

Appendix: Summary of Best Papers Selected for the IMIA Yearbook 2023, Bioinformatics and Translational Informatics

Grazioli F, Siarheyev R, Alqassem I, Henschel A, Pileggi G, Meiser A.

Microbiome-based disease prediction with multimodal variational information bottlenecks

PLoS Comput Biol 2022 Apr 11;18(4):e1010050. doi: 10.1371/journal.pcbi.1010050

In this paper the authors addressed the untapped potential of multimodal machine learning in disease prediction by leveraging the diagnostic potential of gut microbial profiling. Traditionally, microbial species-relative abundances or strain-level markers extracted through shotgun metagenomic sequencing have been separately assessed in disease prediction models. Grazioli et al.'s innovative approach involved the development of a Multimodal Variational Information Bottleneck (MVIB), a deep learning model capable of integrating multiple heterogeneous data modalities into a single predictive framework. MVIB was devised to offer both efficient performance and interpretability. The model creates a joint stochastic encoding of different input data types, thereby integrating a plethora of disease-related markers. Through evaluating the model on 11 publicly available disease cohorts, the researchers achieved high classification performance, with areas under the ROC curve (AUCs) ranging from 0.80 to 0.95 for five cohorts, while maintaining medium performance for the remainder. The versatility of MVIB was demonstrated through cross-study generalization experiments, where training and testing were performed on different cohorts for the same disease. The results were comparable to a benchmark Random Forest model. Moreover, the scalability of MVIB was underscored by its ability to incorporate a third input modality, metabolomic data derived from mass spectrometry, without compromising efficiency or performance.

Kuppe C, Ramirez Flores RO, Li Z, Hayat S, Levinson RT, Liao X, Hannani MT, Tanevski J, Wünnemann F, Nagai JS, Halder M, Schumacher D, Menzel S, Schäfer G, Hoeft K, Cheng M, Ziegler S, Zhang X, Peisker F, Kaesler N, Saritas T, Xu Y, Kassner A, Gummert J, Morshuis M, Amrute J, Veltrop RJA, Boor P, Klingel K, Van Laake LW, Vink A, Hoogenboezem RM, Bindels EMJ, Schurgers L, Sattler S, Schapiro D, Schneider RK, Lavine K, Milting H, Costa IG, Saez-Rodriguez J, Kramann R

Spatial multi-omic map of human myocardial infarction

Nature 2022 Aug;608(7924):766-77. doi: 10.1038/s41586-022-05060-x

In this paper the authors leveraged single-cell -omics profiling to generate a temporal and spatial map of cardiac cell types in myocardial infarction patients and controls. Remodeling of cardiac tissues after myocardial infarction significantly contributes to late-stage mortality and is not well-addressed by current therapies. Limiting the negative impacts of cardiac remodeling on patients will require the development of new therapeutic approaches enabled by a more precise molecular understanding of the cell types involved in the process. The authors combined single-cell approaches, including single nucleus RNA sequencing and single nucleus assay for transposase-accessible chromatin sequencing, with spatial transcriptomics in 31 samples that spanned multiple clinical timepoints. From these data, they were able to identify major cell types in heart tissue and map these to particular histomorphological regions. Integrating these multi-omics data identified sets of differentially expressed genes in cells that marked the border between injured and uninjured tissues, and characterized the profiles of remodeled versus functional myocardium. The resulting multi-modal atlas of the human heart generates hypotheses that can facilitate new therapeutic advances and provides an important resource for the research community.

Weitz P, Wang Y, Kartasalo K, Egevad L, Lindberg J, Grönberg H, Eklund M,

Rantalainen M

Transcriptome-wide prediction of prostate cancer gene expression from histopathology images using co-expression-based convolutional neural networks.

Bioinformatics 2022 Jun 27;38(13):3462-9. doi: 10.1093/bioinformatics/btac343

The authors developed a novel machine learning to predict gene expression directly from haematoxylin and eosin-stained whole slide images (WSIs) in samples from patients with prostate cancer. Molecular phenotyping, especially using gene expression data, is an increasingly important approach to characterize patient samples, compute clinical scores from biomarkers, and implement precision care. However, assays to generate the required data to conduct molecular phenotyping and compute clinical scores are costly to implement on a large scale. Previous work has shown that molecular phenotypes, including gene expression, can be accurately predicted from histopathology WSIs and these WSIs are routinely collected and digitized during care. In their study, Weitz et al. trained a convolutional neural network (CNN) using prostate cancer samples from TCGA PRAD to predict gene expression levels for clusters of co-expressed genes. They then applied their predictions to compute a predicted cell cycle progression (CCP) score, which correlates with cancer aggressiveness, recurrence, and mortality. The authors found a significant correlation between predicted gene expression and transcript levels measured by RNA-sequencing in more than 6,600 genes. The significantly predicted genes were enriched in pathways relevant to prostate cancer, including those involved in DNA replication, cell cycle, and metabolism. Weitz et al. computed a CNN-based CCP score using these results. The CNN-based CCP scores were prognostic in their preliminary analysis, and were correlated with tumor grade to a similar degree as the RNA-seq-based CCP. Ultimately, this work suggests that deep learning models could provide a scalable solution to quantify gene expression phenotypes directly from imaging, particularly in settings where full molecular phenotyping would be otherwise unattainable.