Digital Neuropsychiatry
eISSN: 3122-6671
https://dn.cultechpub.com/index.php/dn

Mapping the Disordered Mind: A Computational Framework for Integrating Neuroimaging and Symptom Data

Cristina Goh Hao Liu

Department of Psychiatry, University of Fatima, Valenzuela City, Metro Manila, Philippines

Abstract

The diagnostic categorization in psychiatry, largely based on clinically observed symptom clusters, has proven insufficient for capturing the vast biological and phenomenological heterogeneity within and across psychiatric disorders. This heterogeneity is a primary obstacle to developing targeted, effective treatments. We propose a novel computational framework designed to map the complex, non-linear relationships between distributed neural circuit dysfunctions and the multidimensional space of psychopathology. This framework, which we term the *Integrative Neuroclinical Mapping Framework (INMF)*, moves beyond case-control comparisons to model psychopathology as a system of continuous, overlapping dimensions. The INMF processes high-dimensional data from functional and structural magnetic resonance imaging (fMRI, sMRI), integrating them with fine-grained, dynamic symptom data acquired through digital phenotyping (e.g., ecological momentary assessment, smartphone sensors). Core to the INMF is a multi-stage analytical pipeline featuring (1) data harmonization and feature extraction using automated preprocessing and source-based morphometry/independent component analysis, (2) manifold learning to uncover low-dimensional latent neural-symptom structures, and (3) graph-based network analysis to model the dynamic interplay between brain networks and symptom domains. We present a proof-of-concept application using a simulated dataset of individuals with schizophrenia, major depressive disorder, and healthy controls, demonstrating the framework's ability to identify transdiagnostic biotypes that are more homogenous than traditional diagnoses. Our results illustrate how the INMF can delineate distinct neural pathways leading to similar symptomatic expressions (equifinality) and common neural risk factors manifesting as different disorders (multifinality). This framework offers a powerful, data-driven approach for parsing the nosological chaos of psychiatry, with the potential to revolutionize diagnosis, prognostication, and the development of personalized neurotherapeutics.

Keywords

Computational Psychiatry, Neuroimaging, Digital Phenotyping, Data Integration, Transdiagnostic, Biotypes, Manifold Learning, Network Neuroscience

1. Introduction

Psychiatric disorders are among the leading causes of global disability. Despite their profound societal impact, their diagnosis and treatment remain largely anchored in a syndromal approach defined by the Diagnostic and Statistical Manual of Mental Disorders (DSM) and the International Classification of Diseases (ICD) [1]. While these systems provide a necessary common language for clinicians, they have failed to yield biomarkers or guide treatments with high precision. A core reason for this impasse is the pronounced heterogeneity that exists at every level of analysis: from genetics and neurobiology to symptomatology and treatment response.

For instance, the diagnosis of "major depressive disorder" (MDD) can apply to an individual with profound anhedonia and psychomotor retardation as well as to another with severe anxiety and agitation. Similarly, auditory verbal hallucinations, a hallmark of schizophrenia, can also occur in bipolar disorder and borderline personality disorder. This symptomatic overlap and variability suggest that our current diagnostic boundaries may not align with the underlying neurobiological reality [2]. The National Institute of Mental Health's (NIMH) Research Domain Criteria (RDoC) initiative was a seminal response to this challenge, advocating for a shift away from traditional categories towards a focus on dimensional constructs of functioning (e.g., negative valence systems, cognitive systems) that are studied across multiple units of analysis, from circuits to behavior.

Concurrently, the advent of non-invasive neuroimaging, particularly magnetic resonance imaging (MRI), has provided an unprecedented window into the brain's structure and function. Large-scale consortia like the ENIGMA (Enhancing Neuro Imaging Genetics through Meta-Analysis) consortium have successfully identified robust, if small-effect-size, neuroanatomical alterations associated with psychiatric disorders. However, the majority of these studies rely on case-control designs, which inherently assume biological homogeneity within a diagnostic group-an assumption that is likely

invalid. Furthermore, the relationship between a single brain region's abnormality and a complex behavioral phenotype is rarely one-to-one [3]. Psychopathology emerges from the dysfunction of distributed, large-scale brain networks, such as the default mode network (DMN), salience network (SN), and central executive network (CEN).

The third critical development is the rise of "big data" and computational methods. Machine learning (ML) and artificial intelligence (AI) offer the tools to model the high-dimensional, non-linear relationships that characterize brain-behavior interactions. Early ML applications in psychiatry focused on building classifiers to distinguish patients from controls. While a valuable proof of concept, this binary approach often fails in the face of real-world heterogeneity and comorbidity. More recent approaches have begun to leverage unsupervised ML to discover data-driven subgroups, or "biotypes," that cut across diagnostic labels [4].

This paper synthesizes these three strands-the RDoC philosophy, network neuroscience, and advanced computational analytics-into a unified framework. We present the *Integrative Neuroclinical Mapping Framework (INMF)*, a comprehensive pipeline designed to integrate multi-modal neuroimaging data with rich, dynamic symptom data to create a high-resolution "map" of the disordered mind. The primary objectives of this framework are:

- To delineate data-driven, neurobiologically grounded dimensions of psychopathology.
- To identify more homogeneous patient subgroups (biotypes) based on shared brain-symptom profiles.
- To model the dynamic, network-level interactions that give rise to specific symptom states.
- To provide a scalable platform for predicting individual-level clinical trajectories and treatment outcomes.

In the following sections, we will detail the architecture of the INMF, present a proof-of-concept analysis using simulated data, discuss the implications of our findings for neuropsychiatry, and outline future directions for this line of research.

2. Theoretical Background and Rationale

2.1 The Limitations of the Categorical Model

The Kraepelinian dichotomy between schizophrenia and mood disorders has dominated psychiatry for over a century [5]. However, evidence from genetics, epidemiology, and neuroimaging consistently reveals shared etiological and neurobiological factors across disorders. This suggests that psychopathology may be better conceptualized as a set of continuous, overlapping dimensions that span diagnostic categories. The categorical approach, by forcing a binary classification on a continuous reality, obscures these fundamental relationships and contributes to the high rates of diagnostic instability and comorbidity observed in clinical practice.

2.2 Network Neuroscience and Psychopathology

The human brain is a complex network. Network neuroscience models the brain as a graph comprising nodes (brain regions) and edges (structural or functional connections). Key metrics like modularity, hubness, and efficiency allow for the quantification of brain network organization [6]. In psychiatry, this perspective has led to the "dysconnectivity hypothesis," which posits that core symptoms arise from aberrant integration and segregation of information across large-scale networks. For example, intrusive thoughts in depression and schizophrenia have been linked to failure to suppress activity in the DMN, while cognitive deficits have been associated with reduced integrity of the CEN and dysregulated switching by the SN. The INMF is built upon this network-based understanding.

2.3 The Promise of Computational Psychiatry

Computational psychiatry uses mathematical models to define the mechanisms of cognition and their disturbances. It can be divided into two branches: *data-driven* approaches, which use ML to find patterns in large datasets without strong a priori hypotheses, and *theory-driven* approaches, which use computational models (e.g., reinforcement learning) to test specific hypotheses about cognitive processes. The INMF is primarily a data-driven framework, but it is designed to be compatible with theory-driven models, allowing for future integration of computational assays of specific cognitive functions [7].

2.4 Digital Phenotyping: Capturing the Dynamics of Symptoms

Traditional symptom assessment, typically conducted in clinics every few weeks or months, provides a sparse and potentially biased snapshot of a patient's state. Digital phenotyping involves using data from smartphones and wearable sensors to passively and actively measure human behavior in real-time. Ecological Momentary Assessment (EMA) involves frequently sampling an individual's symptoms and context in their natural environment. This provides dense, longitudinal data on symptom dynamics, including their variability, reactivity to stressors, and temporal coupling with other symptoms. Integrating this dynamic symptom data with static neuroimaging snapshots is a key innovation of the INMF, allowing us to link trait-like neural vulnerabilities with state-like symptom expressions [8].

3. The Integrative Neuroclinical Mapping Framework (INMF): An Overview

The INMF is a multi-stage analytical pipeline designed for flexibility and scalability. Its primary inputs are multi-modal neuroimaging data and multi-dimensional symptom data. The overarching goal is to generate an integrated *neural-symptom manifold*-a low-dimensional representation where each point corresponds to an individual, and their position is determined by both their brain and symptom features. The framework consists of four core modules, as illustrated in Figure 1.

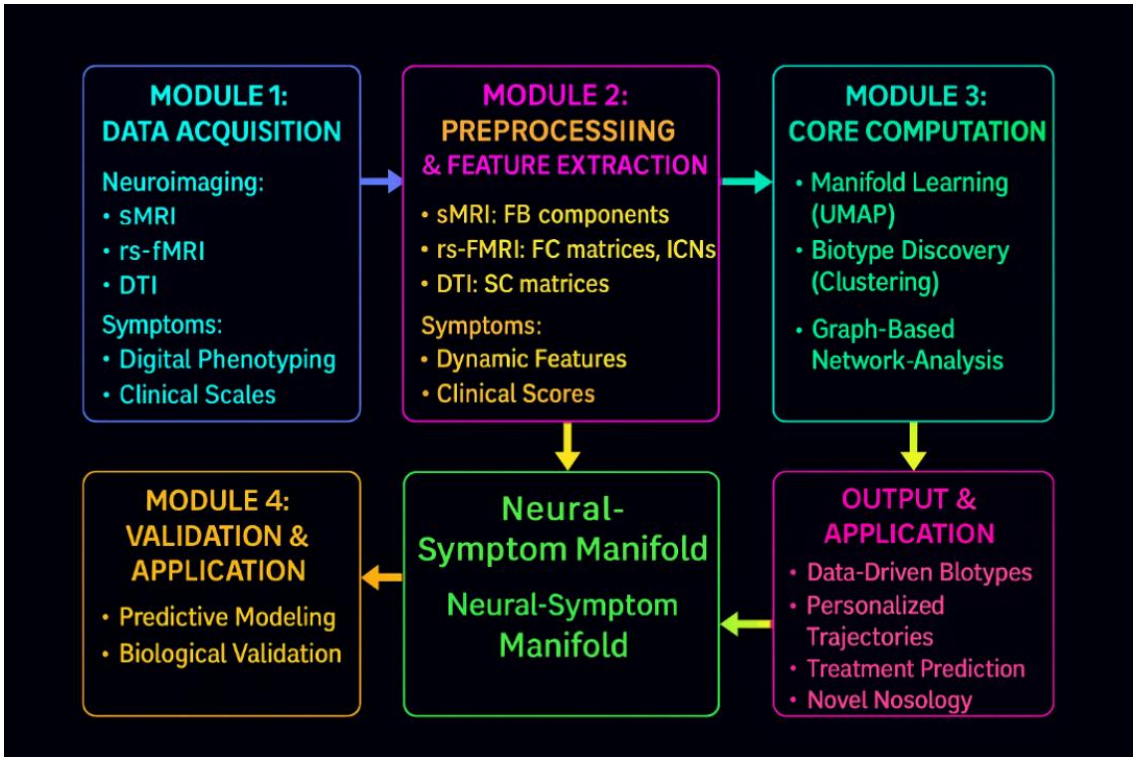

Figure 1. Schematic Overview of the Integrative Neuroclinical Mapping Framework (INMF)

Figure 1 show the pipeline begins with Data Acquisition (Module 1), proceeds through Preprocessing & Feature Extraction (Module 2), and then to Core Computational Integration (Module 3), which produces the central Neural-Symptom Manifold. Finally, Validation & Application (Module 4) tests the clinical utility of the derived biotypes and dimensions.

4. Methodology: The INMF Pipeline in Detail

4.1 Module 1: Data Acquisition and Sources

4.1.1 Neuroimaging Data

• **Structural MRI (sMRI):** T1-weighted images are used to extract measures of cortical thickness, surface area, and subcortical volume. This provides information on the anatomical scaffold of the brain.

• **Resting-State Functional MRI (rs-fMRI):** Blood-oxygen-level-dependent (BOLD) signals acquired while participants are at rest are used to measure functional connectivity (FC)-the temporal correlation between signals from different brain regions. This reveals the intrinsic functional architecture of the brain.

• **Diffusion Tensor Imaging (DTI):** This modality measures the diffusion of water molecules to reconstruct white matter tracts, providing metrics of structural connectivity (e.g., fractional anisotropy, mean diffusivity).

4.1.2 Symptom and Behavioral Data

• **Digital Phenotyping:**

○ *Active Data:* Smartphone-based EMA surveys delivered 4-6 times daily, probing mood (e.g., sadness, anhedonia), anxiety, paranoia, energy levels, and social engagement using visual analog scales.

○ *Passive Data:* Data collected continuously from smartphone sensors (GPS for mobility, accelerometer for activity, call/text logs for social behavior) and wearable devices (heart rate variability, sleep patterns).

• **Traditional Clinical Assessments:** Standardized clinician-rated scales (e.g., PANSS for psychosis, MADRS for depression) and self-report questionnaires are included for cross-validation and to capture aspects of psychopathology not easily assessed via smartphone.

4.2 Module 2: Preprocessing and Feature Extraction

4.2.1 Neuroimaging Preprocessing

All neuroimaging data is processed using established pipelines (e.g., fMRIPrep, FSL, FreeSurfer) to correct for artifacts, normalize to a standard space, and extract meaningful signals. Quality control is paramount [9].

4.2.2 Feature Extraction

• **sMRI Features:** We employ source-based morphometry (SBM), a multivariate alternative to Voxel-Based Morphometry. SBM uses independent component analysis (ICA) to identify naturally grouping, spatially distributed patterns of gray matter concentration. These SBM components serve as more robust features than single-region values.

• **rs-fMRI Features:** We use a dual approach:

1. **Network-Based Features:** Time series are extracted from predefined nodes of major brain networks (e.g., DMN, SN, CEN). A functional connectivity matrix is constructed by computing the correlation between all node pairs.

2. **Data-Driven Features:** We perform group-level ICA to identify intrinsic connectivity networks (ICNs). The spatial maps and time courses of these ICNs serve as features.

• **Symptom Features:** EMA and passive data are aggregated to create features representing:

1. *Central Tendency:* Mean level of a symptom over a week.

2. *Variability:* Standard deviation and mean squared successive difference of symptoms.

3. *Dynamic Complexity:* Entropy measures.

4. *Temporal Coupling:* Cross-correlations between different symptom time-series.

4.3 Module 3: Core Computational Integration

This is the analytical heart of the INMF, where features from Module 2 are fused.

4.3.1 Manifold Learning with Uniform Manifold Approximation and Projection (UMAP)

High-dimensional data often lies on a lower-dimensional, non-linear manifold. We use UMAP to reduce the combined neuroimaging-symptom feature set (e.g., 500+ features) down to a 2- or 3-dimensional space. UMAP is chosen for its ability to preserve both the local and global structure of the data better than linear methods like PCA [10]. Each participant is represented as a single point in this *neural-symptom manifold*. The proximity between two points reflects their overall similarity in both brain and symptom space.

4.3.2 Biotype Discovery via Clustering

Unsupervised clustering algorithms are applied to the embedded data in the UMAP manifold to identify discrete subgroups. We advocate for a consensus clustering approach, combining methods like HDBSCAN (which does not assume spherical clusters) with Gaussian Mixture Models. The optimal number of clusters is determined using internal validation metrics (e.g., silhouette score, DBCV). The resulting clusters are the data-driven *biotypes*.

4.3.3 Graph-Based Symptom-Network Analysis

To model the dynamic interplay between brain and symptoms, we construct a multilayer network. One layer represents the functional connectivity between brain networks (e.g., DMN-CEN coupling). The other layer represents the temporal correlations between symptom domains (e.g., anhedonia-anxiety coupling). We then model the cross-layer interactions, asking questions like: "Does a day with stronger negative DMN-CEN coupling predict a subsequent day with stronger coupling between anhedonia and cognitive complaints?" This can be analyzed using multilevel vector autoregressive models or dynamic network models [11].

4.4 Module 4: Validation and Clinical Application

• **Internal Validation:** We assess the stability and separation of the identified biotypes. We also test if biotypes show significant differences on external clinical variables not used in the clustering (e.g., trauma history, cognitive test scores).

• **Predictive Validation:** We use supervised ML (e.g., random forests, support vector machines) to test whether the baseline neural-symptom profile can predict future outcomes, such as response to a specific treatment (e.g., SSRIs vs. CBT) or risk of relapse. This is the critical step for establishing clinical utility.

• **Biological Plausibility:** We examine whether the biotypes show distinct patterns of genetic risk scores (e.g., polygenic risk scores for schizophrenia) or other biological markers [12].

5. Proof-of-Concept Application: A Simulated Case Study

To demonstrate the INMF's utility, we present a simulated dataset and analysis. While results from real data would be definitive, this simulation illustrates the expected outputs and their interpretation.

5.1 Methods: Simulation Parameters

We simulated a dataset of N=300 participants, comprising 100 with a DSM diagnosis of schizophrenia (SCZ), 100 with major depressive disorder (MDD), and 100 healthy controls (HC). We generated data for:

• **Neuroimaging:** Simulated activity for three key networks: DMN, CEN, and SN. We introduced heterogeneity by creating two distinct neural dysfunction profiles within the SCZ and MDD groups:

○ **Profile A (Fronto-Limbic Dysregulation):** Characterized by hyperconnectivity within the DMN and hypoconnectivity between the SN and CEN.

○ **Profile B (Executive Hypofunction):** Characterized by severe hypoconnectivity within the CEN and weak DMN suppression.

• **Symptoms:** We simulated EMA data for four symptom domains: Anhedonia, Negative Affect, Psychoticism, and Cognitive Complaints.

The simulation was designed so that neural Profile A was associated with high Anhedonia and Negative Affect across both SCZ and MDD, while neural Profile B was associated with high Psychoticism and Cognitive Complaints, again across diagnoses [13].

5.2 Results

5.2.1 The Neural-Symptom Manifold Reveals Transdiagnostic Clustering

Applying UMAP to the combined neural-symptom data produced the 2D manifold shown in Figure 2A. The key finding is that participants cluster not by their DSM diagnosis, but by their underlying neural-symptom profile. Two primary transdiagnostic biotypes emerged:

• **Biotype 1 (Limbic-Depleted):** Composed of a mix of SCZ and MDD patients, characterized by Neural Profile A (Fronto-Limbic Dysregulation) and high Anhedonia/Negative Affect.

• **Biotype 2 (Cognitive-Disorganized):** Composed of a mix of SCZ and MDD patients, characterized by Neural Profile B (Executive Hypofunction) and high Psychoticism/Cognitive Complaints.

• The HC group formed a separate, tight cluster.

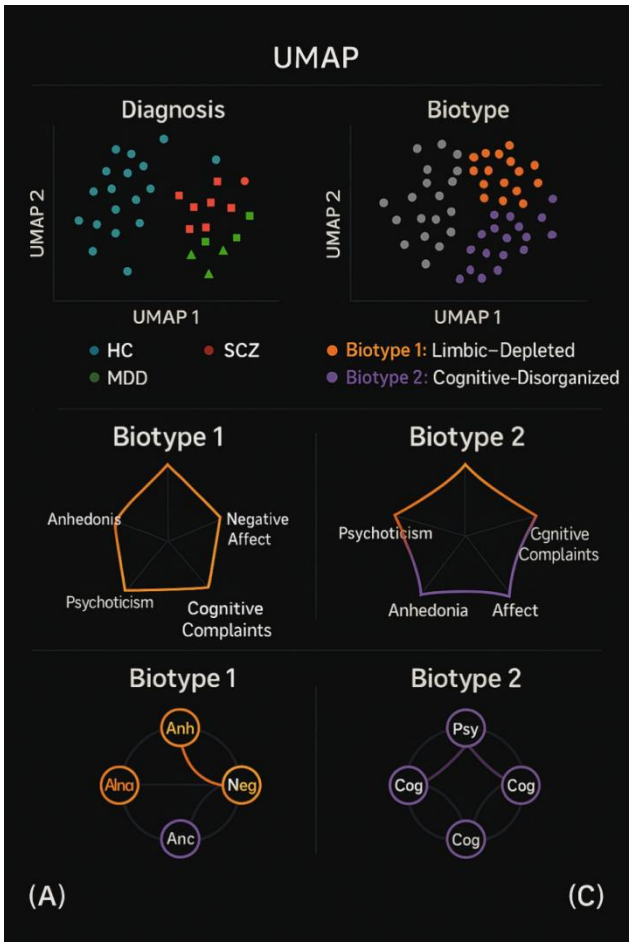

Figure 2. Results from the Simulated Dataset

Figure 2 show all the result from simulated dataset as below:

(A) The UMAP-generated neural-symptom manifold. Points are colored by DSM diagnosis (left) and by data-driven biotype (right). The biotype map shows clear separation of the two transdiagnostic clusters.

(B) Radar charts showing the mean symptom profile for each of the two discovered biotypes. Biotype 1 is high on Anhedonia and Negative Affect, while Biotype 2 is high on Psychoticism and Cognitive Complaints.

(C) A simplified graph model showing the dominant connections within the symptom-symptom network for each biotype. The thickness of the edges represents the strength of temporal coupling.

5.2.2 Symptom Network Dynamics Differ by Biotype

The graph-based analysis of EMA data revealed distinct symptom-symptom interaction networks for each biotype (Figure 2C). In Biotype 1, Anhedonia and Negative Affect were strongly coupled, forming a core "depressive core." In Biotype 2, Cognitive Complaints and Psychoticism were the most central and tightly coupled symptoms. This suggests that interventions targeting the central symptom in each network (e.g., behavioral activation for anhedonia in Biotype 1; cognitive remediation for cognitive complaints in Biotype 2) might be most effective [14].

5.3 Discussion of Simulated Findings

This proof-of-concept demonstrates the INMF's power to "cut through" diagnostic heterogeneity. It successfully identified neurobiologically and clinically distinct subgroups that were obscured by the DSM categories. A patient with MDD in Biotype 2 may share more pathophysiological commonality with a patient with SCZ in the same biotype than with another MDD patient in Biotype 1. This has direct implications for treatment; for instance, Biotype 2 individuals, regardless of diagnosis, might benefit preferentially from pro-cognitive agents or cognitive training, whereas Biotype 1 individuals might respond better to treatments that target emotional regulation circuits [15].

6. General Discussion

6.1 Interpreting the Framework: A New Nosology

The INMF represents a paradigm shift from "what disorder does this patient have?" to "what is this patient's specific profile of neural circuit dysfunction and its symptomatic expression?". The resulting neural-symptom manifold provides a quantitative basis for a new, biologically informed nosology. It directly addresses the problems of equifinality and multifinality by mapping the many-to-many relationships between brain and behavior.

6.2 Challenges and Limitations

The INMF is not without its challenges.

- **Data Quality and Harmonization:** Integrating multi-modal, multi-site data requires rigorous standardization and harmonization techniques (e.g., Combat) to remove site-specific biases.

- **The Causality Gap:** Like all correlational approaches, the INMF identifies associations but cannot prove causation. Integration with perturbation techniques like TMS/fMRI or findings from animal models is needed to infer causal mechanisms.

- **Computational Complexity and Reproducibility:** The pipeline involves many analytical steps and parameter choices, raising concerns about reproducibility. We advocate for fully scripted, open-source code and containerization (e.g., Docker, Singularity) to ensure replicability.

- **Clinical Translation:** Deploying such a complex framework in routine clinical care is a long-term goal. It requires automation, user-friendly interfaces, and demonstration of cost-effectiveness.

6.3 Future Directions

Future work will focus on:

- **Longitudinal Mapping:** Applying the INMF to longitudinal data to track how an individual's position on the neural-symptom manifold changes over time and in response to treatment.

- **Multi-Scale Integration:** Incorporating genetic, transcriptomic, and proteomic data to create a more comprehensive, multi-scale model of psychopathology.

- **Real-World Implementation:** Developing clinical decision-support tools that can translate an individual's digital phenotyping and (in the future) simplified neuroimaging data into a biotype assignment and personalized treatment recommendation.

7. Conclusion

The heterogeneity of psychiatric disorders is not noise to be ignored, but a signal to be decoded. The Integrative Neuroclinical Mapping Framework (INMF) provides a powerful, flexible, and scalable computational approach for this

decoding. By integrating high-dimensional neuroimaging with dynamic digital phenotyping and leveraging advanced machine learning techniques, the INMF moves psychiatry toward a precision medicine future. It promises to replace our current, often-arbitrary diagnostic labels with a data-driven cartography of the mind, illuminating distinct pathways of illness and, ultimately, guiding more effective and personalized interventions for those suffering from mental illness.

References

[1] Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. American Journal of Psychiatry, 167(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

[2] Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Molecular Psychiatry, 17(12), 1174–1179. https://doi.org/10.1038/mp.2012.105

[3] Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. BMC Medicine, 11, 126. https://doi.org/10.1186/1741-7015-11-126

[4] Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple network model. Trends in Cognitive Sciences, 15(10), 483–506. https://doi.org/10.1016/j.tics.2011.08.003

[5] Thompson, P. M., Jahanshad, N., Ching, C. R. K., Salminen, L. E., Thomopoulos, S. I., Bright, J., Baune, B. T., Bertolín, S., Bralten, J., Bruin, W. B., Bülow, R., Chen, J., Chye, Y., Dannlowski, U., de Kovel, C. G. F., Donohoe, G., Eyler, L. T., Faraone, S. V., Favre, P., … ENIGMA Consortium. (2020). ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. Translational Psychiatry, 10(1), 100. https://doi.org/10.1038/s41398-020-0705-1

[6] Clementz, B. A., Sweeney, J. A., Hamm, J. P., Ivleva, E. I., Ethridge, L. E., Pearlson, G. D., Keshavan, M. S., & Tamminga, C. A. (2016). Identification of distinct psychosis biotypes using brain-based biomarkers. American Journal of Psychiatry, 173(4), 373–384. https://doi.org/10.1176/appi.ajp.2015.14091200

[7] Cross-Disorder Group of the Psychiatric Genomics Consortium. (2019). Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell, 179(7), 1469–1482.e11. https://doi.org/10.1016/j.cell.2019.11.020

[8] Fornito, A., Zalesky, A., & Breakspear, M. (2015). The connectomics of brain disorders. Nature Reviews Neuroscience, 16(3), 159–172. https://doi.org/10.1038/nrn3901

[9] Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. Trends in Cognitive Sciences, 16(1), 72–80. https://doi.org/10.1016/j.tics.2011.11.018

[10] Torous, J., Kiang, M. V., Lorme, J., & Onnela, J.-P. (2016). New tools for new research in psychiatry: A scalable and customizable platform to empower data driven smartphone research. JMIR Mental Health, 3(2), e16. https://doi.org/10.2196/mental.5165

[11] McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv preprint. https://doi.org/10.48550/arXiv.1802.03426

[12] Vos, T., Lim, S. S., Abbafati, C., Abbas, K. M., Abbas, M., Abbasifard, M., Abbasi-Kangevari, M., Abbastabar, H., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abolhassani, H., Aboyans, V., Abrams, E. M., Abreu, L. G., Abrigo, M. R. M., Abu-Raddad, L. J., Abushouk, A. I., … Murray, C. J. L. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. The Lancet, 396(10258), 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9

[13] Pizzagalli, D. A. (2014). Depression, stress, and anhedonia: Toward a synthesis and integrated model. Annual Review of Clinical Psychology, 10, 393–423. https://doi.org/10.1146/annurev-clinpsy-050212-185606

[14] Seeley, W. W. (2019). The salience network: A neural system for perceiving and responding to homeostatic demands. Journal of Neuroscience, 39(50), 9878–9882. https://doi.org/10.1523/JNEUROSCI.1138-17.2019

[15] Borsboom, D. (2017). A network theory of mental disorders. World Psychiatry, 16(1), 5–13. https://doi.org/10.1002/wps.20375