Precision Medicine: Multi-modal and multi-scale methods to promote mechanistic understanding of disease

Hannah Carter
Department of Medicine, University of California San Diego, 9500 Gilman Dr. La Jolla, CA 92093, USA
Email: hkcarter@ucsd.edu

Steven Brenner
Department of Plant & Microbial Biology, University of California Berkeley, 461 Koshland Hall, Berkeley, California 94720-3102, USA
Email: brenner@compbio.berkeley.edu

Yana Bromberg
Department of Biology, Emory University, 1510 Clifton Rd, USA
Department of Computer Science, Emory University, 400 Dowman Dr. Atlanta, GA 30317, USA
Email: yana.bromberg@emory.edu

Precision medicine focuses on developing treatments and preventative strategies tailored to an individual's genomic profile, lifestyle, and environmental context. The Precision Medicine sessions at the Pacific Symposium on Biocomputing (PSB) have consistently spotlighted progress in this domain. Our 2025 manuscript collection features algorithmic innovations that integrate data across scales and diverse data modalities, presenting novel techniques to derive clinically relevant insights from molecular datasets. These studies highlight recent advances in technology and analytics and their application toward realizing the potential of precision medicine to enhance human health outcomes and extend lifespan.

Keywords: Precision medicine; variants; GWAS; genomics; proteomics; machine learning.

1. Introduction

Precision medicine leverages data-driven approaches to personalize medical decisions and treatments for individual patients. Novel technologies enabling rapid and systematic production of molecular measurements and imaging data have facilitated new mechanistic insights into complex disease processes, ultimately translating into better clinical decision-making. Breakthroughs in genomic technologies, such as spatial transcriptomics and single-cell analysis, have enabled the discovery of genetic biomarkers associated with drug responses, disease susceptibility, and other key medical outcomes. Concurrently, the vast scale of these data has spurred the development of

novel computational techniques, exemplified by the surge in deep learning-based approaches for biological and medical data analysis.

As the richness of datasets characterizing molecules, cells, and tissues grows, there are new opportunities to combine them across data modalities and scales. Methods to synthesize these data into mechanistic understanding and better biomarkers for clinically relevant outcomes are needed. In our 2025 session, we highlight ground-breaking research from a wide range of disciplines that integrate divergent data to offer novel insight into disease mechanisms, diagnosis, and treatment. These methods provide a preview of how computation can continue to advance precision medicine in the coming years.

2. Session Contributions

2.1. *Integrating imaging data with genotype to investigate mechanism*

Imaging data is increasingly available and can provide valuable information about how genetic factors are linked to disease mechanism through effects on the organization of biological systems. Two papers in this collection showcase the use of imaging to study genotypes. In one, Blennemann *et al* use live cell imaging to obtain longitudinal and spatially resolved information about T cell interaction with tumor cells across 3 genotypes[1]. In the other, Chandio and colleagues use diffusion MRI-based tractometry to obtain 3D quantitative measurements of white matter tracts across individuals with different risk genotypes and diagnoses, finding different structural features in carriers of risk versus protective genotypes[2].

2.2. *Biological understanding through linked diseases*

This collection also includes two studies where investigators use known disease associations as a starting point to identify genes and pathways underlying shared mechanisms. To gain insight into genes driving the link between Down Syndrome and obesity, Nandi *et al* derive latent variables from RNA sequencing analysis and analyze them in the context of karyotype and BMI using causal inference[3]. Ball and team use multi-disease modeling to analyze transcriptomic data from brain samples of mouse models of AD, T2D, both simultaneously and postmortem human brain to uncover a link between these diseases mediated by estrogen and inflammatory pathways[4].

2.3. *Finding drug targets and mediators of adverse drug responses*

Identification of disease-specific drug targets and understanding of the mechanisms mediating drug resistance and adverse responses are essential to inform drug development and clinical matching of patients to drugs. Three papers address these aspects of precision therapy. Orlenko *et al* implicate putative drug targets for Alzheimer’s disease by integrating information about interaction partners of known Alzheimer's genes with known gene-drug associations from a drug database[5]. By integrating methylation QTLs with GWAS loci associated with drug response phenotypes, Smith *et al* identify a number of candidate genes where genetics and epigenetics converge to generate adverse drug response[6]. Wen *et al* propose spherical PCA for single cell imaging data as a strategy to identify

cancer cells that evade cell cycle blocking drugs[7]. They find that cells that evade arrest after treatment express key genes that may represent additional drug targets.

2.4. *Using polygenic scores for molecular and intermediate phenotypes to uncover disease mechanisms*

Polygenic risk scores (PRS) have featured prominently in precision medicine research but have provided limited mechanistic understanding of disease due to their associative nature. Three papers in this session show that polygenic scores linked to protein levels or cellular activities, as opposed to disease incidence, can provide new biological insight. Phillips *et al* use genotype and RNA sequencing data to develop a polygenic score quantifying astrocyte activation, then use the PRS to study Alzheimer's-associated characteristics in a large cohort of elderly individuals[8]. They find that the score correlates with effects on memory and high-level cognition. By integrating information across genotype associations with protein versus gene expression levels, Moore *et al* reveal pathways underlying MRI-derived characteristics of heart function linked to heart failure[9]. Woerner *et al* show that PRS and polygenic predictors of protein levels can be combined to improve prediction of inflammatory bowel disease[10]. They find that polygenic protein scores are even more predictive when polygenic risk is high.

2.5. *Identifying environmental modifiers of traits and risk*

Environmental variables can modify traits and disease risk. Two papers in our session use computational strategies to better understand environmental factors. The study by Rico *et al* uses "environment by environment" associations and lipid measurements to explore interactions among environmental factors that affect lipid phenotypes[11]. They find several cases where combinations of two environmental exposures associate with significant differences in HDL levels. To study the effects of salt intake on risk of chronic kidney disease (CKD), Shivakumar *et al* use polygenic risk scores for CKD to stratify individuals in the UKBioBank, then examine the association between salt consumption and incidence of CKD in each subgroup[12].

2.6. *Methods addressing computational challenges for multimodal health data analysis*

More broadly, the availability of rich multi-modal and sensitive health data presents new computational challenges for advancing precision medicine. Two papers in this session propose solutions to specific challenges in this space. Colombo and team develop a strategy to predict cancer type from a low dimensional representation of 2 data types, SNVs and CNVs, while ensuring preservation of spatial relationships between genes in CNV regions[13]. They further demonstrate that they can operate on encrypted data to ensure patient privacy which remains a major concern for precision medicine approaches that require genomic data. Golovanevsky *et al* propose a one-versus-others attention approach to address computational bottlenecks in neural network-based integration across the rich array of data modalities available for clinical applications[14]. These advances address more general barriers to scaling and implementing computational approaches for precision medicine.

References

1. M. Blennemann, A. Verma, S. Bachl, J. Carnevale, B. E. Engelhardt, Understanding TCR T cell knockout behavior using interpretable machine learning *Pacific Symposium on Biocomputing* (2024)
2. B. Q. Chandio, J. E. Villalon-Reina, T. M. Nir, S. I. Thomopoulos, Y. Feng, S. Benavidez, N. Jahanshad, J. Harezlak, E. Garyfallidis, P. M. Thompson, Amyloid, Tau, and APOE in Alzheimer's Disease: Impact on White Matter Tracts *Pacific Symposium on Biocomputing* (2024)
3. S. Nandi, Y. Z., L. A Gillenwater, M. S. Ganes, H. Zhang, N. Janani, C. Greene, M. Pividori, M. Chikina, J. C. Costello, A Pathway-Level Information ExtractoR (PLIER) framework to gain mechanistic insights into obesity in Down syndrome *Pacific Symposium on Biocomputing* (2024)
4. B. K. Ball, E. A. Proctor, D. K. Brubaker, Cross-Species Modeling Identifies Gene Signatures in Type 2 Diabetes Mouse Models Predictive of Inflammatory and Estrogen Signaling Pathways Associated with Alzheimer's Disease Outcomes in Humans *Pacific Symposium on Biocomputing* (2024)
5. A. Orlenko, M. Venkatesan, L. Shen, M. D. Ritchie, Z. P. Wang, T. Obafemi-Ajayi, J. H. Moore, Biologically Enhanced Machine Learning Model to uncover Novel Gene-Drug Targets for Alzheimer's Disease *Pacific Symposium on Biocomputing* (2024)
6. D. A. Smith, S. A. Arteaga, M.C. Sadler, R. B. Altman, Identifying DNA methylation sites affecting drug response using electronic health record–derived GWAS summary statistics *Pacific Symposium on Biocomputing* (2024)
7. O. Wen, S. C. Wolff, W. Stallaert, D. Li, J. E. Purvis, T. M. Zikry, Spherical Manifolds Capture Drug-Induced Changes in Tumor Cell Cycle Behavior *Pacific Symposium on Biocomputing* (2024)
8. J. M Phillips, J. A. Schneider, D. A. Bennett, P. K. Crane, S. L. Risacher, A. J. Saykin, L. C. Dumitrescu, T. J. Hohman, Astrocyte Reactivity Polygenic Risk Score May Predict Cognitive Decline in Alzheimer's Disease *Pacific Symposium on Biocomputing* (2024)
9. A. Moore, R. Venkatesh, M. G. Levin, S. M. Damrauer, N. Reza, T. P. Cappola, M. D. Ritchie, Connecting intermediate phenotypes to disease using multi-omics in heart failure *Pacific Symposium on Biocomputing* (2024)
10. J. Woerner, T. Westbrook, S. Jeong, M. Shivakumar, A. R. Greenplate, S. A. Apostolidis, S. Lee, Y. Nam, D. Kim, Genomics and Computational Biology Graduate Group, University of Pennsylvania, "Plasma protein-based and polygenic risk scores serve complementary roles in predicting inflammatory bowel disease *Pacific Symposium on Biocomputing* (2024)
11. A. L. G. Rico and N. Palmiero, M. D. Ritchie, M. A. Hall, Integrated exposomic analysis of lipid phenotypes: leveraging GE.db in environment by environment association studies *Pacific Symposium on Biocomputing* (2024)
12. M. Shivakumar, Y. Kim, S. Jung, J. Woerner, D. Kim, Salt intake is a stronger predictor of chronic kidney disease in genetically susceptible population *Pacific Symposium on Biocomputing* (2024)

13. A. A. F. Colombo, L. Colombo, A. Falcetta, and M. Roveri, Enhancing Privacy-Preserving Cancer Classification with Convolutional Neural Networks *Pacific Symposium on Biocomputing* (2024)
14. M. Golovanevsky, E. Schiller, A. Nair, E. Han, R. Singh, C. Eikchoff, One-Versus-Others Attention: Scalable Multimodal Integration for Biomedical Data *Pacific Symposium on Biocomputing* (2024)