & Artigas, F. Serotonin modulation of cortical neurons and
662 networks. *Front. Integr. Neurosci.* **7**, 25 (2013).

663 50. William Moreau, A., Amar, M., Le Roux, N., Morel, N. & Fossier, P. Serotoninergic
664 Fine-Tuning of the Excitation–Inhibition Balance in Rat Visual Cortical Networks. *Cereb.*
665 *Cortex* **20**, 456–467 (2010).

666 51. Hasselmo, M. E. The role of acetylcholine in learning and memory. *Curr. Opin.*
667 *Neurobiol.* **16**, 710–715 (2006).

668 52. Wilson, R. I. & Nicoll, R. A. Endocannabinoid signaling in the brain. *Science* **296**,
669 678–682 (2002).

670 53. Kandel, E. R., Koester, J. D., Mack, S. H. & Siegelbaum, S. A. *Principles of Neural*
671 *Science, Sixth Edition*. (McGraw Hill LLC, 2021).

672 54. Abbott, L. F. Lapicque's introduction of the integrate-and-fire model neuron (1907).
673 *Brain Res. Bull.* **50**, 303–304 (1999).

674 55. Friston, K. J., Mechelli, A., Turner, R. & Price, C. J. Nonlinear responses in fMRI: the
675 Balloon model, Volterra kernels, and other hemodynamics. *NeuroImage* **12**, 466–477 (2000).

676 56. Isaacson, J. S. & Scanziani, M. How Inhibition Shapes Cortical Activity. *Neuron* **72**,
677 231–243 (2011).

678 57. Deco, G. *et al.* How Local Excitation–Inhibition Ratio Impacts the Whole Brain
679 Dynamics. *J. Neurosci.* **34**, 7886–7898 (2014).

680 58. Haider, B. & McCormick, D. A. Rapid neocortical dynamics: cellular and network
681 mechanisms. *Neuron* **62**, 171–189 (2009).

682 59. Yu, Y. *et al.* A 3D atlas of functional human brain energetic connectome based on
683 neuropil distribution. *Cereb. Cortex* **33**, 3996–4012 (2023).

684 60. Baroni, F. & Fulcher, B. Synchrony, oscillations, and phase relationships in collective
685 neuronal activity: A highly comparative overview of methods. *PLOS Comput. Biol.* **21**,
686 e1013597 (2025).

687 61. Suzuki, A., Hara, H. & Kimura, H. Role of the AMPA receptor in antidepressant
688 effects of ketamine and potential of AMPA receptor potentiators as a novel antidepressant.
689 *Neuropharmacology* **222**, 109308 (2023).

690 62. Dakwar, E. *et al*. A Single Ketamine Infusion Combined With Motivational
691 Enhancement Therapy for Alcohol Use Disorder: A Randomized Midazolam-Controlled Pilot
692 Trial. *Am. J. Psychiatry* **177**, 125–133 (2020).

693 63. Riss, J., Cloyd, J., Gates, J. & Collins, S. Benzodiazepines in epilepsy: pharmacology
694 and pharmacokinetics. *Acta Neurol. Scand.* **118**, 69–86 (2008).

695 64. Haber, S. N. & Knutson, B. The Reward Circuit: Linking Primate Anatomy and
696 Human Imaging. *Neuropsychopharmacology* **35**, 4–26 (2010).

697 65. Buckner, R. L., Andrews-Hanna, J. R. & Schacter, D. L. The brain's default network:
698 anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* **1124**, 1–38 (2008).

699 66. Cole, M. W. *et al*. Multi-task connectivity reveals flexible hubs for adaptive task
700 control. *Nat. Neurosci.* **16**, 1348–1355 (2013).

701 67. Power, J. D. *et al*. Functional network organization of the human brain. *Neuron* **72**,
702 665–678 (2011).

703 68. Lynch, C. J. *et al*. Frontostriatal salience network expansion in individuals in
704 depression. *Nature* **633**, 624–633 (2024).

705 69. Reid, A. T. *et al*. Advancing functional connectivity research from association to
706 causation. *Nat. Neurosci.* **22**, 1751–1760 (2019).

70. Friston, K. J. Functional and Effective Connectivity: A Review. *Brain Connect.* **1**, 13–36 (2011).

71. Jirsa, V. *et al.* Personalised virtual brain models in epilepsy. *Lancet Neurol.* **22**, 443–454 (2023).

72. Wang, P. *et al.* Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain. *Sci. Adv.* **5**, eaat7854 (2019).

73. Stringaris, A. *et al.* The Brain's Response to Reward Anticipation and Depression in Adolescence: Dimensionality, Specificity, and Longitudinal Predictions in a Community-Based Sample. *Am. J. Psychiatry* **172**, 1215–1223 (2015).

74. Cole, M. W., Ito, T., Cocuzza, C. & Sanchez-Romero, R. The Functional Relevance of Task-State Functional Connectivity. *J. Neurosci.* **41**, 2684–2702 (2021).

75. Ozeki, H., Finn, I. M., Schaffer, E. S., Miller, K. D. & Ferster, D. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron* **62**, 578–592 (2009).

76. Murphy, B. K. & Miller, K. D. Balanced Amplification: A New Mechanism of Selective Amplification of Neural Activity Patterns. *Neuron* **61**, 635–648 (2009).

77. Joglekar, M. R., Mejias, J. F., Yang, G. R. & Wang, X.-J. Inter-areal Balanced Amplification Enhances Signal Propagation in a Large-Scale Circuit Model of the Primate Cortex. *Neuron* **98**, 222-234.e8 (2018).

78. Murray, J. D. *et al.* Linking Microcircuit Dysfunction to Cognitive Impairment: Effects of Disinhibition Associated with Schizophrenia in a Cortical Working Memory Model. *Cereb. Cortex* **24**, 859–872 (2014).

79. Yizhar, O. *et al.* Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* **477**, 171–178 (2011).

80. Ahmadian, Y. & Miller, K. D. What is the dynamical regime of cerebral cortex? *Neuron* **109**, 3373–3391 (2021).

81. Bradley, C., Nydam, A. S., Dux, P. E. & Mattingley, J. B. State-dependent effects of neural stimulation on brain function and cognition. *Nat. Rev. Neurosci.* **23**, 459–475 (2022).

Figures

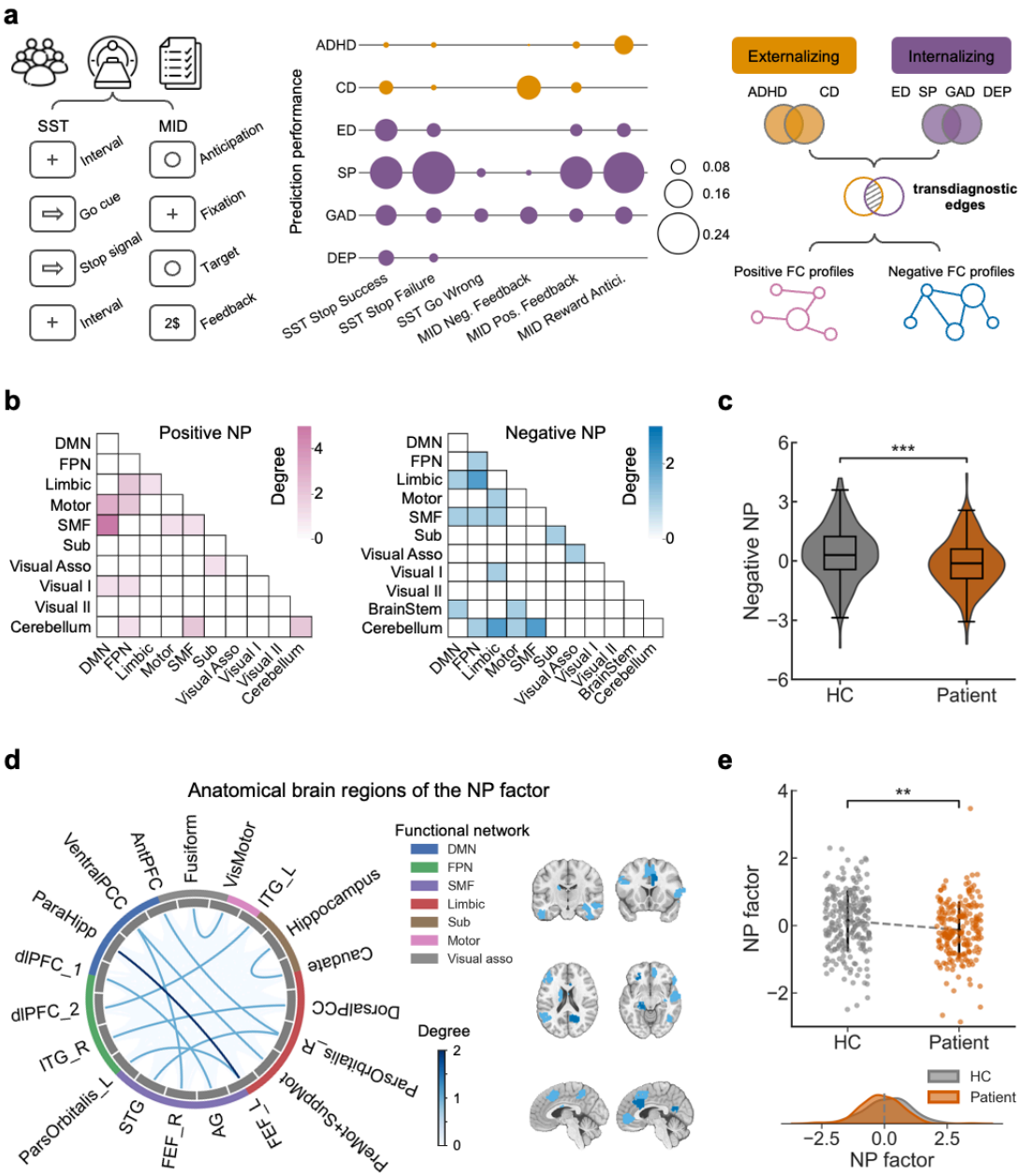

Figure 1 Identification of a compact functional network phenotype of transdiagnostic psychopathology.

a, Task-based fMRI data and behavioural symptom scores from the IMAGEN cohort (n = 1,050) were analysed using connectome-based predictive modelling (CPM) to identify task-

evoked functional connectivity (FC) profiles associated with transdiagnostic psychopathology. Predictive performances of task-specific FC derived from the Stop-Signal Task (SST) and Monetary Incentive Delay (MID) were shown in a bubble plot, where bubble size reflects Spearman's correlation between predicted and empirical symptom scores. The mean *P* values across 1,000 repetitions were corrected for multiple comparisons across 36 models using false discovery rate (FDR) correction, corresponding to a minimum significant effect size of Spearman's $\rho = 0.07$. Task-specific functional connectivity (FC) matrices were then selected if they showed significant prediction performance for at least three behavioural symptom domains. SST conditions included stop-success, stop-failure, and go-wrong; MID conditions included positive feedback, negative feedback, and reward anticipation. Behavioural measures comprised two externalising domains (attention-deficit/hyperactivity disorder, ADHD, and conduct disorder, CD) and four internalising domains (eating disorder, ED, specific phobia, SP, generalized anxiety disorder, GAD, and depression, DEP). A transdiagnostic FC profile was identified that significantly predicted both externalising and internalising symptoms. These transdiagnostic associated edges were divided into positive (pink) or negative (blue) FC profiles based on the direction of their associations with behavioural symptoms**. b,** Heatmaps show the canonical functional network affiliations of positive and negative transdiagnostic FC profiles. Positively predictive connections were primarily distributed across the default mode network (DMN), frontoparietal network (FPN), somatomotor–frontal (SMF) network, and cerebellum, whereas negatively predictive connections were concentrated in limbic, DMN, FPN, SMF, and cerebellar systems. **c,** Negative NP scores (summed strength of the negative FC profile) were compared between healthy controls (HC, n = 225) and patients (n = 209) from the STRATIFY cohort. Group difference was assessed using an independent-samples *t* test with Bonferroni correction (t =

3.94, $P_{corrected}$ = 0.0003). Values represent residuals after regressing out sex, scanning site, and head motion (mean framewise displacement). Violin plots show kernel density estimates with embedded boxplots indicating the median and interquartile range (IQR); whiskers extend to 1.5× IQR. **d,** After excluding cerebellar and brainstem connections, the negative FC profile was reduced to a compact 12-edge NP circuit comprising cortico–subcortical connections. Edges in the circos plot represent connection strength and negative associations with psychopathology; corresponding brain regions are shown on cortical and subcortical renderings. **e,** Individual summed 12-edge-NP factor scores were compared between HC and patients from the STRATIFY cohort. Group difference was confirmed using an independent-samples *t* test with Bonferroni correction (t = 3.27, $P_{corrected}$ = 0.0036). Scores represent residuals after regressing out sex, scanning site, and head motion. Each dot denotes one participant; black markers and vertical lines indicate group means and standard deviations, respectively. Kernel density estimates illustrate the distributional shift in NP factor scores between groups. **, $P_{corrected} < 0.01$; ***, $P_{corrected} < 0.001$.

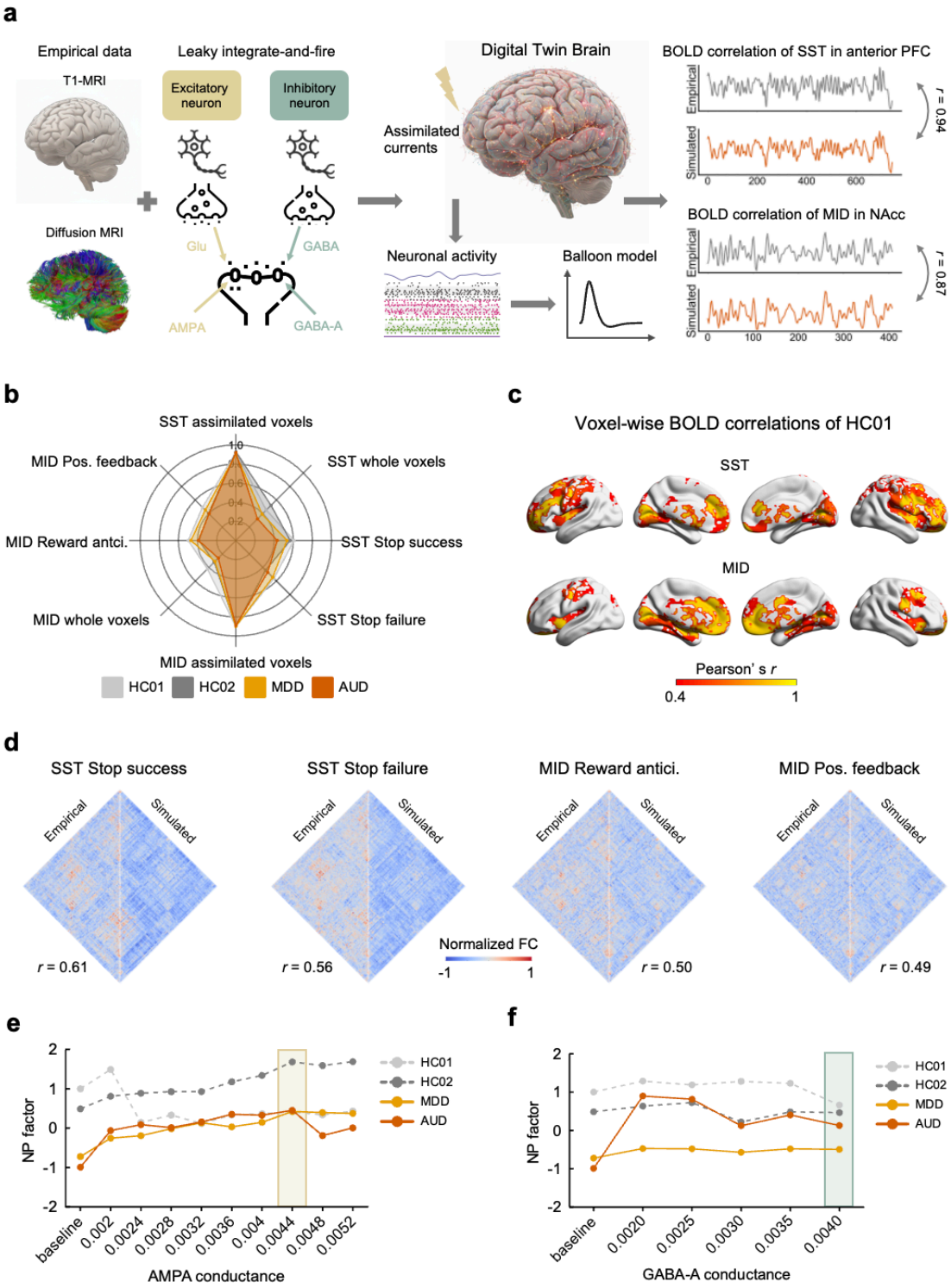

Figure 2 Individualised digital twin brains (DTBs) reveal excitation/inhibition-sensitive NP modulation.

a, Computational workflow for constructing individualised DTBs from multi-modal participant-specific MRI data. DTB model received assimilation hyperparameters as external current inputs and generated simulated neuronal firing rates, which were converted to BOLD signals using Balloon model. The simulation performance was assessed by Pearson's *r* between empirical and simulated BOLD signals. **b,** Radar plots showing simulation performances of 1-billion-neuron models for two patients (one MDD; one AUD) and two healthy controls (HC01; HC02). For each task, both the mean voxel-wise Pearson's correlations in assimilated voxels (assimilated) and the mean correlations across the whole voxels (whole). Task-specific FC matrices included stop-success and stop-failure for the SST, and reward anticipation and positive feedback for the MID task. **c,** Voxel-wise Pearson's correlations between empirical and simulated BOLD signals from the 1-billion-neuron DTB model for the MID and SST tasks of HC01. Bright yellow regions indicate higher similarity. **d,** Empirical and simulated FC matrices across four task conditions of SST and MID from the 1-billion-neuron DTB model of the HC01. **e–f,** In silico E/I modulation sweeps. Curves depict NP factor responses to systematic changes in AMPA (**e**) and GABA-A (**f**) synaptic conductance. For AMPA modulation, NP factor scores increased in patients and HC02 but decreased in HC01. For GABA-A modulation (with AMPA = 0.0044, the parameter yielding maximal NP factor improvement), the NP factor scores increased in two patients, while slightly decreased in HCs. The shaded areas indicate optimal E/I parameters for future perturbations (AMPA = 0.0044; GABA-A = 0.0040). **, $P_{corrected} < 0.01$.

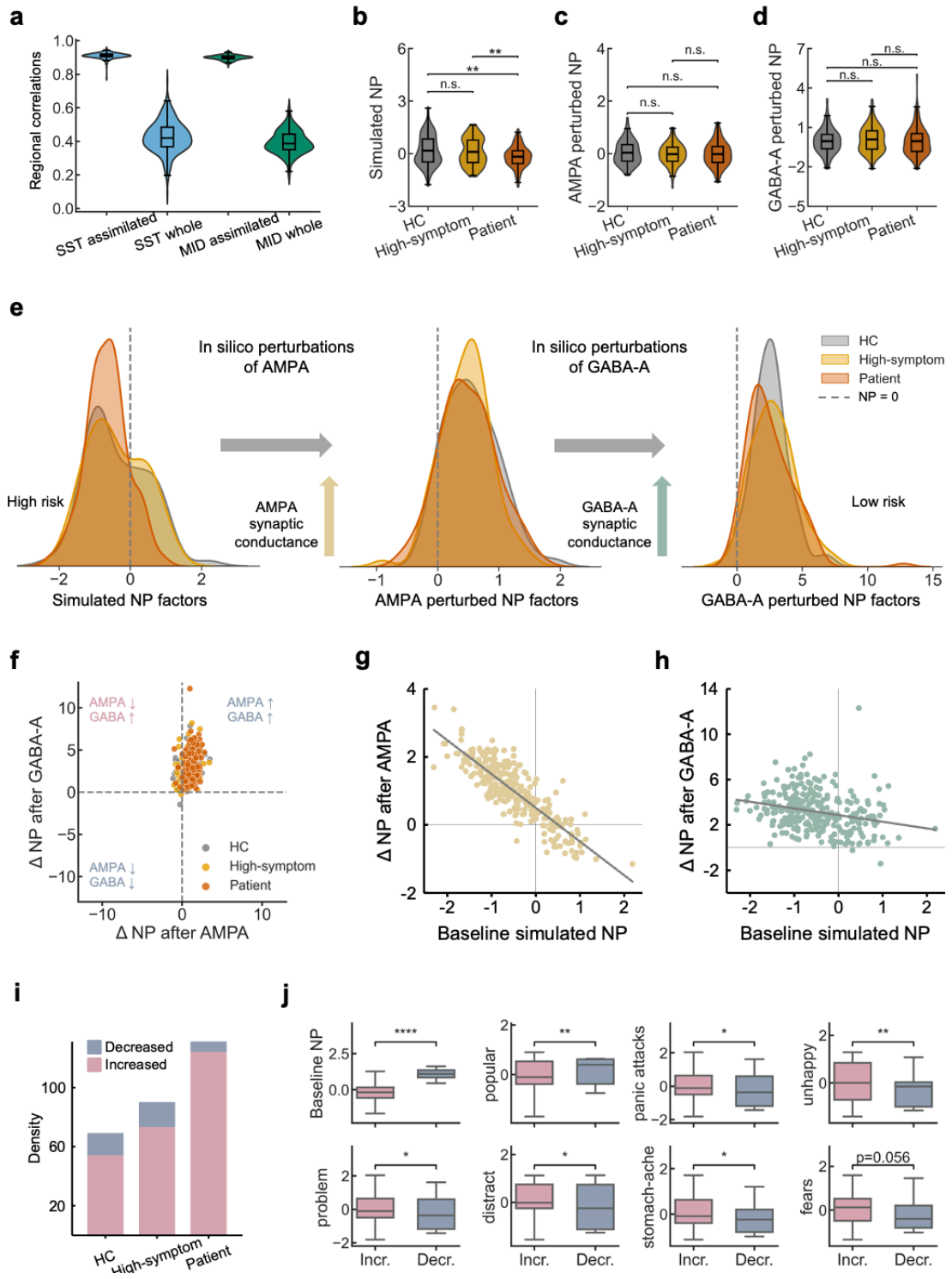

Figure 3 Population-scale in silico perturbations and mechanistically grounded stratification.

a, Regional mean Pearson's correlations between empirical and simulated BOLD signals during SST and MID tasks from 3-million-neuron DTB models for 290 participants. For each task, Violin plots show both the mean correlations within assimilated task-related regions (assimilated) and the mean correlations across all regions (whole). The boxplots indicating the median and IQR; whiskers extend to 1.5× IQR. **b-d,** Simulated NP factor scores under baseline (**b**), AMPA-mediated (**c**), and GABA-A-mediated (**d**) virtual perturbations in healthy controls (HC, n = 69), high-symptom individuals (n = 90), and patients (n = 131). Group differences were assessed using ANOVA and post-doc tests with Bonferroni correction for multiple comparisons (n.s., not significant; **, $P_{corrected} < 0.01$). Values are residuals after regressing out sex, site, and head motion. Violin plots show kernel density estimates with embedded boxplots indicating the mean and IQR; whiskers extend to 1.5× IQR. **e,** Kernel density distributions of NP factor scores across healthy controls (HC, grey), high-symptom individuals (yellow), and patients (orange) under three simulation conditions: (left) baseline, (middle) AMPA modulation, and (right) GABA-A modulation. The x-axis represents the simulated NP factor scores, and shaded areas denote the distribution density for each group. Across the three panels, the degree of overlap between group distributions and the relative shift of scores from negative to positive values correspond to the transition from baseline to specific receptor manipulations. **f,** Individual changes in NP factor scores following the AMPA and GABA-A perturbations. Dashed lines indicate zero change. Each dot represents an individual subject. Individuals with ΔNP>0 (perturbed-baseline) under both perturbations were classified as the increased-response group (n = 251, pink); others formed the decreased-response group (n = 39, blue). **g-h,** Baseline NP factor scores were inversely correlated with AMPA- (**g**, Pearson's $r = -0.86$, $P < 0.0001$) and GABA-A- (**h**, Pearson's $r = -0.25$, $P < 0.0001$) induced changes, with a permutation test of 10,000 times. **i,** Distribution of response types

across groups (Chi-square test: χ^2 = 14.05, P = 0.007). Increased-response proportions: patients, 124/131 (94.66%); high-symptom, 73/90 (81.11%); HC, 54/69 (78.26%). **j,** The increased-response group exhibited lower baseline NP factors, and higher symptom burden, including panic attacks, negative affect and reduced social support, than the decreased-response group (Mann–Whitney U tests, controlling for sex, site, and head motion; *, P<0.05; **, P<0.01; ****, P<0.0001). Boxplots represent the median and IQR; whiskers extend to 1.5× IQR.

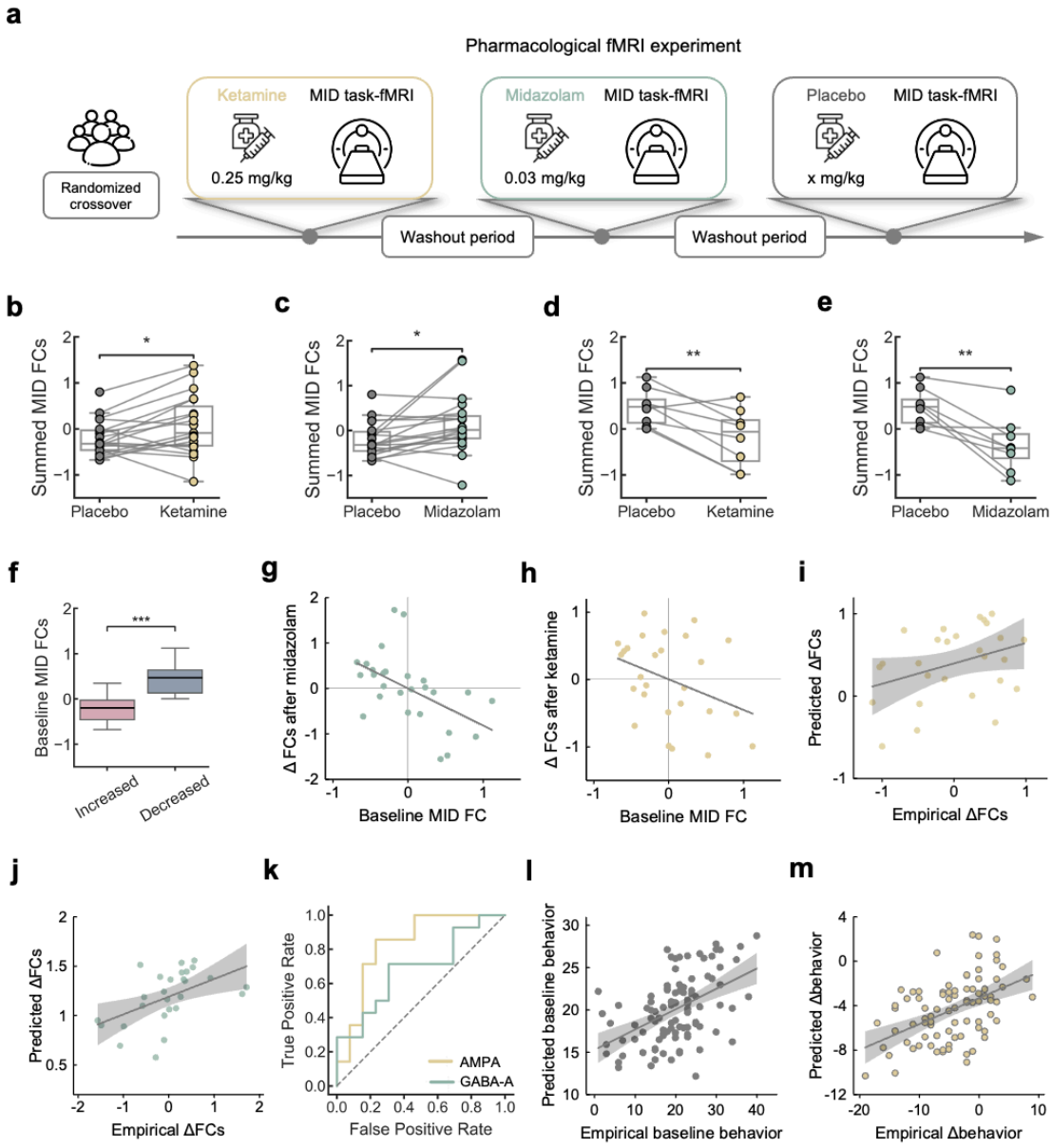

Figure 4 Pharmacological fMRI validation and prospective symptom prediction.

a, Pharmacological fMRI experiment in 27 male healthy volunteers using a three-way randomized, controlled crossover design. **b-e,** Individual- and group-level changes in summed NP-related MID functional connectivity (summed MID FCs) following ketamine and midazolam administration in subgroup 1 (**b**, **c**, n = 19) and subgroup 2 (**d**, **e**, n = 8). Each dot represents one individual and is colored by drug condition: placebo (grey), ketamine

847 (gold), and midazolam (teal). Statistical significance was assessed using paired *t*-tests or
848 Wilcoxon signed-rank tests, as appropriate. Values represent residuals after regressing out
849 head motion. **f,** Subgroup 1, exhibiting drug-induced increased responses, showed lower
850 baseline summed MID FCs under placebo compared with subgroup 2 (Mann–Whitney U
851 tests, controlling for head motion; ***, $P<0.001$). **g-h,** Baseline summed MID FC was
852 inversely correlated with midazolam-induced change (**g**, Pearson's r = -0.52, 10,000
853 permutations, P = 0.0056), and showed a similar trend for ketamine-induced change (**h**,
854 Pearson's r = -0.35, 10,000 permutations, P = 0.069). **i-j,** Prediction of empirical drug-
855 induced FC changes from baseline summed MID FC using linear mappings between DTB-
856 simulated E/I perturbations and simulated baseline MID FC strength. Scatter plots show
857 predicted versus observed FC changes induced by ketamine (**i**, Pearson's r = 0.35, 1,000
858 permutations, P = 0.039) and midazolam (**j**, Pearson's r = 0.52, 1,000 permutations, P =
859 0.001). Each dot represents one individual; solid lines indicate linear regression fits with 95%
860 confidence intervals. **k,** Prediction of drug-induced response direction (ΔNP>0 or <0) from
861 baseline summed MID FCs using DTB-derived models (AMPA model AUC = 0.69; GABA-A
862 model AUC = 0.83). **l,** Linear association between empirical NP factor scores and
863 behavioural symptom severity at age 19 (n = 100, r^2 = 0.24, F = 2.3, P = 0.014). Each dot
864 represents one individual; solid lines indicate linear regression fits with 95% confidence
865 intervals. **m,** Prospective prediction of behavioural symptom changes over 4 years. Scatter
866 plot shows predicted versus empirical behavioural changes from age 19 to 23. The prediction
867 model explained a significant proportion of variance (n = 85, r^2 = 0.23, F = 4.80, permutation
868 test of 5,000 times, P = 0.0006). Each dot represents one individual; solid line indicates the
869 linear fit with 95% confidence intervals. All boxplots show the median and interquartile
870 range (IQR); whiskers extend to 1.5× IQR. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Methods

Participants and MRI dataset.

IMAGEN dataset

The discovery dataset used to idenitify brain phenotype of transdiagnostic symptoms was derived from the IMAGEN cohort (https://www.imagen-project.org/), a longitudinal, population-based study comprising behavioural, neuroimaging, environmental, and genetic data from approximately 2,000 participants assessed at ages 14, 19, and 23[1,2]. IMAGEN participants were recruited from eight research sites across the United Kingdom, Germany, France, and Ireland, and were cognitively normal with no history of psychiatric diagnoses. To reduce brain developmental confounds, we focused on early adulthood, specifically the follow-up 2 timepoint (aged 19y) of the IMAGEN study. At this time point, task-fMRI data were available for the Monetary Incentive Delay (MID, n = 1,390) and Stop Signal Task (SST, n = 1,394). After excluding participants with excessive head motion (mean framewise displacement > 0.5 mm) and ensuring matched multimode data, the final sample included 1,050 participants (mean age = 18.41 ± 0.67 years; female/male = 564/486).

The task-fMRI paradigms used in this study are summarized below. The MID task comprised 42 trials across three reward conditions (no-win, small-win, and big-win). Each trial consisted of a reward cue (250 ms), a fixation period (4-4.5 s), a target response, and reward feedback (1,450 ms). The SST included 480 go trials (respond to arrow direction) and 80 stop trials (inhibit response when a stop signal appeared ~300 ms after the go cue). The Emotional Face Task (EFT) presented angry, neutral, and happy human faces in a passive-viewing paradigm without responses. Each emotional condition comprised four trials, with each trial lasting 18 s. More details were available in the original paper of IMAGEN study [2].

894 *STRATIFY dataset*

895 We validated brain phenotype of transdiagnostic symptoms in independent samples from
896 the STRATIFY cohorts[3], which used identical MRI protocols and behavioural assessments to
897 IMAGEN. These datasets included patients with psychiatric diagnoses and demographically
898 matched healthy controls. After the same screening procedure with IMAGEN, the final
899 sample included 209 patients (age = 22.26 ± 2.21; female/male = 135/74) and 225 healthy
900 controls (age = 21.40 ± 1.72; female/male = 129/96). Specific diagnoses included alcohol
901 use disorder (AUD, female/male = 58/40), major depressive disorder (MDD, female/male =
902 75/29), psychosis (female/male = 2/4), and attention-deficit/hyperactivity disorder (ADHD,
903 female/male = 0/1) for STRATIFY.

904 Clinical diagnoses were determined based on established thresholds from standardized
905 assessment tools consistent with DSM-5 criteria. Individuals with AUD were defined by an
906 Alcohol Use Disorders Identification Test (AUDIT)[4] score greater than 15. MDD was defined
907 by a Patient Health Questionnaire-9 (PHQ-9)[5] score greater than 15. Psychosis was defined
908 using ICD-10 diagnosis, with inclusion requiring at least one episode of schizophreniform
909 or chronic schizophrenia. Healthy controls were screened using rigorous inclusion criteria
910 to minimize confounding factors. Participants were excluded if they had: a) any current or
911 past mental health disorder[6], b) first- or second-degree relatives with mental health diagnoses,
912 c) regular use of medication for serious physical health conditions, d) learning difficulties
913 (e.g., Wechsler Adult Intelligence Scale verbal comprehension score below the 10th
914 percentile), or e) any history of recreational drug use.

915 *Pharmacological dataset*

To validate the effects of in sillco neurotransmitter manipulation in vivo, we utilized pharmacological task-fMRI data[7] from 30 healthy male participants (mean age 27.3 ± 6.2 years). Each participant underwent three separate scanning sessions in a randomized, single-blind, placebo-controlled, three-way crossover design, receiving ketamine, midazolam, or placebo. A minimum 48-hour interval was maintained between sessions to prevent carryover effects. The experimental procedure included a 16-minute resting-state fMRI scan, during which drug administration began at the 7-minute mark, followed by task-fMRI acquisition during the MID task. Three participants were excluded due to missing MID behavioural recording files (n = 2) and T1-weighted structural MRI data (n = 1). The MID task used here comprised 90 trials across two conditions: no-win (45 trials) and win (45 trials). Details on administration protocols are available in the original study[7].

All above studies received approval from the local ethics committee, and written informed consent was obtained from all participants.

Behavioural symptoms assessments.

We assessed behavioural symptoms for each individual from IMAGEN and STRATIFY cohorts using the Development and Well-Being Assessment[8] (DAWBA) and Strengths and Difficulties Questionnaire[9] (SDQ). To characterise externalising and internalising symptoms, we summed the scores of relevant items from DAWBA and SDQ[3]. The externalising symptoms included ADHD (8 items) and conduct disorder (CD, 7 items), while the internalising symptoms encompassed eating disorder (ED, 5 items), specific phobia (SP, 13 items), general anxiety disorder (GAD, 7 items), and depression (DEP, 8 items). A detailed list of the specific items is provided in Table S2.

MRI acquisition and preprocessing.

939 The structural and functional MRI protocols for the IMAGEN and STRATIFY studies were
940 harmonised across sites and scanner manufacturers (Siemens: 6 sites, Philips: 2 sites, General
941 Electric: 1 site, and Bruker: 1 site), Standardized hardware for visual and auditory stimulus
942 presentation (Nordic Neurolabs, Norway) was used at all sites.

943 *Structural T1-weighted images*

944 T1-weighted images (T1-w) of the IMAGEN and STRATIFY cohorts were acquired using a
945 protocol based on those from the ADNI study (https://adni.loni.usc.edu/data-samples/adni-
946 data/neuroimaging/mri/mri-scanner-protocols/). Several parameters in these protocols
947 deliberately differ between scanner models, in order to produce consistent image contrast
948 and quality despite implementation differences between manufacturers; all scans were
949 however collected with a sagittal slice plane and voxel size = 1.1 × 1.1 ×1.1 mm^3.

950 The structural T1-w images were preprocessed using fMRIPrep[10] (version 20.2.3), a robust
951 and standardized pipeline based on Nipype, supported by Advanced Normalization Tools
952 (ANTs, version 2.3.3), FMRIB Software Library (FSL, version 5.0.9), Analysis of Functional
953 NeuroImages (AFNI, version 20160207), and FreeSurfer (6.0.1) packages. Anatomical
954 preprocessing included bias field correction, skull-stripping, tissue segmentation and
955 nonlinear registration to standard Montreal Neurological Institute (MNI) space.

956 *Diffusion-weighted images*

957 Diffusion-weighted images (DWIs) of IMAGEN and STRATIFY cohorts were acquired using
958 a 2D diffusion weighted EPI sequence: TR = 15,000 ms; TE = 104 ms; FOV = 307 × 307
959 mm^2; number of slices = 60; scan time = 9 min 45 s; voxel size = 2.4 × 2.4 × 2.4 mm^3; 36
960 optimal non-colinear diffusion-weighted directions with b = 1,300 s/mm^2.

The structural DWIs were preprocessed using the FSL and MRtrix3 package. Eddy current-induced distortions and head motion were corrected using FSL's eddy tool, and the rotated b-vectors were updated using eddy_rotated_bvecs. The preprocessed DWI data was then converted to the MRtrix3 format (.mif) with the mrconvert tool. The mean b0 image was extracted using the mrmath command (mean), followed by brain extraction with the bet tool (fractional intensity threshold = 0.3) to generate a brain mask. The white matter fiber orientation distributions (FODs) within each voxel were estimated using constrained spherical deconvolution (CSD), with response function estimation performed using the Tournier method. The anatomical T1-weighted image was registered to the native diffusion space using FSL's flirt tool. A nonlinear transformation to MNI space was then computed using FSL's fnirt, and the inverse warp was generated with invwarp. The resulting warp was subsequently converted into an MRtrix-compatible format for fiber transformation. A five-tissue-type segmentation was carried out with the 5ttgen command. The gray matter-white matter interface (GMWMI) was extracted using the 5tt2gmwmi tool to serve as the seeding region for tractography. Whole-brain fiber tracking was performed using the tckgen command with the iFOD2 algorithm, seeding from the extracted GMWMI and applying Anatomically Constrained Tractography (ACT). The generated fiber tracks were transformed from diffusion space to MNI space and generate voxel-wise and regional-level structural connectivity matrices.

Resting-state and task-state functional MRI images

The resting-state fMRI scans of IMAGEN and STRATIFY cohorts were acquired using a 2D gradient echo EPI sequence with the following parameters: TR = 2,200 ms; TE = 30 ms; FA = 75°; FOV = 220 × 220 mm^2; number of slices = 40; number of volumes = 164; and voxel size = 3.4 × 3.4 × 2.4 mm^3. Task-state fMRI scans were conducted with identical scanning

parameters, except for the number of volumes, that is 191 volumes for the MID task, 349 volumes for SST task, and 202 volumes for EFT task .

Addtionally, the pharamacological MID task-state fMRI were acquired on a 3T scanner (Siemens Skyra, Erlangen, Germany), using the EPI sequence: TR = 2,200 ms; TE = 27 ms; FA = 79°; FOV = 215 × 215 mm^2; number of slices = 30; number of volumes = 410; and voxel size = 3 × 3 × 3 mm^3.

All fMRI data were preprocessed using fMRIPrep (version 20.2.3), including motion correction, slice timing correction, and susceptibility distortion correction using fMRIPrep's fieldmap-less approach. Functional MRI images were co-registered to the T1-weighted image using boundary-based registration and then normalized to MNI152 space. Confound regressors were extracted to control for physiological and motion-related artifacts, including framewise displacement, DVARS, and CompCor components. All transformations were applied in a single interpolation step to minimize resampling effects. Then the preprocessed data underwent temporal detrending to remove low-frequency drifts, followed by spatial smoothing using a full-width at half maximum (FWHM) of 6 mm Gaussian kernel and a band-pass filtering (0.01 ~ 0.1 Hz). Quality control was performed by excluding subjects with incomplete demographic or scanning timepoints; excessive head motion (mean framewise displacement > 0.5 mm); or unsuccessful spatial normalization.

Identification of a compact network phenotype of psychopathology in IMAGEN.

Followed by previous work identifing the neuropsychopathology (NP) factor in the IMAGEN baseline at age 14[3], we constructed MID and SST task-specific functional connectivity (FC) matrics using the CONN toolbox (version 16.h) and a well-established 268-node whole-brain functional parcellation[11], applying weighted generalised linear models after regressing

out task condition regressors and nuisance variables. Subsequently, we employed a Connectome-based Predictive Modeling (CPM) [12] with a 50-fold cross-validation to predict each behavioural symptom scores (including ADHD conduct disorder, eating disorder, specific phobia, general anxiety disorder, and depression) from each task-specific functional connectomes (SST conditions included stop-success, stop-failure, and go-wrong; MID conditions included positive feedback, negative feedback, and reward anticipation). The model performance was assessed using the Spearman's correlation between predicted and empirical behavioural symptoms scores. This procedure was repeated 1,000 times, and only edges selected in over 95% of models were used for further analysis. The mean *P* values across 1,000 repetitions were corrected for multiple comparisons across 36 predictive models using false discovery rate (FDR) correction ($q < 0.05$). Task-specific functional connectivity (FC) matrices were then selected if they showed statistically significant prediction performance for at least three behavioural symptom domains, ensuring robustness across multiple diagnostic categories.

From these selected task-specific FC matrices, we extracted edges that significantly predicted both externalising and internalising symptoms. Edges that were consistently associated with behavioural symptoms across all selected task conditions were defined as transdiagnostic associated edges. To improve interpretability, we stratified these transdiagnostic associated edges into positive or negative FC profiles based on the direction of their associations with behavioural symptoms. The summed strengths of these two profiles were termed as positive and negative NP scores, respectively.

Case-control comparison of NP factors in STRATIFY.

To test whether the NP scores generalise to clinical populations, we re-extracted these transdiagnostic associated FC profiles in the independent STRATIFY dataset and re-calculated NP scores for each individual. We then compared NP scores between patients and healthy controls, as well as among depression and alcohol use disorder subgroups, using independent-samples *t* tests while covarying for sex, site, and head motion (mean framewise displacement). Multiple comparisons were corrected using the Bonferroni method.

Permutation testing of PET receptor maps enrichment in NP-related regions.

To assess the neurochemical context of the NP factor, we examined whether the NP-related regions exhibited higher receptor map values relative to non-NP regions using published whole-brain PET-derived receptor density maps[13] (https://github.com/netneurolab/hansen_receptors/tree/main/data/PET_nifti_images). For a given PET map, we first extracted regional expressive levels using the 268-node functional parcellation[11], and then computed the empirical difference in mean values between NP (n = 32) and non-NP (n = 236) regions. Statistical significance was assessed using permutation testing (10,000 times), in which region labels (NP-related vs NP-unrelated) were randomly reassigned while preserving the original group sizes. For each permutation, the difference in mean receptor values between the permuted groups was recomputed, yielding a null distribution expected under no spatial specificity. Two-sided permutation *P* values were calculated as the proportion of permuted differences whose absolute value exceeded the empirical difference. Resulting *P* values were corrected for multiple comparisons across 39 receptor maps using FDR correction.

Multimode neuroimaging data for Digital Twin Brain (DTB) models.

To constructing individualised DTB models, we extracted voxel-wise multimodel neuroimaging data, including grey matter volume, white matter structural connectivity, and functional BOLD signals from resting-state and task-state fMRI.

Gray matter volume was estimated using voxel-based morphometry (VBM) in SPM12 (Matlab R2020b). T1-weighted images were segmented into tissue maps, normalized to MNI space using DARTEL, modulated by Jacobian determinants, and smoothed with an 8 mm FWHM Gaussian kernel at $3 \times 3 \times 3$ mm^3 resolution.

White matter structural connectivity was derived from whole-brain tractography using the iFOD2 algorithm in MRtrix3, with anatomical constraints (ACT). Tracking used 5 million streamlines seeded from the GMWMI, with a step size of 0.2 mm, curvature threshold of 45°/step, and length range of 3–250 mm.

To extract BOLD signals, we first generated individualised brain mask by selecting voxels that: a) overlapped with both the 268-node functional parcellation[11] and the MNI152 template, b) had existing white matter structural connections, and c) were located in the cortex or subcortex. On average, ~12,000 voxels per subject were retained. The time series from these voxels were then extracted as simulation targets for the DTB models.

Importantly, incorporating specific neuronal models for the cerebellum and brainstem would substantially increase model complexity and computational cost, particularly given that the DTB simulations are implemented at the scale of hundreds of millions to billions of neurons. We therefore adopted a parsimonious modeling strategy focused on cortical and subcortical circuits, while acknowledging that this exclusion represents a limitation and that future extensions incorporating differentiated regional parameterization may enable more comprehensive characterization of NP network dynamics.

Construction of individualised task-state DTB models.

Followed by the framework developed by Lu et al.[14,15], we firstly constructed a excitation-inhibition balanced spiking neural network constrained by individual multimode neuroimaging data. In this network, each voxel was modeled as a sub-unit comprising one excitatory and one inhibitory neuronal population. Each voxel included both excitatory and inhibitory neurons at a fixed ratio of 4:1[16]. The number of neurons within each voxel was determined by its proportional grey matter volume. Within-voxel synaptic architecture followed a 4:1:2 ratio of internal excitatory, internal inhibitory, and external excitatory synapses, based on anatomical organization of the cat visual cortex[17]. Between-voxel connections were weighted by row-normalized structural connectivity derived from diffusion MRI, and exclusively excitatory[18,19]. Neuronal activity was simulated using the leaky integrate-and-fire model[20], incorporating α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic-acid (AMPA; excitatory-leaning) and γ-aminobutyric-acid-A (GABA-A; inhibitory-leaning) synaptic conductances to represent glutamatergic and GABAergic signaling:

$$C_i \frac{dV_i}{dt} = -g_{L,i}(V_i - V_L) + \sum_u I_{syn,i} + I_{bg,i} + I_{ext,i},\ V_i < V_{th,i}$$

where C_i is the capacitance of the neuron membrane, $g_{L,i}$ is the leakage conductance, V_i is the membrane potential of neuron i, V_L is the leakage voltage, $I_{syn,i}$ is the synaptic currents of two synapse types (AMPA and GABA-A), $I_{bg,i}$ is the background current as noises, and $I_{ext,i}$ is the external current input for the task. It should be noted that the $I_{ext,i}$ are the independent parameters, which are estimated using empirical task-fMRI data. The background noise is given by independent Ornstein–Uhlenbeck processes and described as in the original paper[15].

Simulated neuronal activity was quantified as mean firing rate of neuronal population within each voxel, and was transformed into simulated BOLD signals using the Balloon-Windkessel model[21]. The model performance was quantified by Pearson's correlation between simulated and empirical task-state BOLD signals.

To optimize the fit between simulated and empirical BOLD signals, we applied the Hierarchical Mesoscale Data Assimilation (HMDA) method[15] to estimate voxel-wise, time-dependent hyperparameters. This assimilation procedure iteratively alternated between forward simulating neuronal activity and parameter updating to reduce the mismatch between simulated and observed BOLD signals. At each time point for each assimilated voxel, the parameters of synaptic conductances for AMPA and GABA-A were sampled from gamma distributions, which were constrained by voxel-specific hyperparameters. These synaptic conductances governed the simulated neuronal dynamics, which were transformed into BOLD signals. Hyperparameters were then iteratively updated so that the resulting simulated BOLD signals progressively converged toward the empirical fMRI data. Parameter updating was implemented using a diffusion ensemble Kalman filter (EnKF), which estimates latent states and parameters by propagating an ensemble of 30 parallel model realizations. More methodological details of HMDA are describedin Lu et al.'s original paper[15].

Resting-state digital twin brain models were first obtained by assimilating empirical resting-state fMRI BOLD signals for all voxels using the HMDA framework. This procedure yielded individualised resting-state DTBs with converged hyperparameters and stable neuronal dynamics. To construct task-state DTB models, we next assimilated empirical task-state fMRI BOLD signals from task-engaged regions to infer digital external input currents. These inferred digital input currents were then injected into the corresponding voxels of the resting-

state DTB to induce task-related neuronal activity. For regions not directly engaged by the task, model hyperparameters were fixed to the mean values estimated during the resting-state assimilation, ensuring a stable baseline while allowing task-specific dynamics to propagate through the network.

Task-relevant regions were identified based on prior literature and meta-analytic activation maps from the Neurosynth database. For the MID task, regions where received assimilated currents included the orbitofrontal cortex, dorsal anterior cingulate cortex, insula, amygdala, dorsal striatum, and nucleus accumbens, associated with reward processing. For the SST task, assimilated regions involved the orbitofrontal cortex, anterior and dorsolateral prefrontal cortex, dorsal ACC, insula, and thalamus, related to executive control. For the EFT task, assimilated regions included the amygdala, dorsal striatum, hippocampus, insula, thalamus, anterior and posterior cingulate cortex, dorsolateral prefrontal cortex, and orbitofrontal cortex, which are related to emotional recognition and processing. Primary visual were included for all three tasks, and sensorimotor cortices for first two tasks. A full list of assimilated regions is provided in Table S6.

Assessment of subject-specific information in DTB-assimilated hyperparameters.

To evaluate whether DTB-assimilated hyperparameters capture subject-specific neural dynamics independent of empirical data, we analyzed EFT task-based fMRI data from an independent subset of four male participants, matched for scanning site (Berlin) and age (one MDD and one AUD, both aged 21 years, and two HCs aged 22 years). Each participant completed four trials per emotional block (angry, neutral, and happy), yielding 12 trials per subject.

We first independently estimated hyperparameters for each trial, without using any other trial of the same participant. Each trial-specific hyperparameter was then injected into noise-initialized DTB models to generate simulated BOLD time series of matched duration. Trial-level activation maps were computed from simulated and empirical BOLD signals using identical preprocessing and statistical modeling procedures. To quantify subject specificity, we correlated activation maps from all simulated and empirical trials across subjects using Pearson correlation coefficients, yielding a 48 × 48 similarity matrix (4 subjects × 12 trials), and compared correlations between simulated and empirical trials from the same individual (self–self) with those between trials from different individuals (self–other) using a Mann–Whitney *U* test.

In addition, we tested whether simulated activation maps, generated without re-assimilation of empirical data, were sufficiently distinctive to support individual identification. For each participant, we created subject-specific empirical templates by averaging the first two trials of each emotional condition. Hyperparameters estimated from these first-half trials were used to simulate BOLD time series for the remaining trials without assimilation. We then correlated activation maps from simulated second-half trials (4 subjects × 6 trials) with all subject-specific templates, and aligned each simulated trial to the subject with the highest correlation. The same identification procedure was applied to empirical activation maps from the second-half trials as a benchmark.

Calibration experiment of DTB model scales.

The DTB framework allows flexible specification of total neuronal count across a wide range, from approximately 1 million to 86 billion neurons[14,15]. To balance biological fidelity, perturbational controllability, and computational feasibility, we performed calibration

experiments to determine optimal model resolutions for three complementary analytical purposes: high-fidelity simulation, controlled perturbational exploration, and population-level inference.

We first evaluated the influence of model scale using the DTB of an illustrative healthy control participant (HC01), by varying neuronal counts from 10 million to 5 billion and assessing simulation fidelity across resolutions. Model performance was quantified using the mean voxel-wise Pearson's correlation between simulated and empirical BOLD signals, as well as correlations between simulated and empirical whole-brain functional connectivity matrices (268-node functional parcellation[11]). These experiments established the relationship between neuronal scale and task-state simulation accuracy, providing an empirical basis for selecting biologically realistic yet computationally tractable model resolutions.

Second, to identify the optimal resolution for the population-scale simulations, we constructed reduced-scale DTBs ranging from 1 to 9 million neurons for the same participant. In these models, individual MRI data were represented at the regional rather than voxel-wise level. The number of neurons per region was assigned in proportion to the regional grey matter volume. While voxel-wise grey matter volumes could in principle be used, the total number of neurons in millions of neuron models would make it impractical to allocate neurons to every individual voxel, as some voxels have a minimal ratio of grey matter volume. By using regional-level allocation, we ensured that the neuron distribution faithfully reflected grey matter proportions while remaining computationally tractable. Between-regional connections were weighted by row-normalized regional structural connectivity derived from diffusion MRI. The other details were consistent with the methodology used in our voxel-wise simulation models. Model performance was quantified using the mean regional-level Pearson's correlation between simulated and empirical BOLD signals, as well as correlations

between simulated and empirical whole-brain functional connectivity matrices (268-node functional parcellation[11]).

Together, these calibration procedures ensured that DTB simulations maintained high fidelity while enabling scalable individualised and population-level analyses.

DTB Simulations for four illustrative participants.

Following the calibration experiments described above (see Supplementary Results), high-fidelity 1-billion-neuron DTBs were constructed for four illustrative participants: two patients (one MDD and one AUD) and two healthy controls (HC01 and HC02; demographics information in Table S7). The patients were selected as those with the highest negative NP scores within their respective diagnostic groups, indicating the highest symptom burden. Among the healthy controls, HC01 and HC02 were chosen from the subset of participants with the lowest behavioural symptom summed scores in the IMAGEN cohort. HC01 had the lowest negative NP score, whereas HC02 had the highest negative NP score, illustrating the range of NP values even among individuals with minimal behavioural symptoms. All selected participants had complete multi-modal imaging data, including T1-weighted, DTI, and fMRI scans.

In addition to billion-neuron simulations, 100-million-neuron DTBs were constructed for the same participants to enable computationally efficient perturbational analyses while retaining individualised model characteristics. This intermediate resolution was selected by calibration results showing no clear performance inflection point across neuronal scales (Fig.S4; Table S8), suggesting that reduced-scale models could preserve essential network properties while improving controllability for systematic parameter exploration.

1213 To evaluate whether 100-million-neuron DTBs retained subject-specific network phenotypes,
1214 we extracted simulated NP factor scores from each individualised model and compared them
1215 with empirical NP factors derived from all individuals in the IMAGEN and STRATIFY datasets.
1216 Similarity was quantified using mean-squared error (MSE) between each pair of simulated
1217 and empirical NP factors. For each illustrative participant, self-to-self MSE (simulated versus
1218 own empirical NP factor) was contrasted against cross-participant MSE (simulated versus all
1219 other empirical scores). Statistical significance was evaluated using permutation testing
1220 across all 1,575 simulated - empirical pairings.

1221 Systematic parameter sweeps of virtual AMPA and GABA-A perturbations.

1222 Based on 100-million-neuron DTB models of four illustrative participants, we performed
1223 systematic perturbations of AMPA- and GABA-A-mediated synaptic conductances. AMPA
1224 conductance was varied from 0.0020 to 0.0052 in steps of 0.0004 (baseline = 0.0008), and
1225 GABA-A conductance from 0.0015 to 0.0040 in steps of 0.0005 (baseline = 0.0015). The
1226 baseline levels of AMPA and GABA-A in the model were determined across two key
1227 dimensions: the mean firing rate of the neuronal population (<10 Hz), consistent with
1228 physiological observations[22], and the synchrony measure, which represents the degree of
1229 oscillatory behaviour. These metrics identified the optimal parameter window where the
1230 model maintains robust, reasonable neuronal firing while avoiding excessive
1231 synchronization[23], thereby ensuring a stable and biologically plausible baseline for the
1232 model.

1233 Crucially, the upper limit for the parameter of AMPA conductance was constrained by the
1234 stability of the simulated BOLD signal: beyond a specific threshold, excessive neuronal firing
1235 led to a cessation of dynamic fluctuations in the BOLD signal. We considered such states as

1236 "over-firing" regimes, where the hemodynamic response reaches a non-biological plateau
1237 and no longer reflect meaningful neural dynamics[24]. During GABA-A manipulations, AMPA
1238 conductance was held fixed at the value that had produced the maximal NP factor
1239 enhancement during prior AMPA sweeps, allowing controlled assessment of excitatory-
1240 inhibitory balance, and preventing excessive inhibition that could silence neuronal activity
1241 and abolish intrinsic BOLD fluctuations[25].

1242 These synaptic conductance parameters were implemented as global control variables,
1243 applied uniformly across the whole-brain model rather than specific regions. Importantly,
1244 they were directly applied to DTB models that had already been calibrated to individual
1245 empirical fMRI data; therefore, no additional data assimilation or parameter refitting was
1246 required during virtual perturbations. The NP network responses were obtained by forward
1247 simulation under systematically varied synaptic levels. This design allows direct assessment
1248 of perturbational sensitivity around individualised baseline operating points. To ensure
1249 reproducibility, each perturbation was repeated five times, and the mean NP factor scores
1250 across repetitions were used for subsequent analyses. Finally, while multiple combinations
1251 of synaptic parameters can give rise to similar network outputs, our goal was not to recover
1252 exact physiological parameter values, but to use biologically interpretable synaptic
1253 conductance parameters as control variables to probe regime-dependent network responses.

1254 Evaluation of cross-task generalisability of virtual perturbations.

1255 To test whether the effects of AMPA- and GABA-A–mediated perturbations on the NP factor
1256 generalise across cognitive contexts, we conducted the same virtual perturbation to the DTB
1257 models that simulate EFT task activity of the same four illustrative participants. Specifically,
1258 we first calibrated DTB models at a 100-million-neuron scale to simulate BOLD signals of

EFT task-based fMRI data, and then applied the optimal perturbational parameters identified in the parameter sweeps to these DTB models. For each participant, NP-related functional connectivity was extracted from baseline and perturbed simulated EFT BOLD signals using the same computational pipeline as the original MID and SST tasks, enabling direct comparison of perturbation effects across cognitive tasks.

Population-scale virtual perturbations.

To assess whether virtual E/I modulations in DTB models generalise across the cohort, we first constructed DTB models for 290 participants, including 72 patients with MDD (age = 22.33 ± 2.20, female/male = 48/24), 59 with AUD (age = 22.41 ± 2.02, female/male = 37/22), and 69 healthy controls (age = 21.45 ± 1.38, female/male = 38/31) from STRATIFY dataset, as well as 90 subclinical participants with high behavioural symptom scores (sum of six externalising and internalising symptoms ≥ 20) from IMAGEN (age = 18.42 ± 0.65, female/male = 74/16). We then applied the optimal perturbational parameters identified in the parameter sweeps to these DTB models.

For each participant, we calculated simulated NP factor scores derived from baseline and perturbed BOLD signals. Group differences in empirical, simulated, and perturbed NP factors were assessed using ANOVA analysis and post-doc tests, controlling for sex, site, and head motion (mean framewise displacement). The paired-sample *t* tests were used to compare simulated and empirical, and simulated and perturbed NP factors, among subgroups. Multiple comparisons were corrected using the Holm-Bonferroni method.

Prediction of task performance from DTB simulations.

To assess the behavioural validity of digital twin brain simulations, we evaluated whether simulated task-state functional connectivity metrics, derived from population-scale simulations, significantly predicted individual differences in task performance during the MID task.

Analyses were conducted in the 287 participants with 3-million-neuron DTB simulations and MID task behavioural data. Individualised task-state functional connectomes were extracted separately for the anticipation-hit and feedback-hit stages of the MID task. Behavioural performance was quantified as mean response time (RT) under three incentive conditions: big-win, small-win, and no-win. Prediction of behavioural performance was performed using a connectome-based predictive modeling framework with 10-fold cross-validation. For each fold, linear models were trained on task-state FC matrices from nine folds to predict RT measures and evaluated on the held-out fold. Prediction accuracy was quantified as the Spearman's correlation between predicted and observed RT values. The entire cross-validation procedure was repeated 100 times to ensure stability, and prediction accuracies were averaged across repetitions. Separate models were constructed for each combination of task stage (anticipation-hit, feedback-hit) and behavioural measure (big-win, small-win, no-win), resulting in six independent prediction models.

We also examined the prediction performances of empirical task-state FC for establishing a benchmark. Prediction accuracies obtained using empirical FC and DTB-simulated FC were compared using paired-sample *t* tests across the six model combinations.

Stratification of NP network responses to virtual perturbations.

In the population-level analysis, we quantified individual responses to virtual neurotransmitter perturbations in the DTB models. For each participant, we computed the

1303 change in simulated NP factor following AMPA and GABA-A modulation relative to baseline.
1304 Participants were classified as "increasers" (ΔNP>0 following both modulations) or
1305 "decreasers" (ΔNP ≤ 0 for any given perturbation). We compared the distribution of response
1306 types across diagnostic groups (patients, high-symptom, and healthy controls), sex, and site
1307 using Chi-square tests.

1308 We also assessed the relationship between baseline simulated NP factor and the AMPA- or
1309 GABA-A–induced changes using Pearson's correlation. Statistical significance was assessed
1310 using permutation testing (10,000 permutations), in which ΔNP values were randomly
1311 reassigned across individuals to generate a null distribution. For each permutation, the
1312 correlation between simulated baseline NP and permuted ΔNP was recomputed. Two-sided
1313 permutation *P* values were calculated as the proportion of null correlations whose absolute
1314 value exceeded the observed correlation, directly testing whether modulation-induced
1315 changes reflect systematic baseline dependence rather than symmetric fluctuations or
1316 regression-to-the-mean effects.

1317 Finally, we compared simulated baseline NP factors and behavioural symptom measures
1318 using Mann–Whitney *U* tests, given unequal subgroup sizes and non-normal distributions.
1319 All analyses controlled for sex, scanning site, and mean framewise displacement.
1320 Behavioural measures were derived from the 69 entry items of the DAWBA and the SDQ.

1321 Stratification of individual pharmacological responses.

1322 To validate DTB-predicted effects of virtual AMPA and GABA-A modulations on the NP
1323 factor, we analyzed an independent pharmacological fMRI dataset in which healthy
1324 participants completed MID task-fMRI under placebo, ketamine, and midazolam in a

randomised cross-over design. For each participant and drug condition, we computed NP-related MID FC strengths (six edges) and summed them as a "summed MID FC" metric.

We first tested group-level drug effects by comparing the summed MID FC between placebo and drug conditions using paired-sample *t* tests. To further capture individual variability, we computed the drug-induced change in summed MID FC relative to placebo for each participant and applied k-means clustering (500 iterations) to identify two response subgroups. Within each subgroup, paired-sample *t* tests, controlling for head motion, were used to compare placebo and drug conditions. Wilcoxon signed-rank tests were applied when normality or variance assumptions were violated.

We also compared summed MID FC under placebo between response subgroups using a independent-samples *t* test. Finally, the relationship between baseline summed MID FC in placebo and drug-induced FC change across all participants was assessed using Pearson's correlations. Statistical significance was assessed using permutation testing (10,000 permutations), in which Δsummed MID FC values were randomly reassigned across individuals to generate a null distribution. For each permutation, the correlation between baseline summed MID FC and permuted Δsummed MID FC was recomputed. Two-sided permutation *P* values were calculated as the proportion of null correlations whose absolute value exceeded the observed correlation, directly testing whether drug-induced changes reflect baseline dependence rather than regression-to-the-mean effects.

Prediction of pharmacological responses from DTB-derived virtual perturbations.

To test whether DTBs capture systematic relationships between baseline network configuration and drug-induced NP-related connectivity changes, we first constructed a linear regression model linking simulated baseline summed MID FC and FC changes

following virtual AMPA and GABA-A modulations. We then applied this regression model to predict empirical FC changes under ketamine or midazolam from each participant's baseline FC under placebo. Here, we chose a linear mapping to minimize overfitting and maximize interpretability given the modest sample size. Prediction accuracy was quantified by Pearson's correlation between predicted and observed FC changes, with significance assessed via permutation tests (1,000 permutations of subject labels).

To evaluate whether DTB-based predictions can distinguish individuals with opposite drug responses, participants were classified as "increasers" or "decreasers" based on the their observed drug-induced change in summed MID FC. The predicted probabilities from the DTB-derived regression model were then used to calculate the area under the receiver operating characteristic curve (AUC) as a measure of how well the model correctly identifies individuals in each response category.

Prediction of longitudinal symptom changes from DTB-derived virtual perturbations.

To test whether DTB-derived virtual E/I modulations predict longitudinal symptom trajectories, we first estimated an empirical linear mapping between NP factors and behavioural symptoms using baseline data. Specifically, we fitted a linear regression linking NP factors at age 19 to summed internalising symptom scores at age 19 across four domains (eating disorder, depression, generalized anxiety, and specific phobia). Externalising symptoms (ADHD and conduct disorder) were excluded due to their low prevalence at follow-up (age 23).

Baseline NP factors and NP factors under virtual AMPA and GABA-A modulation were simulated for the full cohort using 3-million-neuron DTBs. From this cohort, we selected individuals with imaging data at age 19 and symptom assessments at both ages 19 and 23.

The empirical NP–symptom mapping derived at baseline was then applied to the DTB-simulated NP factors to generate predicted symptom scores at baseline and after virtual modulation.

For each individual, we defined a DTB-derived "behavioural restoration" index as the difference between predicted baseline and post-modulation symptom scores. Longitudinal symptom change over four years (age 23 minus age 19) was subsequently modeled using multiple linear regression, with baseline symptom scores and the DTB-derived "behavioural restoration" index included as predictors. Prediction accuracy was quantified by Pearson's correlation between predicted and observed behavioural symptom changes. Model fit was quantified using the coefficient of determination (r^2). The incremental variance explained by the DTB-derived index (Δr^2) was assessed by comparing full and reduced models using F tests. Statistical significance of DTB-related effects was further evaluated using 5,000 permutation tests for both r^2 and the regression coefficient associated with the DTB index.

Methods References

1. Mascarell Maričić, L. *et al*. The IMAGEN study: a decade of imaging genetics in adolescents. *Mol. Psychiatry* 25, 2648–2671 (2020).

2. Schumann, G. *et al*. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol. Psychiatry* 15, 1128–1139 (2010).

3. Xie, C. *et al*. A shared neural basis underlying psychiatric comorbidity. *Nat. Med*. 29, 1232–1242 (2023).

4. Allen, J. P., Litten, R. Z., Fertig, J. B. & Babor, T. A review of research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcohol. Clin. Exp. Res*. 21, 613–619 (1997).

5. Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med*. 16, 606–613 (2001).

6. Sheehan, D. V. *et al*. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 59 Suppl 20, 22-33;quiz 34-57 (1998).

7. Forsyth, A. *et al*. Comparison of local spectral modulation, and temporal correlation, of simultaneously recorded EEG/fMRI signals during ketamine and midazolam sedation. *Psychopharmacology (Berl.)* 235, 3479–3493 (2018).

8. Goodman, R., Ford, T., Richards, H., Gatward, R. & Meltzer, H. The Development and Well-Being Assessment: Description and initial validation of an integrated assessement of child and adolescent psychopathology. *J. Child Psychol. Psychiatry* https://doi.org/10.1017/S0021963099005909 (2000) doi:10.1017/S0021963099005909.

9. Goodman, R. Psychometric properties of the strengths and difficulties questionnaire. *J. Am. Acad. Child Adolesc. Psychiatry* 40, 1337–45 (2001).

10. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* 16, 111–116 (2019).

11. Shen, X., Tokoglu, F., Papademetris, X. & Constable, R. T. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage* 82, 403–415 (2013).

12. Shen, X. *et al.* Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat. Protoc.* 12, 506–518 (2017).

13. Hansen, J. Y. *et al.* Mapping neurotransmitter systems to the structural and functional organization of the human neocortex. *Nat. Neurosci.* 25, 1569–1581 (2022).

14. Lu, W. *et al.* Imitating and exploring human brain's resting and task-performing states via resembling brain computing: scaling and architecture. *Natl. Sci. Rev.* nwae080 (2024) doi:10.1093/nsr/nwae080.

15. Lu, W. *et al.* Simulation and assimilation of the digital human brain. *Nat. Comput. Sci.* 4, 890–898 (2024).

16. Brunel, N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* 8, 183–208 (2000).

17. Binzegger, T., Douglas, R. J. & Martin, K. A. C. A Quantitative Map of the Circuit of Cat Primary Visual Cortex. *J. Neurosci.* 24, 8441–8453 (2004).

1426 18. Deco, G., Jirsa, V. K. & McIntosh, A. R. Emerging concepts for the dynamical
1427 organization of resting-state activity in the brain. *Nat. Rev. Neurosci.* 12, 43–56 (2011).

1428 19. Isaacson, J. S. & Scanziani, M. How Inhibition Shapes Cortical Activity. *Neuron* 72,
1429 231–243 (2011).

1430 20. Abbott, L. F. Lapicque's introduction of the integrate-and-fire model neuron (1907).
1431 *Brain Res. Bull.* 50, 303–304 (1999).

1432 21. Friston, K. J., Mechelli, A., Turner, R. & Price, C. J. Nonlinear responses in fMRI: the
1433 Balloon model, Volterra kernels, and other hemodynamics. *NeuroImage* 12, 466–477 (2000).

1434 22. Yu, Y. *et al.* A 3D atlas of functional human brain energetic connectome based on
1435 neuropil distribution. *Cereb. Cortex* 33, 3996–4012 (2023).

1436 23. Baroni, F. & Fulcher, B. Synchrony, oscillations, and phase relationships in collective
1437 neuronal activity: A highly comparative overview of methods. *PLOS Comput. Biol.* 21,
1438 e1013597 (2025).

1439 24. Deco, G. *et al.* Resting-state functional connectivity emerges from structurally and
1440 dynamically shaped slow linear fluctuations. *J. Neurosci. Off. J. Soc. Neurosci.* 33, 11239–
1441 11252 (2013).

1442 25. Li, J., Kells, P. A., Osgood, A. C., Gautam, S. H. & Shew, W. L. Collapse of complexity
1443 of brain and body activity due to excessive inhibition and MeCP2 disruption. *Proc. Natl.*
1444 *Acad. Sci.* 118, e2106378118 (2021).

1445

Acknowledgments

This work received support from the following sources: National Natural Science Foundation of China (32400935 to YX, W2541022 to GS, 82150710554 to GS), National Science and Technology Major Project (2025ZD0215100 to GS), 111 Project (B18015 to JF), Shanghai Municipal Science and Technology Major Project (2018SHZDZX01 to JF), the European Union-funded FP6 Integrated Project IMAGEN (Reinforcement-related behaviour in normal brain function and psychopathology) (LSHM-CT- 2007-037286 to GS), the Horizon 2020 funded ERC Advanced Grant 'STRATIFY' (Brain network based stratification of reinforcement-related disorders) (695313 to GS), Horizon Europe 'environMENTAL', grant no: 101057429 to GS, UK Research and Innovation (UKRI) Horizon Europe funding guarantee (10041392 and 10038599) to SD, Human Brain Project (HBP SGA 2, 785907, and HBP SGA 3, 945539) to GS. The German Center for Mental Health (DZPG), the Bundesministerium für Bildung und Forschung (BMBF grants 01GS08152; 01EV0711; Forschungsnetz AERIAL 01EE1406A, 01EE1406B; Forschungsnetz IMAC-Mind 01GL1745B), the Deutsche Forschungsgemeinschaft (DFG project numbers 458317126 [COPE] to GS, 186318919 [FOR 1617], 178833530 [SFB 940], 386691645 [NE 1383/14-1], 402170461 [TRR 265], 454245598 [IRTG 2773]), the Medical Research Foundation and Medical Research Council (grants MR/R00465X/1 and MR/S020306/1) to SD, the National Institutes of Health (NIH) funded ENIGMA-grants 5U54EB020403-05, 1R56AG058854-01 and U54 EB020403 as well as NIH R01DA049238, the National Institutes of Health, Science Foundation Ireland (16/ERCD/3797) to ALWB and RW.

Consortia

DTB Consortium

Yubin Bao, Boyu Chen, Siming Chen, Zhongyu Chen, Fei Dai, Weiyang Ding, Xin Du, Jianfeng Feng, Yubo Hou, Mingda Ji, Peng Ji, Chong Li, Chunhe Li, Xiaoyi Li, Yuhao Liu, Wenlian Lu, Zhihui Lv, Hengyuan Ma, Yang Qi, Edmund Rolls, He Wang, Huarui Wang, Jiexiang Wang, Shouyan Wang, Ziyi Wang, Yunman Xia, Shitong Xiang, Chao Xie, Xiangyang Xue, Leijun Ye, Longbin Zeng, Tianping Zeng, Chenfei Zhang, Jie Zhang, Nan Zhang, Wenyong Zhang, Yicong Zhao & Qibao Zheng

IMAGEN Consortium

Tobias Banaschewski, Gareth J. Barker, Arun L.W. Bokde, Rüdiger Brühl, Sylvane Desrivières, Herta Flor, Penny Gowland, Antoine Grigis, Andreas Heinz, Herve Lemaitre, Jean-Luc Martinot, Marie-Laure Paillère Martinot, Eric Artiges, Frauke Nees, Dimitri Papadopoulos Orfanos, Tomáš Paus, Luise Poustka, Michael N. Smolka, Sarah Hohmann, Nilakshi Vaidya, Henrik Walter, Robert Whelan, Paul Wirsching & Gunter Schumann

STRATIFY Consortium

Nilakshi Vaidya, Zuo Zhang, Lauren Robinson, Jeanne Winterer, Yuning Zhang, Gareth J Barker, Arun L.W. Bokde, Rüdiger Brühl, Hedi Kebir, Hervé Lemaître, Frauke Nees, Dimitri Papadopoulos Orfanos, Ulrike Schmidt, Julia Sinclair, Robert Whelan, Henrik Walter, Sylvane Desrivières & Gunter Schumann

EnvironMENTAL Consortium

Tobias Banaschewski, Sylvane Desrivières, Andreas Heinz, Frauke Nees, Nilakshi Vaidya, Henrik Walter, Vince D. Calhoun & Gunter Schumann

<u>Author contributions</u>

Conceptualization: G.S., J.F., V.J., W.L., Y.X., S.P.. Methodology: W.L., Y.X., S.P., C.X., S.X., S.Pet., Y.L., J.W., L.Z., H.W., Q.Z., and DTB Consortium. Investigation: Y.X., S.P., S.X., Z.L.. Data acquisition: J.D., J.H., S.M., A.F., T.B., G.J.B., A.L.W.B., R.B., S.D., H.F., P.G., A.G., A.H., H.L., J.-L.M., M.-L.P.M., E.A., F.N., D.P.O., T.P., L.P., M.N.S., S.H., N.V., H.W., R.W., P.W., Z.Z., L.R., J.W., Y.Zh., H.K., U.S., J.S., G.S.; Data were acquired and provided by the IMAGEN and STRATIFY consortiums. Supervision: G.S., J.F., W.L., V.J., S.Pet., C.Z., A.M., J.D., T.J., T.L.. Writing—original draft: Y.X., S.P., G.S. Writing—review & editing: Y.X., G.S., S.P., J.D., S.Pet., L.Q., S.D., Z.Z., J.F., W.L..

Data availability

The data analysed in this study were obtained from the IMAGEN consortium and associated cohorts. Due to ethical and data protection restrictions, the raw data are not publicly available. Researchers may request access to the data through the official data access procedures of the respective consortia. Processed data supporting the findings of this study are available from the corresponding author upon reasonable request and subject to consortium approval.

Code availability

The code used to perform the analyses in this study is publicly available at GitHub (https://github.com/Shmily94/Digital_twin_brain-psychiatry-.git). Custom codes developed in PyTorch for personal computing environments and in C++ for high-performance computing (HPC) systems of the DTB platform is available via the DTB Consortium GitHub repository (https://github.com/DTB-consortium/Digital_twin_brain-open) and Zenodo (https://doi.org/10.5281/zenodo.13995756). Please note that parts of the HPC

implementation depend on specific hardware and software configurations. Further technical information can be requested from the DTB Consortium via dtb.fudan@gmail.com.

Supplementary Materials

Digital Twin Brain simulation and and manipulation of a functional brain network underlying mental illness.

This file includes: Supplementary Results and Figures. S1 to S15

Supplementary Results

The Supplementary Results are organized to systematically demonstrate the DTB framework's performance, fidelity, and applicability. We first presented calibration experiments to select appropriate neuronal model scales for multiple analytical aims, followed by the evaluation of individualised model fidelity. We then illustrate controlled E/I perturbational analyses and assess the generalisation of bidirectional effects across tasks. We then demonstrate the behavioural relevance of DTB simulations and examine the robustness of NP network responses against potential confounds and network definition. Finally, we show that DTBs preserve subject-specific signatures, even without state-specific re-assimilation. This structure clarifies how DTB simulations support both mechanistic insights, behavioural predictions, and population-level inference.

Calibration experiments for DTB model scale selection.

To balance computational efficiency with simulation fidelity, we conducted a series of calibration experiments to determine appropriate neuronal scales for high-fidelity voxel-wise simulations and population-scale regional-level analysis.

We first evaluated model performance across neuronal scales ranging from 10 million to 5 billion neurons using an individualised DTB constructed for an illustrative healthy control participant (HC01, Main Text). Model fidelity was quantified as the mean voxel-wise Pearson correlation between simulated and empirical BOLD signals as well as functional connectivity matrices. Performance increased with neuronal scale, exceeding 0.40 at the 1-billion-neuron resolution for both tasks (Stop Signal Task (SST): r = 0.45, Monetary Incentive Delay (MID): r = 0.42; Fig. S4; Table S8), with minimal improvement at higher resolutions. Moreover, global functional connectivity matrices generated by the 1-billion-neuron model exhibited high correspondence with empirical data, with correlations ranging from 0.49 to 0.61 across four task conditions (Fig. S4; Table S8). These results indicate performance saturation near the billion-neuron scale, supporting its selection for demonstrating biological fidelity and simulation feasibility.

Because billion-neuron simulations are computationally prohibitive for large-cohort analyses, we next evaluated reduced-resolution DTBs in which individualised MRI data were represented at the regional level. Using the same participant, we conducted a second calibration experiment in which the neuronal scale was varied from 1 to 9 million neurons. Regional mean correlations saturated at approximately 3 million neurons, exceeding 0.6, with negligible gains at higher resolutions (Fig. S4; Table S14). Accordingly, a 3-million-neuron scale was adopted for population-scale simulations.

Together, these calibration experiments demonstrate scale-dependent performance saturation in DTB simulations, enabling principled selection of neuronal resolution across individualised high-fidelity and population-scale modeling regimes.

Preservation of subject-specific NP factors in 100-million-neuron DTB models.

Although 1-billion-neuron DTBs provide better simulation fidelity, the calibration curve did not show a clear inflection point, and exploring the perturbational parameter space at this scale would be computationally prohibitive. Therefore, we implemented 100-million-neuron DTBs for the same four participants to enable controlled perturbational exploration while preserving individualised network structure.

Despite the reduced resolution, 100-million-neuron DTBs preserved high accuracy within assimilated regions, with mean voxel-wise correlations of 0.91 for the SST and 0.90 for the MID task across participants (Table S10). As expected, whole-brain mean correlations were lower (0.30 for SST and 0.27 for MID; Table S10), reflecting the reduced simulation performance outside task-constrained regions.

To confirm that 100-million-neuron models preserved individualised network phenotypes, we compared simulated NP factors derived from 100-million-neuron DTBs with empirical NP factors from all individuals in the IMAGEN and STRATIFY cohorts. For each illustrative participant, self-to-self mean-squared error (MSE) between simulated and empirical NP factors was significantly lower than cross-participant comparisons (MDD, $P_{perm} = 0.02$; AUD, $P_{perm} = 0.001$; HC01, $P_{perm} < 0.001$; HC02, $P_{perm} < 0.001$, permutation tests across 1,575 simulated–empirical pairings; Fig. S7), demonstrating retention of subject-specific NP structure at this intermediate resolution.

Task-generalisation of bidirectional E/I modulation effects.

To test whether bidirectional NP connectivity changes following excitation–inhibition (E/I) modulation generalise across cognitive contexts, we repeated AMPA- and GABA-A–based perturbations in the Emotional Face Task (EFT), using 100-million-neuron DTBs calibrated for the same four individuals. NP-related functional connectivity was computed from simulated BOLD signals at baseline and after each virtual perturbation.

Consistent with the results observed in the MID and SST contexts, virtual E/I modulation produced heterogeneous and bidirectional NP responses across individuals. Both patient models (one MDD and one AUD) showed increases in NP-related connectivity following AMPA and GABA-A perturbations. However, one healthy control (HC01) showed a similar decrease in NP-related connectivity, and the other one (HC02) exhibited minimal change following AMPA modulation (Table S13; Fig. S8). These findings indicate that the heterogeneous, bidirectional modulation effects captured by the DTBs is not task-specific, but generalises across distinct cognitive contexts.

We also observed variability in the responses of the two healthy controls to GABA-A modulation, likely reflecting task-state dependent shifts in the network operating points, such that the optimal perturbation parameters for SST and MID tasks may not directly translate to the EFT.

Prediction of task performances from DTB simulations.

To provide behavioural validation beyond the NP connectivity network, we tested whether simulated functional connectivity derived from DTB simulations significantly predicted task performance, with prediction accuracy exceeding that based on empirical functional connectivity.

Using empirical MID task-state FC, both anticipation-hit and feedback-hit global connectivity significantly predicted response times under big-win, small-win, and no-win conditions (n = 287, all $P_{corrected} < 0.0001$, Table S16; Fig. S10), confirming that task-state FC captures behaviourally relevant variance. When the same prediction framework was applied to DTB-simulated functional connectivity, simulated FC also significantly predicted response times across all incentive conditions (all $P_{corrected} < 0.0001$, Table S16; Fig. S10). Importantly, prediction accuracy based on simulated FC consistently exceeded that obtained using empirical FC. A direct comparison across the six prediction models revealed significantly higher prediction performance for simulated FC than empirical FC

(paired-sample t test: t = -4.47, P = 0.0066, Fig. S10), with a maximum increase in explained variance of up to 9% (Δr^2).

These results indicate that DTB simulations preserve behaviourally relevant neural information and, in some cases, enhance predictive signal relative to empirical functional connectivity alone, consistent with the view that DTBs distill task-relevant network dynamics while reducing measurement noise.

Assessment of potential confounds in NP network responses.

To exclude potential confounding effects on individual NP responses toward virtual modulations, we examined whether responses types differed by sex or site. Results suggested that response distributions did not differ across sites (χ^2 = 11.02, P = 0.20). Although a nominal sex effect was observed (χ^2 = 7.66, P = 0.0056), the association between response type and sex was no longer significant after adjusting for diagnostic group (t = –1.94, P = 0.054), indicating that the apparent sex difference was driven by the higher proportion of females in the patient group.

Robustness of bidirectional, baseline-dependent responses in MID-specific NP connectivity.

To test whether the bidirectional response pattern generalises beyond the full NP definition, we re-analysed virtual perturbations in the population-scale simulations and restricted the network phenotype to MID-specific NP connectivity. This reproduced a similar bidirectional, baseline-dependent response pattern. We also divided participants into subgroups: increased FC after both AMPA and GABA-A modulation (n = 257), increased after AMPA but decreased after GABA-A (n = 12), decreased after AMPA but increased after GABA-A (n = 17), and decreased after both (n = 4). Baseline simulated summed MID FC score was inversely correlated with AMPA- (Pearson's r = -0.74, P < 0.0001) and GABA-A- (Pearson's r = -0.24, P < 0.0001) induced changes in NP connectivity (Fig. S13), and was significantly lower in the increased group compared the decreased group (Fig. S13).

These findings demonstrate that the bidirectional, baseline-dependent E/I response pattern is robust to both task restriction and network definition, confirming the core dynamic property of the NP circuit.

DTBs preserve subject-specific signatures without re-assimilation.

Although DTBs are configured to assimilate state-specific functional brain activity, we tested whether individualised DTB models retain subject-specific information in the absence of re-assimilation. To minimize potential confounds, we selected four participants matched for sex, scanning site, and age, and constructed 1-billion-neuron DTBs for them to simulate BOLD signals during the EFT. Trial-specific hyperparameters estimated from each EFT trial were used, without any re-optimization, to simulate BOLD signals of all remaining trials across the four participants. Trial-level activation maps were computed from both simulated and empirical time series using identical procedures.

Correlations between simulated and empirical trials from the same individual (self–self) were significantly higher than those between trials from different individuals (self-other) (Mann-Whitney U test, $P < 0.0001$, Fig. S15). Moreover, when hyperparameters derived from the first half of the EFT were used to simulate BOLD signals in the second half, simulated trials achieved 100% fingerprinting accuracy, compared with 95.8% for empirical second-half trials (Fig. S15). These analyses indicate that DTB models, even without re-assimilation, preserve subject-specific information and reliably identify individual BOLD dynamics. More details were available in Methods.

All detailed statistical information is available in Table S21.

Figures

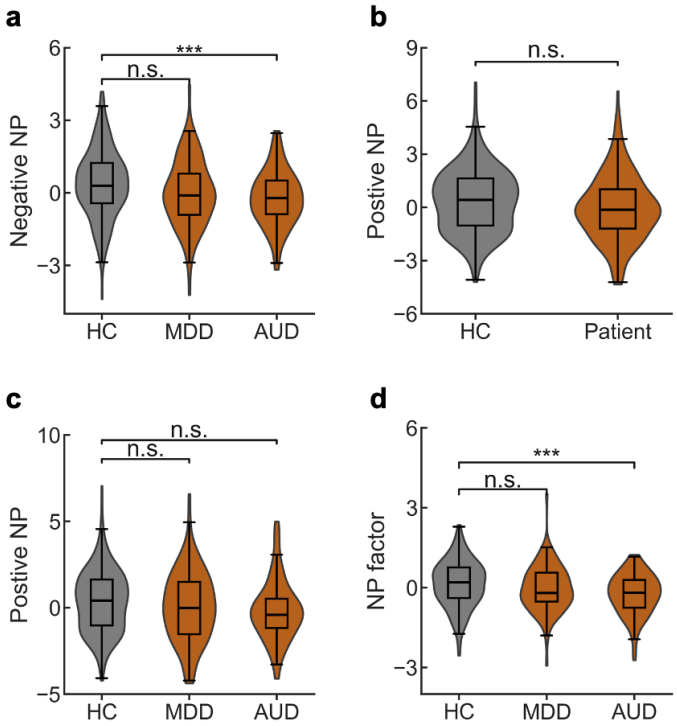

Fig. S1 Comparison of NP profiles between diagnostic groups and healthy controls. a, Negative NP scores (residuals corrected for sex, site, and head motion) in patients with MDD (n = 104, t = 2.34, $P_{corrected}$ = 0.059) and AUD (n = 98, t = 3.82, $P_{corrected}$ = 0.0005) compared with healthy controls (n = 225) from the STRATIFY cohort. **b**, Positive NP scores (summed strength of the positive FC profile) in HCs (n = 225) and patients (n = 209), regressing out sex, site, and head motion. Group difference was assessed using an independent-samples t test with Bonferroni correction (t = 2.29, $P_{corrected}$ = 0.068). **c**, Positive NP scores in each clinical group compared with the HC group. No significant differences were observed for MDD (t = 1.32, $P_{corrected}$ = 0.56) and AUD (t = 2.41, $P_{corrected}$ = 0.05). **d**, Summed 12-edge-NP factor scores were lower in AUD (t = 3.88, $P_{corrected}$ = 0.0004), but not in the MDD (t = 1.62, $P_{corrected}$ = 0.32). All violin plots show kernel density estimates with embedded boxplots indicating the median and interquartile range (IQR); whiskers extend to 1.5× IQR. n.s., not significant; *, $P_{corrected}$ <0.05; ***, $P_{corrected}$ <0.001. P values were corrected for multiple comparisons using a Bonferroni procedure with three tests.

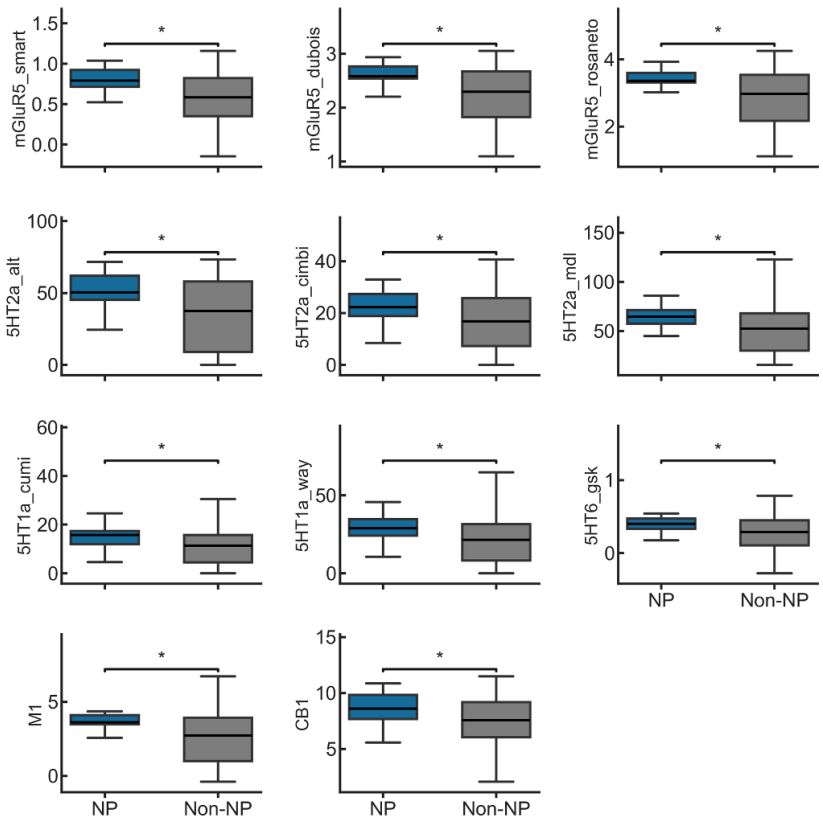

Fig. S2 Enrichment of PET-derived receptor maps in NP-related regions. Distributions of regional values from eleven PET-derived neurotransmitter receptor maps are shown for NP-related (blue) and non-NP (grey) cortical regions. Maps include glutamatergic receptors (mGluR5: SMART, Dubois, and Rosaneto), serotonergic receptors (5-HT_{2A}: Alt, Cimbi, and MDL; 5-HT_{1A}: Cumi and Way; 5-HT_6: GSK), cholinergic receptors (M1), and cannabinoid receptors (CB1). Multiple maps for the same receptor reflect independent PET datasets or radioligands rather than distinct receptor subtypes. Statistical significance was assessed using permutation testing (10,000 permutations), in which regional labels (NP-related vs. NP-unrelated) were randomly reassigned while preserving the original group sizes. Two-sided permutation P values were corrected for multiple comparisons across 39 receptor maps using false discovery rate (FDR $q < 0.05$) correction. The boxplots indicate the mean and interquartile range (IQR); whiskers extend to 1.5× IQR. *, $P_{corrected}<0.05$.

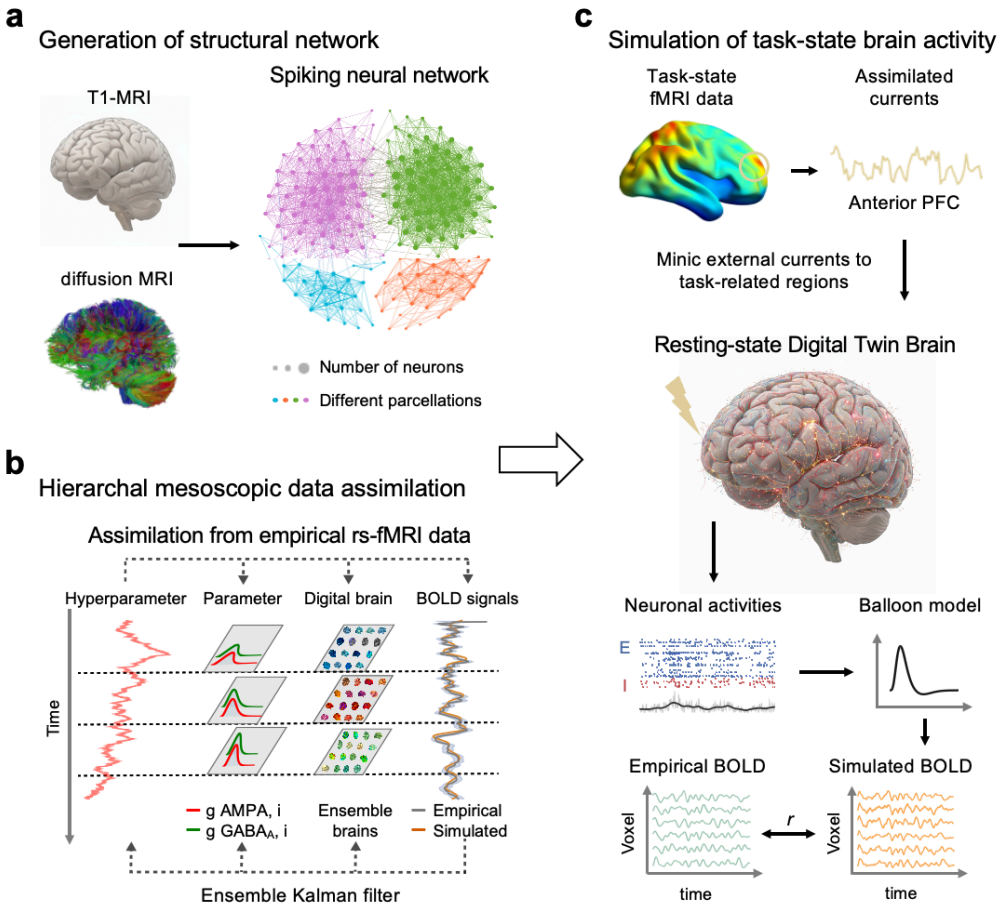

Fig. S3 Construction of individualised task-state DTB models. **a**, Multimodal MRI data were integrated to construct individualised spiking neural network. Diffusion tensor imaging (DTI) was used to estimate voxel-/regional- wise white-matter connectivity, defining inter-unit connection probabilities, while T1-weighted images were used to assign voxel-/regional- specific neuronal population sizes according to grey matter volume. Neuronal and synaptic structural networks were constructed at the voxel or region-wise unit, after specifying the total neuron number and the average in-degree. Network initialization procedures were applied to prevent artificial synchronization of neuronal spiking. Neuronal activity was simulated using the leaky integrate-and-fire model (LIF)[23], incorporating AMPA (excitatory-leaning) and GABA-A (inhibitory-leaning) synaptic conductances. **b**, A hierarchical mesoscopic data assimilation framework was employed to fit simulated BOLD signals to empirical fMRI data. Model hyperparameters were estimated by iteratively alternating between forward simulation and state filtering using a diffusion ensemble Kalman filter (EnKF). The EnKF used an ensemble of 30 parallel model realizations to estimate the variance of hidden states. For each voxel, AMPA and GABA-A synaptic conductances were sampled from gamma distributions, which were constrained by voxel-specific hyperparameters. As assimilation progressed

over time, simulated BOLD signals converged toward empirical observations. Three representative time points are illustrated, showing the sampled parameters and ensemble models. After assimilations for all voxels, a resting-state DTB was obtained. **c**, Task-state DTBs were generated by assimilating task-state BOLD signals from task-engaged regions to estimate digital external input currents. These inferred inputs were then injected into the corresponding voxels of the resting-state DTB to induce task-state dynamics. For regions not directly engaged by the task, hyperparameters were fixed at their mean values of the hyperparameters estimated in the resting state. Voxel-averaged neuronal activity (mean firing rates, E means excitatory, and I means inhibitory) generated by the task-state DTB were then transformed into simulated task-state BOLD signals using the Balloon–Windkessel hemodynamic model. Model performance was quantified by Pearson's correlation (*r*) between simulated and empirical task-state BOLD signals. More details about the model construction can be found in Methods.

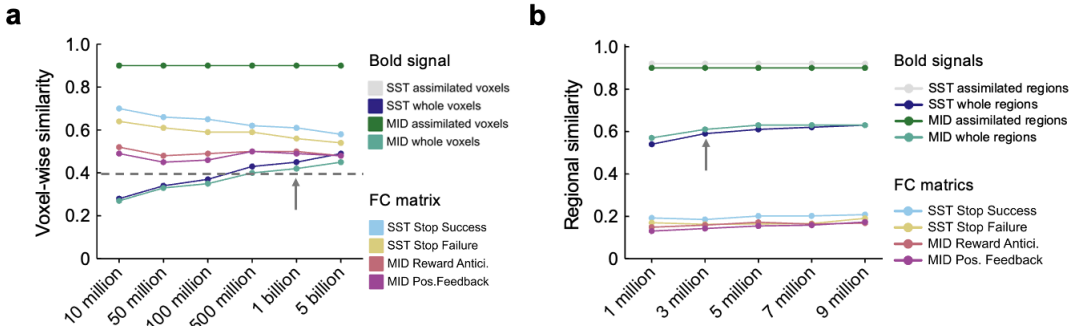

Fig. S4 Calibration of DTB neuronal scale for individualised and population-scale simulations. a, Individualised DTB models for a illustrative healthy control (HC01) were constructed across neuronal scales (10 million–5 billion) and evaluated by voxel-wise Pearson's correlations between simulated and empirical BOLD signals and task-specific functional connectivity matrics during the MID and SST tasks (lines indicate mean Pearson's *r* across scales). Simulation performance plateaued at the 1-billion-neuron scale (mean correlations across the whole voxels: *r* of MID= 0.42; *r* of SST = 0.45), with no further improvement at higher resolutions. **b,** For population-scale simulations, regional DTB models spanning 1–9 million neurons showed saturation of regional BOLD correlations at approximately 3 million neurons (mean correlations across the whole voxels of both tasks > 0.6). These results motivated the use of 1-billion-neuron DTBs to demonstrate individualised simulation fidelity and 3-million-neuron DTBs for population-level analyses.

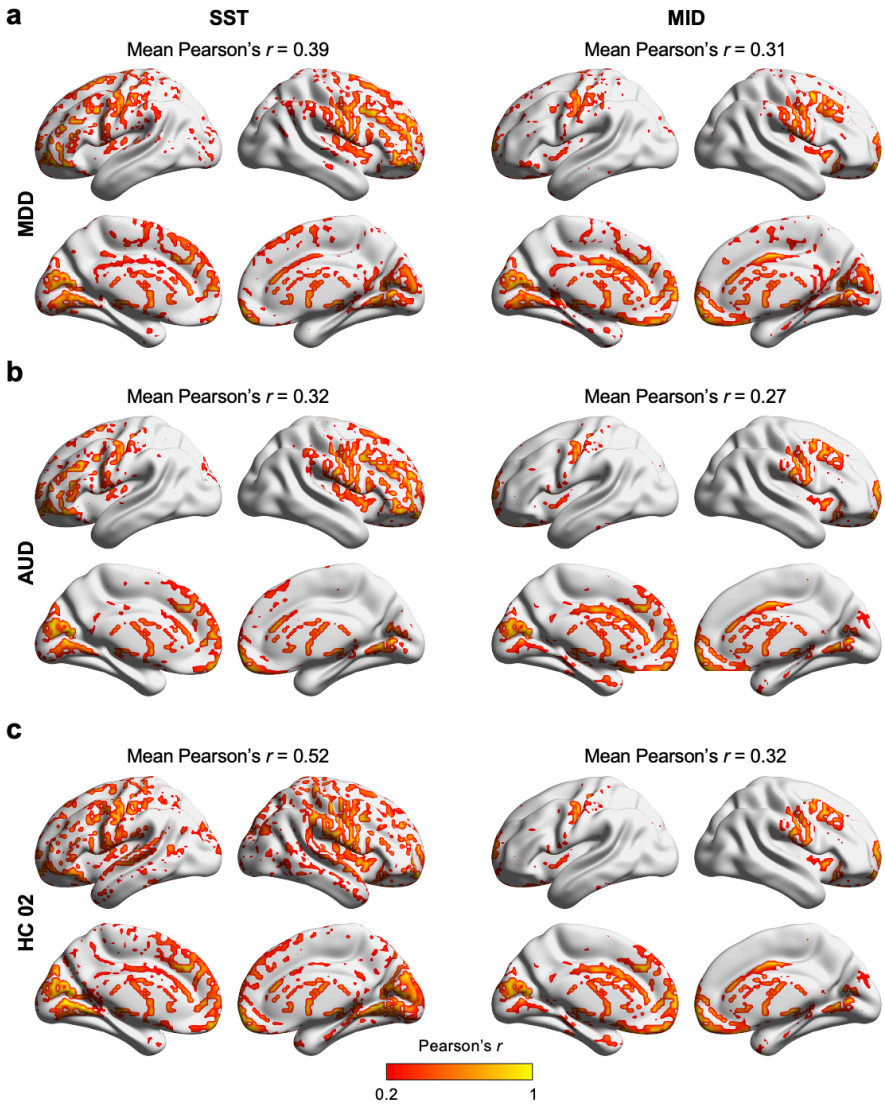

Fig. S5 Simulation performances of 1-billion-neuron models for three illustrative participants. Voxel-wise Pearson's correlations between empirical and simulated BOLD signals from the 1-billion-neuron DTB model for the MID and SST tasks of MDD, AUD, and HC02.

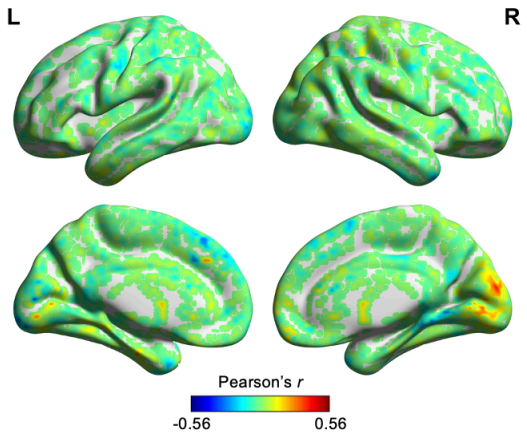

Fig. S6 A reference simulation of a 1-billion-neuron DTB with a random parameter shuffle. Voxel-wise correlations between empirical and simulated BOLD signals from the 1-billion-neuron DTB model for the MID task of HC01, using randomly shuffled hyperparameters for simulation. The whole average voxel-wise Pearson's $r = 0$.

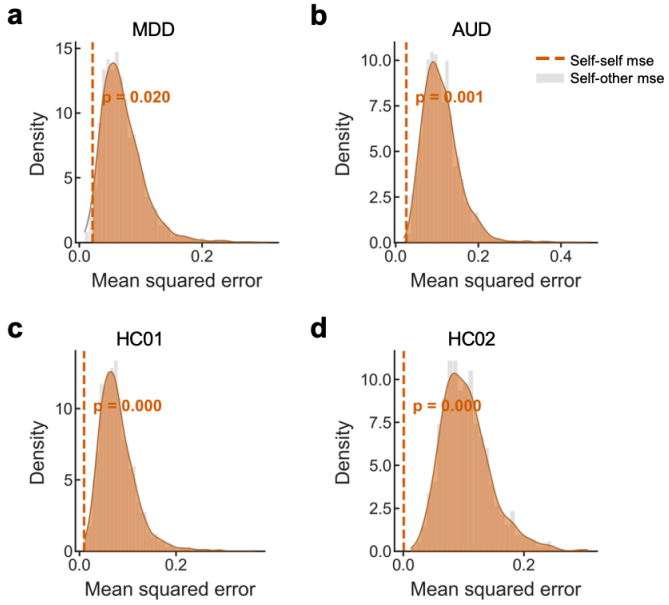

Fig. S7 Subject-specific correspondence between simulated and empirical NP factors. For each of four illustrative participants, the simulated NP factor showed the closest correspondence to own empirical NP score, yielding a significant lower mean squared error (MSE) than comparisons with other individuals from IMAGEN and STRATIFY cohorts (**a**, MDD, $P_{perm} = 0.02$; **b**, AUD, $P_{perm} = 0.001$; **c**, HC01, $P_{perm} < 0.001$; **d**, HC02, $P_{perm} < 0.001$; permutation tests across 1,575 simulated–empirical pairings). Kernel density distributions show the null distribution of MSE values obtained from self–other simulated–empirical pairings across all individuals (grey). The vertical dashed line indicates the observed self–self MSE for each participant (orange). Shaded areas represent the proportion of the null distribution exceeding the observed value.

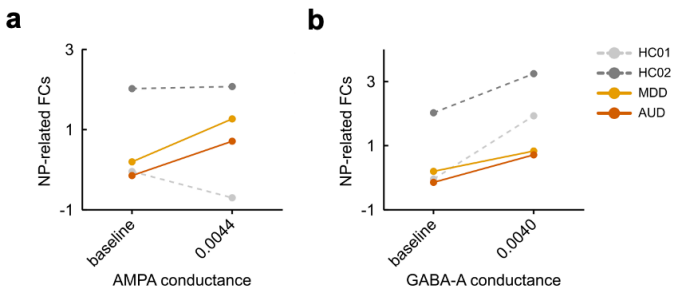

Fig. S8 Generalisation of bidirectional NP responses across task contexts. Following **a**, AMPA-mediated (increased to 0.0044) and **b**, GABA-A–mediated (increased to 0.0040) virtual perturbations, NP-related functional connectivity derived from simulated EFT BOLD signals exhibited heterogeneous, bidirectional responses across the same four illustrative participants. Patient models (MDD and AUD) showed increases in NP connectivity, one healthy control (HC01) showed a similar decrease, consistent with the bidirectional modulation patterns observed in the main task analyses.

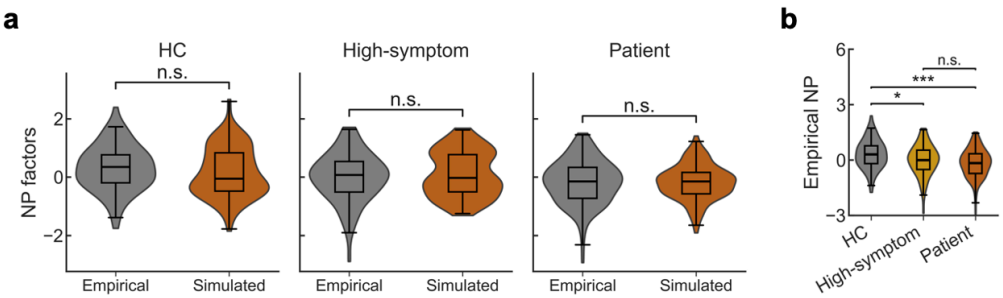

Fig. S9 Comparisons of simulated and empirical NP factor scores. a, Within each diagnostic group, no significant differences between empirical and simulated NP factor scores (healthy controls (HC): n = 69, $t = 1.08$, $P = 0.28$; high-symptom: n = 90, $t = -0.96$, $P = 0.34$; patients: n = 131, $t = 0.014$, $P = 0.99$, paired-sample t tests). Violin plots show kernel density estimates with embedded boxplots indicating the median and IQR; whiskers extend to 1.5× IQR. **b**, Empirical NP factor scores in healthy controls, high-symptom individuals, and patients. Group differences were assessed using ANOVA and post-doc tests with Bonferroni correction for multiple comparisons. Values are residuals after regressing out sex, site, and head motion. Violin plots show kernel density estimates with embedded boxplots indicating the mean and IQR; whiskers extend to 1.5× IQR. n.s., not significant; *, $P_{corrected}<0.05$; ***, $P_{corrected}<0.001$.

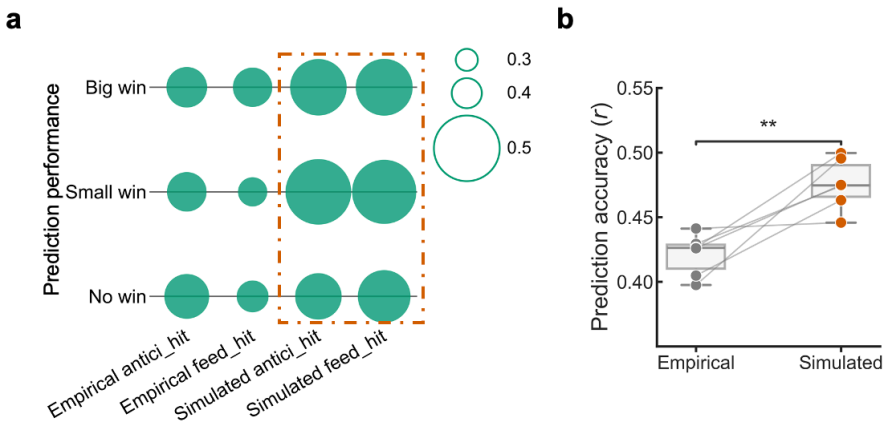

Fig. S10 Prediction of task performance from empirical and DTB-simulated functional connectivity. a, Predictive performances of task-specific functional connectivity (FC) derived from the Monetary Incentive Delay task were shown as a bubble plot, where bubble size indicates Spearman's correlation between predicted and empirical response times. Predictions were performed separately for big-win, small-win, and no-win conditions using whole-brain task-state FC derived from empirical fMRI data and DTB simulations. Both empirical and simulated FC significantly predicted response times across all conditions (n = 287; all $P_{corrected} < 0.0001$; Table S16). **b,** Prediction accuracy was consistently higher for DTB-simulated FC than for empirical FC across six prediction models (paired-sample *t* test, $t = -4.47$, $P = 0.0066$), with a maximal increase in explained variance of up to 9% (Δr^2). **, $P < 0.01$.

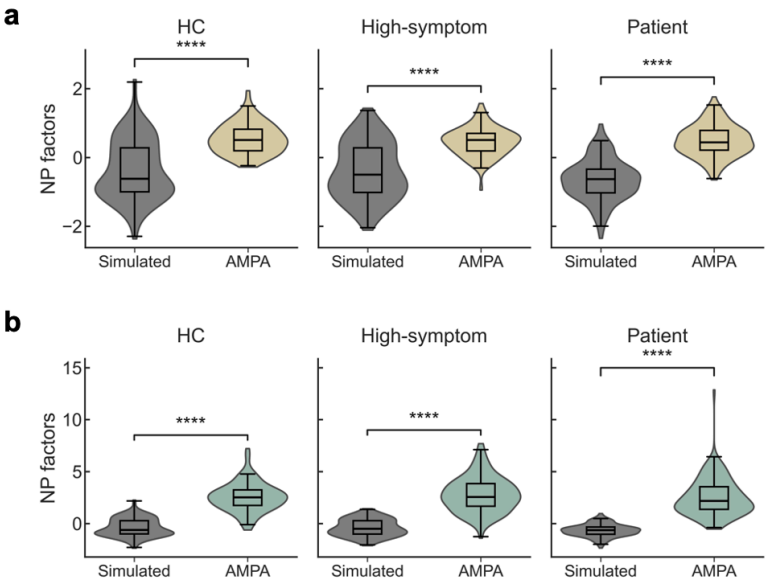

Fig. S11 Within-group comparisons of simulated NP factors between baseline and following AMPA and GABA-A modulations. **a,** Within each diagnostic group, simulated AMPA modulation significantly increased NP factor scores (HC: $t = -7.59$; high-symptom: $t = -8.74$; patients: $t = -18.60$, all P values < 0.0001). **b,** Similarly, GABA-A–mediated modulation also produced significant within-group increases (HC: $t = -15.57$; high-symptom: $t = -17.78$; patients: $t = -20.54$, all P values < 0.001; paired-sample t tests). Each panel shows the distribution of NP factor scores at simulated baseline (grey), following AMPA (gold) and GABA-A (teal) modulation. Violin plots depict kernel density estimates with embedded boxplots indicating the median and interquartile range (IQR). Asterisks denote within-group significance relative to baseline. The shared y-axis enables direct comparison of modulation magnitude across groups. ****, $P < 0.0001$.

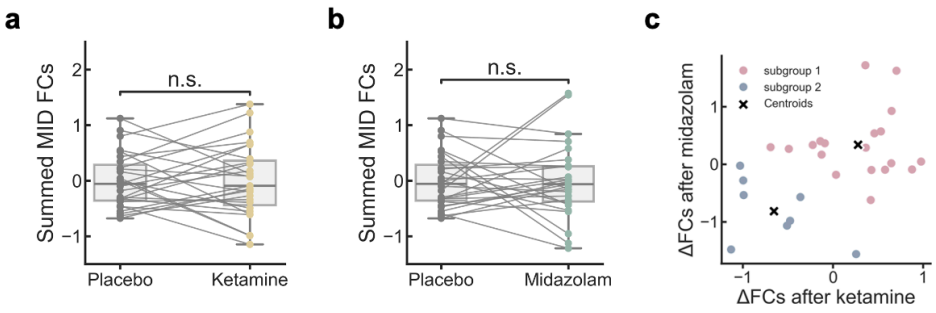

Fig. S12 Paired pharmacological modulations of NP-related MID functional connectivity. a,b, Paired comparisons of summed NP-related MID functional connectivity (FC) under placebo versus ketamine (**a**) and placebo versus midazolam (**b**). Each dot represents one individual, colored by drug condition (placebo, grey; ketamine, gold; midazolam, teal). Statistical significance was assessed using paired-sample *t* tests. **c**, K-means clustering of individual pharmacological response profiles based on drug-induced changes in summed NP-related MID FC. Each point represents one participant, color-coded by cluster assignment (subgroup 1, pink; subgroup 2, blue). Black crosses indicate cluster centroids. Values represent residuals after regressing out head motion. n.s., not significant.

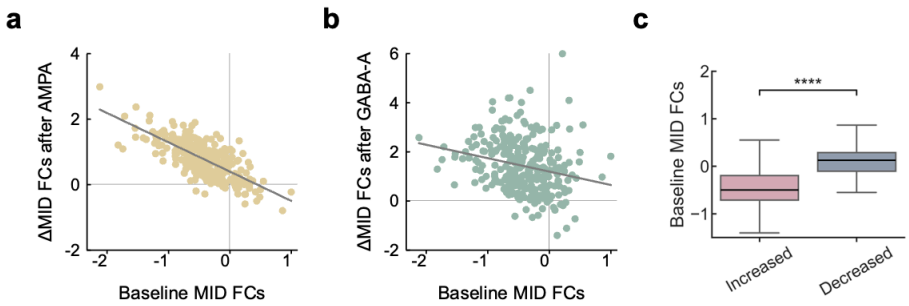

Fig. S13 Robustness of baseline-dependent, bidirectional responses in MID-specific NP connectivity. Virtual AMPA- and GABA-A-mediated perturbations were re-analysed in population-scale DTB simulations with the network phenotype restricted to NP-related MID-specific NP connectivity (summed MID FCs). **a,b**, Baseline summed MID FCs score was inversely correlated with AMPA- (**a**, Pearson's r = -0.74, $P < 0.0001$) and GABA-A- (**b**, Pearson's r = -0.24, $P < 0.0001$) induced changes in NP connectivity. **c**, The increased-response group exhibited lower simulated baseline summed MID FCs scores than the decreased-response group (Mann–Whitney U test, $P < 0.0001$). Boxplots indicate the median and interquartile range; whiskers extend to 1.5× IQR. ****, $P < 0.0001$.

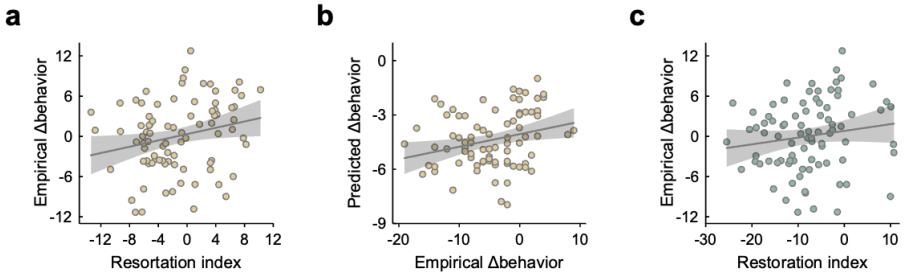

Fig. S14 Prediction of longitudinal symptom trajectories from DTB-derived virtual perturbations. a, Scatter plot showing the relationship between DTB-derived restoration index (difference between predicted symptom scores of baseline-simulated and AMPA-perturbed conditions) and empirical symptom changes after removing baseline symptom effects (residuals), illustrating the independent predictive contribution of the restoration index (t = 2.22, P = 0.029). **b,** Prospective prediction of behavioural symptom changes over 4 years (ages 19-23, n = 85) using restoration index for AMPA perturbation alone. Scatter plot shows predicted versus empirical behavioural changes from age 19 to 23. The prediction model explained a significant proportion of variance (r^2 = 0.048, F = 4.91, P = 0.029). **c,** Scatter plot showing the relationship between DTB-derived restoration index for GABA-A perturbation and empirical symptom changes after removing baseline symptom effects (residuals, t = 1.43, P = 0.16). Each dot represents one individual; solid line indicates the linear fit with 95% confidence intervals.

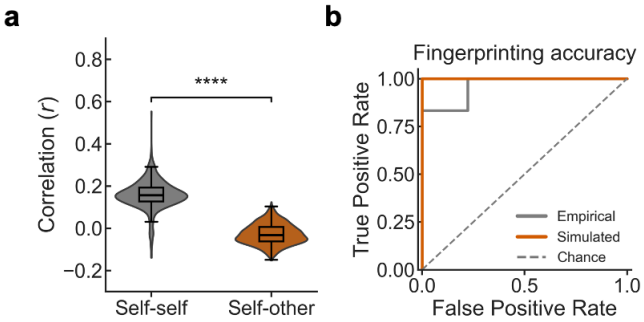

Fig. S15 Preservation of subject-specific signatures by DTBs without re-assimilation. One-billion-neuron digital twin brain models were constructed for four matched participants and used to simulate BOLD signals during the Emotional Face Task (EFT). **a,** Hyperparameters estimated from a single EFT trial were used to simulate all remaining trials across all participants. Trial-level activation maps were computed from simulated and empirical time series using identical procedures. Correlations between empirical and simulated trials using hyperparameters derived from the same participant (self–self condition: 12 trials × 4 participants × 12 trials) were significantly higher than those obtained using hyperparameters derived from other participants (self-other condition: 12 trials × 4 participants × 36 trials; Mann–Whitney U test, $P < 0.0001$). Violin plots show kernel density estimates with embedded boxplots indicating the median and interquartile range (IQR); whiskers extend to 1.5× IQR. **b,** Hyperparameters estimated from the first half of the EFT were applied to simulate the second half. Correlating the second half simulated and empirical activation maps (6 trials × 4 participants) with individual-specific templates yielded 100% fingerprinting accuracy for DTB simulations (AUC = 1.00, orange), compared with 95.8% for empirical trials (AUC = 0.958, grey). ****, $P <$ 0.0001.