PheMIME: An Interactive Web App and Knowledge Base for Phenome-Wide Multi-Institutional Multimorbidity Analysis

Siwei Zhang[1], Nick Strayer[2], Tess Vessels[3,4], Karmel Choi[5,6], Geoffrey W Wang[7], Yajing Li[1], Cosmin A Bejan[8], Ryan S Hsi[9], Alexander G. Bick[10], Justin M Balko[11], Douglas B Johnson[11], Digna R Velez Edwards[12], Michael R Savona[11], Elizabeth J Philips[13,14], Dan M Roden[15], Jordan Smoller[5,6,16], Douglas M Ruderfer[3,4,17], Yaomin Xu[1,8 *]

1. Department of Biostatistics, Vanderbilt University, Nashville, TN, USA

2. Posit PBC, Boston, MA, USA

3. Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA

4. Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

5. Psychiatric & Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston MA

6. Center for Precision Psychiatry, Department of Psychiatry, Massachusetts General Hospital, Boston MA

7. Department of Statistics, North Carolina State University

8. Department of Biomedical informatics, Vanderbilt University, Nashville, TN, USA

9. Department of Urology, Vanderbilt University Medical Center, Nashville, TN, USA

10. Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

11. Division of Hematology and Oncology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

12. Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, USA

13. Center for Drug Safety and Immunology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

14. Institute for Immunology and Infectious Diseases, Murdoch University, Murdoch, Western Australia, Australia

15. Department of Pharmacology, Vanderbilt University Medical Center, Nashville, TN, USA

16. Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA

1. Background

Multimorbidity, the simultaneous occurrence of multiple diseases within an individual, is posing significant challenge to healthcare systems [1–4]. In depth understanding of intricate multimorbidity reveals shared molecular mechanism among various diseases [5–7], offering opportunities for innovative prevention strategies, targeted interventions, and personalized treatments for patients with multimorbidities [8–11]. Large-scale electronic health record (EHR) systems have significantly enhanced the statistical power to examine robust multimorbidity patterns that mirror real-world scenarios[7,12–16]. Network analysis has been a powerful approach to decipher complex multimorbidity patterns [17–19]. Integrating multiple EHR systems can improve the generalizability of multimorbidity characterization. However, current standards and tools limit the characterization of multimorbidity patterns across different populations, particularly in phenome-wide analyses using large-scale electronic health record (EHR) systems[1,20]. In this study, we build a phenome-wide multimorbidity knowledge base compiled from three major EHR databases: Vanderbilt University Medical Center (VUMC), Massachusetts General Brigham Hospital (MGB), and UK Biobank (UKB). And we developed an interactive web application called Phenome-wide Multi-Institutional Multimorbidity Explorer (PheMIME), which enables researchers to interactively explore, compare and discover multimorbidity patterns.

2. Methods

2.1. Data Integration and Multimorbidity Knowledge Base

We accessed three EHR databases that included individual-level data for 250,000 random patients each from VUMC and MGB, as well as data from 431,105 subjects in the UKB. The ICD 10 codes were mapped to phecodes and logistic regressions adjusting for patient age at last

recorded visit, sex, race, and the number of unique phecodes present in patients' records were run for each pair of two phecodes with two conditions of either phecode A or phecode B regarded as the outcome and the other one treated as the independent variable [19]. The averaged test statistic from the two regression analyses is then used as an estimate for the multimorbidity strength between the phecode pair A and B. Multimorbidity strengths of all pairs are subsequently calculated and used to construct a phenome-wide multimorbidity network, which represents an undirected weighted networks with disease as nodes and disease-disease connections as edges, weighted by the multimorbidity strengths. In addition, Pearson correlation of the common multimorbidity patterns between each pair of phecodes is used as a similarity score, which was also used to generate another undirected weighted network using the similarity scores as weights with disease as nodes and connections as edges. We call this a multimorbidity similarity network. We finally consolidate all the summarized data from three institutions into a database.

2.2. Design Scheme

PheMIME incorporates five primary modules: (1) "Disease Selection" module facilitates an interactive phecode table where users can seamlessly search, filter, and select a disease phecode of interest. (2) "Multimorbidity Consistency Inspection" module enables users to assess the overall consistency of multimorbidity strength measurements from the knowledge base and compare them across multiple institutions**.** Additionally, this module incorporates features that underscore significant multimorbidity strengths linked to the selected phecode of interest, aids in assessing their distribution amidst all multimorbidity measurements, and enables comparison across institutions. **(**3) "Multimorbidity Network Visualization" module presents interactive visual representations of the multimorbidity networks constructed based on the multimorbidity

strength measurements. By integrating a dynamic network visualization and clustering methodology called associationSubgraphs [17], this module permits exploration of the network's subgraph structures and dynamic clustering for any multimorbidity network from a single institution or an amalgamation of multiple institutions. Moreover, this module enables users to apply filters and emphasize any significant multimorbidities and investigate their interconnections and enriched subgraphs. (4) "Reproducible Multimorbidities Exploration" module provides an interactive environment for examining a customizable subset of reproducible multimorbidities across the institutions. This interface allows users to visualize the interconnections among chosen phecodes and the enriched multimorbidity subgraphs within the combined multimorbidity networks. Furthermore, this module accommodates pairwise comparisons between all institutions. (5) "Multimorbidity Similarities Exploration" module, uses multimorbidity similarity measurements as the strength measurement. It permits visualization of interconnected phecodes and the multimorbidity subgraphs enriched in the combined multimorbidity similarity networks and enables pairwise comparisons between all institutions.

3. Results

The PheMIME knowledge base and web application are accessible at https://prod.tbilab.org/PheMIME/. A comprehensive tutorial, including a use-case example, is available at https://prod.tbilab.org/PheMIME_supplementary_materials/. Furthermore, the source code for PheMIME can be freely downloaded from https://github.com/tbilab/PheMIME. Once a disease of interest is selected, the tool generates interactive visualization and tables that users can delve into multimorbidity networks within a single system or compare across multiple systems. Figure 1A shows an interactive Manhattan plot and a scatter plot that enables users to select and highlight a consistent set of the same phecodes based on the magnitude and consistency of

disease multimorbidities. The data table in Figure 1B shows comorbid phecodes of the user-selected phecode, its description, disease categories and corresponding multimorbidity strengths among three institutions. This table is interactive with the Manhattan and Scatter plots, allowing users to add or remove phecodes by clicking on the rows in the table. If a disease multimorbidity exhibits both a large magnitude and high consistency across different systems, it strongly indicates a robust disease multimorbidity across the systems. Figure 1C presents the interactive network visualization to explore subgraph structure of multimorbiditeis. For dynamic network analysis, the associationSubgraphs method [17] has been enhanced to provide an interactive visualization to rapidly explore subgraph structures. As shown in Figure 1C, network nodes are annotated into two groups, with the selected phecodes color-filled based on disease categories and the other unselected phecodes (nodes) color-filled in grey. Users have the ability to visualize the subgraphs that enhances the selected phecodes within the combined multimorbidity networks and users can easily identify the diseases present in these enriched subgraphs.

4. Conclusion

We have introduced PheMIME, an interactive visualization tool specifically designed for analyzing multimorbidities across multiple institutions, which simultaneously presents an extensive multimorbidity knowledge base consolidating data from three major EHR systems. To our understanding, PheMIME is the first knowledge base of its kind, integrating and comparing data from multiple extensive EHR systems while providing substantial support for efficient online analysis and interactive visualization, aiding in the discovery of complex multimorbidity patterns.

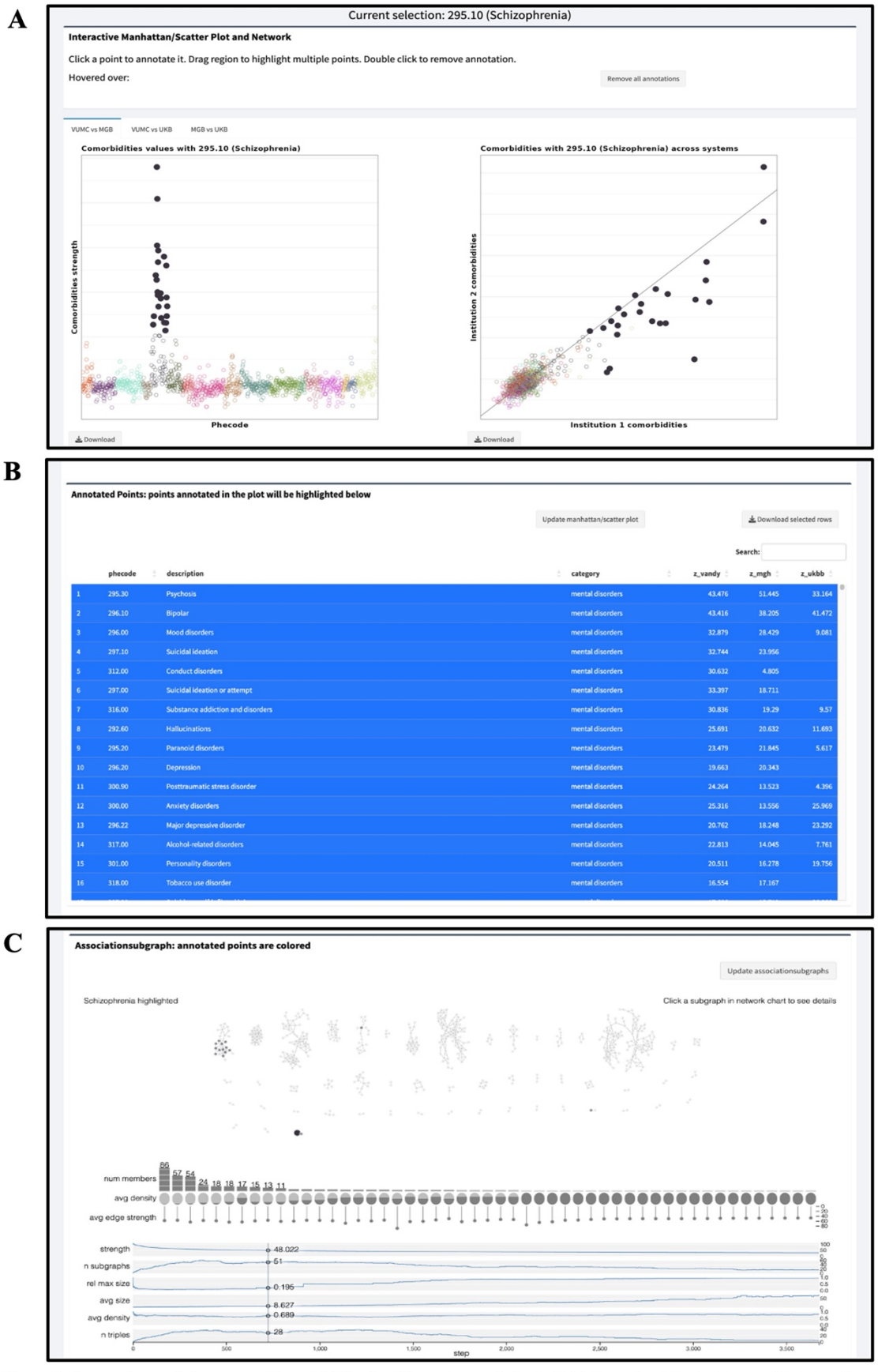

Figure 1: Comparison of disease multimorbidities in different populations. (A) Interactive Manhattan and Scatter plots comparing different cohorts, with schizophrenia (Phecode 295.10) selected as the phecode of interest. (B) Data table displaying comorbid phecodes associated with schizophrenia and the corresponding comorbidity strength values. The selected phecodes from part A are highlighted, and the table allows for user interaction with the plots. (C) Dynamic network analysis using associationSubgraphs, highlighting the user-selected phecodes from part A and their corresponding subgraphs.

References

1. Pearson-Stuttard, J., Ezzati, M. & Gregg, E. W. Multimorbidity—a defining challenge for health systems. *Lancet Public Health* **4**, e599–e600 (2019).
2. Skou, S. T. *et al.* Multimorbidity. *Nat Rev Dis Primers* **8**, 48 (2022).
3. Morris, J. E. *et al.* Treatment burden for patients with multimorbidity: cross-sectional study with exploration of a single-item measure. *British Journal of General Practice* **71**, e381–e390 (2021).
4. Hounkpatin, H. O. *et al.* Change in treatment burden among people with multimorbidity: a follow-up survey. *British Journal of General Practice* **72**, e816–e824 (2022).
5. Fabbri, E. *et al.* Aging and Multimorbidity: New Tasks, Priorities, and Frontiers for Integrated Gerontological and Clinical Research. *J Am Med Dir Assoc* **16**, 640–647 (2015).
6. Koch, L. Shared genetic components of multimorbidity. *Nat Rev Genet* **22**, 624–624 (2021).
7. Dong, G., Feng, J., Sun, F., Chen, J. & Zhao, X.-M. A global overview of genetically interpretable multimorbidities among common diseases in the UK Biobank. *Genome Med* **13**, 110 (2021).
8. Smith, S. M., Wallace, E., Clyne, B., Boland, F. & Fortin, M. Interventions for improving outcomes in patients with multimorbidity in primary care and community setting: a systematic review. *Syst Rev* **10**, 271 (2021).
9. Nicholson, K. *et al.* Multimorbidity and comorbidity revisited: refining the concepts for international health research. *J Clin Epidemiol* **105**, 142–146 (2019).
10. Jakovljević, M. & Ostojić, L. Comorbidity and multimorbidity in medicine today: challenges and opportunities for bringing separated branches of medicine closer to each other. *Psychiatr Danub* **25 Suppl 1**, 18–28 (2013).
11. Kuan, V. *et al.* Identifying and visualising multimorbidity and comorbidity patterns in patients in the English National Health Service: a population-based study. *Lancet Digit Health* **5**, e16–e27 (2023).
12. van Driel, M. A., Bruggeman, J., Vriend, G., Brunner, H. G. & Leunissen, J. A. M. A text-mining analysis of the human phenome. *European Journal of Human Genetics* **14**, 535–542 (2006).
13. Calvin, C. M., Conroy, M. C., Moore, S. F., Kuźma, E. & Littlejohns, T. J. Association of Multimorbidity, Disease Clusters, and Modification by Genetic Factors With Risk of Dementia. *JAMA Netw Open* **5**, e2232124 (2022).
14. Hanlon, P. *et al.* Associations between multimorbidity and adverse health outcomes in UK Biobank and the SAIL Databank: A comparison of longitudinal cohort studies. *PLoS Med* **19**, e1003931 (2022).
15. Hassaine, A. *et al.* Learning multimorbidity patterns from electronic health records using Non-negative Matrix Factorisation. *J Biomed Inform* **112**, 103606 (2020).
16. McQueenie, R. *et al.* Multimorbidity, polypharmacy, and COVID-19 infection within the UK Biobank cohort. *PLoS One* **15**, e0238091 (2020).
17. Strayer, N. *et al.* Interactive network-based clustering and investigation of multimorbidity association matrices with associationSubgraphs. *Bioinformatics* **39**, (2023).

18. Fotouhi, B., Momeni, N., Riolo, M. A. & Buckeridge, D. L. Statistical methods for constructing disease comorbidity networks from longitudinal inpatient data. *Appl Netw Sci* **3**, 46 (2018).
19. Aguado, A., Moratalla-Navarro, F., López-Simarro, F. & Moreno, V. MorbiNet: multimorbidity networks in adult general population. Analysis of type 2 diabetes mellitus comorbidity. *Sci Rep* **10**, 2416 (2020).
20. Ho, I. S. S. *et al.* Variation in the estimated prevalence of multimorbidity: Systematic review and meta-analysis of 193 international studies. *BMJ Open* vol. 12 Preprint at https://doi.org/10.1136/bmjopen-2021-057017 (2022).