

Preview

Data-driven precision medicine through the analysis of biological functional modules

Ilan Shomorony^{1,*}¹Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, USA*Correspondence: ilans@illinois.edu<https://doi.org/10.1016/j.xcrm.2022.100876>

Data-driven methods are expected to enable a next generation of personalized, preventative medicine. Zhang and colleagues¹ demonstrate how biological functional modules (BFMs) derived from the analysis of multimodal data can provide detailed quantitative health assessments and inform medical interventions.

Precision medicine is based on the idea that medical care should be tailored to each individual's health profile, which should take into account clinical information, genetics, lifestyle, and environment.² Such a holistic approach to medicine is expected to modernize our understanding of the mechanisms of disease and wellness and to pave the way for a new era of personalized, preventative health care and risk assessment.

Recent technological advances in genomics, the growing availability of health-care data, and the popularization of big data analytics have pushed this vision further, setting the stage for a data-driven precision medicine. But at the core of this new paradigm lies a basic data science task: how do we effectively integrate a large set of disparate data modalities in order to build robust and reliable statistical models for health assessment and risk prediction?

A promising methodology for the task of multimodal health data integration relies on techniques from network science to uncover relationships across different data modalities. This idea was originally proposed by Price et al. on a study with 108 subjects,³ and later further developed in a work by Shomorony et al. on a cohort of 1,253 individuals.⁴ These two studies considered a two-step approach to make sense of multimodal health data, illustrated in Figure 1. First, pairwise statistical tests are performed to identify cross-modality associations. Second, by viewing these associations as forming a cross-modality association network, one can employ community detection algorithms, such as Girvan-Newman,⁵ to extract groups of features that are

densely connected. These groups are termed biological functional modules (BFMs) and are expected to contain features that are implicated on a specific "axis" of human health. For example, the cardiometabolic health BFM may contain features from several modalities such as cholesterol levels, BMI, genetic markers associated with metabolic syndrome, and gut microbiome bacteria that act as probiotics.

The analysis of BFMs can reveal new biological pathways and biomarker signatures for health and disease. For example, the study by Shomorony et al.⁴ identified the metabolite cinnamoylglycine as a biomarker for gut microbiome health and lean mass percentage. From a data analysis perspective, the BFMs perform a kind of unsupervised dimensionality reduction and provide subsets of variables on which health assessment models can be subsequently trained. However, previous BFM studies were based on relatively small cohorts of mostly healthy individuals, limiting the medical potential of the downstream analysis.

A new study by Zhang et al.¹ aims to take the analysis of BFMs to the next level by converting BFM-based insights into medically actionable findings. This new study is the largest of its kind, comprising 4,277 individuals and 1,240 features across a wide range of data modalities, which include whole-genome sequencing, immune repertoire sequencing, metabolomic profiling, clinical labs, gut metagenomic sequencing, body composition analysis, and lifestyle questionnaires, among others. Following the general network analysis approach in Figure 1, several BFMs were extracted

and annotated based on known biomarkers.

With the goal of exploring the medical usefulness of BFMs, Zhang et al. proposed a new method to assess the health status of individuals in the context of each BFM. The method, called BFM-ash, computes a notion of distance between an individual's BFM feature vector and the corresponding population baseline BFM feature vector. This provides a per-BFM score for each individual and can be used to identify anomalous health profiles with respect to a particular BFM. BFM-ash is shown to identify anomalous BFMs in individuals with known conditions such as gastroenteritis and tuberculosis, and the identified anomalous BFMs correspond to axes of health that are known to be connected with the corresponding condition (e.g., a BFM related to gut microbiota is anomalous for gastroenteritis and a BFM related to platelets and lymphocyte counts is anomalous for tuberculosis). In several cases, BFMs (or sub-BFMs) are identified as anomalous even though the BFM's individual features themselves cannot be classified as anomalous. This supports the idea that BFMs have more predictive power than their individual features.

Zhang et al.¹ also explored the use of BFMs to perform a longitudinal dietary intervention study. The study was performed on individuals whose BFM-ash health assessment score was poor for several BFMs associated with gut microbiome health. Individuals in the case group were given grape seed extract as a dietary supplement, while the control group was given a placebo. The multimodal data from case and control groups



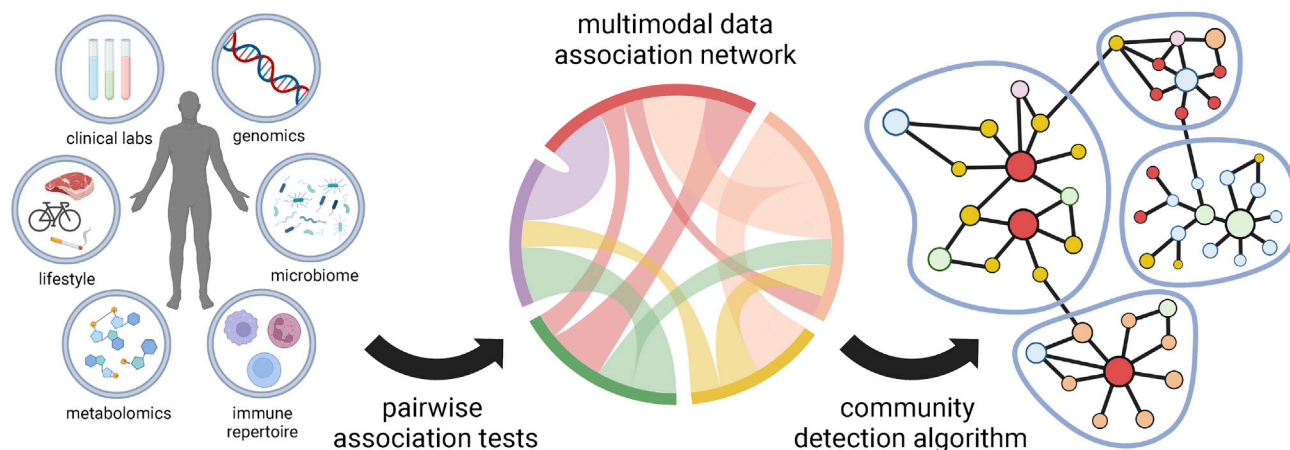


Figure 1. Extracting biological functional modules (BFMs) from multimodal data

Health data are collected across several modalities such as clinical labs, genomics, microbiome metagenomics, immune repertoire sequencing, metabolomic profiling, and lifestyle. Statistical association tests are computed for each pair of features in different modalities, forming an association network. Community detection algorithms are then applied to identify BFMs.

were collected at three different time points, separated by three-month intervals, and their health status was assessed for each time point using BFM-ash. Interestingly, the BFM-ash health status of individuals in the case group shows a clear improvement with all individuals moving closer to the population baseline, while the same effect is not observed in the control group. This demonstrates a concrete way in which a BFM-based health assessment can be used to stratify a population and to evaluate the efficacy of a medical intervention.

Overall, the work of Zhang et al. puts forward new promising ways to harness the power of multimodal health data. BFMs provide a lens to make sense of these rich and complex datasets and a framework to

perform health assessment and risk prediction. While several technical challenges still remain before we can use these techniques in medical practice, studies like the one by Zhang et al. represent meaningful steps toward the next generation of data-enabled medicine.

DECLARATION OF INTERESTS

The author declares no competing interests.

REFERENCES

- Zhang, W., Wan, Z., Li, X., Li, R., Luo, L., Song, Z., Miao, Y., Li, Z., Wang, S., Shan, Y., et al. (2022). A population-based study of precision health assessments using multi-omics network-derived biological functional modules. *Cell Rep. Med.* 3, 100847.

- Hodson, R. (2016 Sep). Precision medicine. *Nature* 537, S49. <https://doi.org/10.1038/537s49a>.
- Price, N.D., Magis, A.T., Earls, J.C., Glusman, G., Levy, R., Lausted, C., McDonald, D.T., Kusebauch, U., Moss, C.L., Zhou, Y., et al. (2017 Aug). A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat. Biotechnol.* 35, 747–756. <https://doi.org/10.1038/nbt.3870>.
- Shomorony, I., Cirulli, E.T., Huang, L., Napier, L.A., Heister, R.R., Hicks, M., Cohen, I.V., Yu, H.C., Swisher, C.L., Schenker-Ahmed, N.M., et al. (2020 Dec). An unsupervised learning approach to identify novel signatures of health and disease from multimodal data. *Genome Med.* 12, 7–4. <https://doi.org/10.1186/s13073-019-0705-z>.
- Girvan, M., and Newman, M.E.J. (2002 Jun 11). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* 99, 7821–7826. <https://doi.org/10.1073/pnas.122653799>.