Pacific Symposium on Biocomputing 26:166-171 (2021)

Computational Challenges and Artificial Intelligence in Precision Medicine

Olga Afanasiev[1], Joanne Berghout[2], Steven Brenner[3,4,5], Martha L. Bulyk[6,7], Dana C. Crawford[8,9], Jonathan H. Chen[10], Roxana Daneshjou[11], Łukasz Kidziński[12,*]

[1]*Sutter Health - Palo Alto Medical Foundation, Palo Alto, California*
[2]*Department of Medicine, University of Arizona Health Science, Tucson, Arizona; currently: Rare Disease Research Unit, Pfizer Inc., Cambridge, Massachusetts*
[3]*Department of Bioengineering, University of California, Berkeley, California*
[4]*Department of Molecular & Cell Biology, University of California, Berkeley, California*
[5]*Department of Plant & Microbial Biology, University of California, Berkeley, California*
[6]*Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts*
[7]*Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts*
[8]*Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio*
[9]*Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, Ohio*
[10]*Department of Biomedical Informatics, Stanford University, Stanford, California*
[11]*Department of Biomedical Data Science, Stanford University, Stanford, California*
[12]*Department of Bioengineering, Stanford University, Stanford, California*
[*]*Corresponding author, e-mail: lukasz.kidzinski@stanford.edu*

Continuously decreasing cost, speed and efficiency of DNA and RNA sequencing, coupled with advances in real-world sensing, storage of electronic health records, publicly available databases, and new data processing techniques enable precision medicine at unprecedented scale. Machine learning and artificial intelligence emerge naturally as tools for analyzing and summarizing data, supporting clinical decisions with data-driven insights and further unlocking genetically driven mechanisms underlying individualized risk. While these computational tools allow modeling of complex relations in large datasets, they pose new challenges especially because a patient's health is at stake. Due to an often black-box nature and high reliance on the training data, these new tools are prone to biases and most commonly provide correlational rather than causal insights. Results of these analyses have been difficult to validate, interpret, and explain to practitioners, and most genetic studies have struggled to encompass the full spectrum of human diversity. In this work, we summarize recent research trends in addressing these issues with examples from submissions to the "Computational Challenges and Artificial Intelligence in Precision Medicine" session at Pacific Symposium on Biocomputing 2021. We observe growing research interest in identifying biases, deriving causal and interpretable relations, tuning parameters of models for production, and using artificial intelligence for quality control. We expect further upsurge in work on interpretability and low-risk applications of advanced computational tools.

Keywords: artificial intelligence, augmented clinical decision making, bioinformatics, genomics, machine learning, personalized medicine, precision medicine, transcriptomics.

1. Introduction

High-volume genetic sequencing and 'omics data collection as well as increasingly accessible data streams from electronic health records (EHRs), clinical imaging, biobanks, wearables and more are opening up new vistas in biomedical and health data research. To integrate and/or identify meaningful insights from these large and typically noisy multi-dimensional data resources, the field has developed and applied novel computational tools, including many based on machine learning and artificial intelligence. Applied to genetics, these new methods have connected DNA variation to molecular functions and cellular perturbations, identified disease or patient subgroups and the biological processes driving these differences, suggested new therapeutic targets, and overall, dramatically increased our understanding of biomedicine. By integrating these datasets with rich clinical data, or developing algorithms to interpret, condense, or transform facets of these data into more interpretable modalities, much of the hidden information and patterns can be revealed and made useful for the practice of medicine.

2. Genomics and multi-omics data for precision medicine

An increasingly recognized problem for both health equity and methods development is the overrepresentation of European-ancestry participants in large-scale biobanks and omics resources[1]. Tools developed and trained on predominantly European-ancestry datasets have largely performed very poorly when generalized to more diverse populations, and this has serious bioethical, social, and scientific consequences that include missed insights, widened health disparities, and predictive inaccuracies[2,3]. In this PSB session, Singh et al. (2021) present a new method to analyze integrated DNA methylation, transcript expression, and sequence data in order to discover methylation-adjusted expression quantitative trait loci (eQTL) in cadaveric liver samples derived from donors of African American genetic ancestry[4]. Intersecting these data with cataloged genome-wide association study (GWAS) summary results presented several new genetic targets underlying GWAS loci in diseases that disproportionately impact African American populations. These targets had not been identified as candidates in previous work, and underscore the need for additional resources, methods development, and work in this area.

Along with better capturing the common variation present in humanity by expanding sample ascertainment to include historically excluded and currently underrepresented populations, one of the most compelling areas of research in precision medicine is to understand the functional impact of rare variation, and indeed, which rare variation is functional at all. Multiple tools have been generated to predict the functional consequences of protein coding variation, but only a few tools exist to analyze rare variation outside these coding regions. The problem is challenging due to incomplete annotation of functional regions and statistical limitations when considering ultra-rare variants that may appear uniquely within a dataset. Dong et al. (2021) have developed the AeQTL tool to identify rare heterogeneous variants that impact on levels of gene expression by aggregating rare variants according to user-specified regions and combining this genetic information with

patient-matched transcriptomic data[5]. They applied their methods to breast cancer sample data and were able to discover associations between aggregated rare germline variants in cis exomic regions with the expression of BRCA1 and SLC25A39.

Moving closer to the clinic, pharmacogenomics has enormous capacity for clinical actionability by bringing genotype-data driven guidance to the task of selecting an appropriate maintenance dose for individual patients. Rapidly determining the correct dose effectively balances the risk of side effects or other adverse outcomes against patient benefit. McInnes and Altman (2021) conducted linear modeling analyses across more than 200,000 participants in the UK Biobank to interrogate the real-world, observational pharmacy evidence that patient genotype at pre-specified loci may influence the maintenance dose of certain drugs prescribed in practice by clinicians[6]. A significant genotype-drug dose relationship was observed across (i) those drugs with Clinical Implementation of Pharmacogenomics Consortium (CIPC) guidance[7], (ii) drugs with a relationship described in DrugBank but no formal practice guideline, and (iii) a discovery set, where six out of 561 tested drugs showed pharmacogenomic potential. They further leveraged the longitudinal nature of UK Biobank to identify associations to side-effects including the appearance of new diagnoses. While the existence of a genotype-dose relationship is an unsurprising result, it previously had not been demonstrated in real prescribing patterns at this scale, and clearly demonstrates how incorporating patient genotype can be an important advance to patient safety.

The fourth paper submitted to this track by Aoki and Ester (2021) presented another new computational tool designed to improve causal inference[8]. Finding relationships between genes and outcomes can lead to better understanding of biological pathways and processes. And so, correlating outcomes and genes is a natural screening tool. However, in purely observational studies, and particularly those with thousands of potential variables, we risk identifying non-causal relations, which are of lower importance for biological discovery or intervention. One approach for narrowing down research targets is to focus on causal relations rather than correlations. Aoki and colleagues have proposed the ParKCa framework which leverages a stacking ensemble meta-learner approach to combine outcomes of multiple causal discovery methods, exploit partially known causes, and predict new ones. They confirmed the efficacy of their approach through simulations and by using their analysis over a real-world dataset[9] to identify cancer driver genes.

3. Artificial intelligence in multi-modal datasets for clinical research and workflows

Massive amounts of clinical data require new methods for quality control, analysis, processing, validation, and deployment of algorithms. Data-driven algorithms are particularly scrutinized due to their black-box nature, and high reliance on the training dataset, resulting in overfitting and biases towards certain populations. As opposed to classical hand-crafted algorithms for which the entire processing pipeline can be diligently monitored, biases and errors cannot be easily removed from data-driven algorithms due to complicated relationships between millions of parameters automatically derived from the data.

Opportunities for novel computational tools in precision medicine are particularly emphasized by the COVID-19 pandemic. COVID-19 disease, caused by a highly infectious SARS-CoV-2[10], can be associated with severe pneumonia resulting in serious complications or death; these poor

outcomes are more likely in patients with compromised immune systems due to other underlying conditions or age[11,12]. Rapid upsurge in the number of cases has exposed problems in healthcare systems across the globe. Moreover, restrictive measures for limiting the spread of the virus has led to the cancellation of face-to-face clinical visits for non-emergency visits. This situation has naturally resulted in the upsurge of telemedicine[13] and research on reading patient data automatically, with the intention of reducing the burden on clinicians. These developments among others are reflected in submissions and accepted papers to "Computational Challenges and Artificial Intelligence in Precision Medicine" session at Pacific Symposium on Biocomputing 2021.

Much of recent work in methods for clinical research has been focused on addressing these deployment issues, as it is exemplified by submissions and accepted papers to the "Computational tools and methods" track of this PSB 2021 session. First, for tuning models to certain populations[14], have analyzed optimization procedures for tuning model parameters. By analyzing a multi-study cohort of patients, they found that Bayesian optimization search is not more efficient than grid search and random sampling approaches despite prior evidence in literature based on single-study cohorts. Second, the bias towards the training set not only affects generalizability when we use it in different hospitals, but also results in varying behavior depending on demographics[15]. Third, data quality has fundamental importance not only for building medical machine learning tools, but also for clinical applications, particularly in telemedicine. Influx of telemedicine data due to COVID-19 motivated researchers to use deep learning for quality control of data[16].

3.1. *Optimization of genomic classifiers*

Machine learning and artificial intelligence allows researchers to identify relationships between patients' gene expression and their outcomes. These techniques can bring clinical benefits, but translation of research models into in-hospital deployment requires proper validation frameworks. While the research community focuses on accuracy metrics within a single cohort, practitioners often fail when attempting to deploy such models in practice.

Mayhew et al. (2021) addressed this problem by providing a framework for benchmarking solutions in the context of real-world deployment[14]. To that end, they built their models on a multi-study cohort of patients. They analyzed Bayesian optimization, grid search, and random search for hyperparameter optimization.

The authors illustrate an application of their framework on data on acute in-hospital infections with data coming from multiple studies. In contrast to previous research, they found that a Bayesian optimization framework was not more efficient and provided only marginal gains in performance of the final model. The study emphasized the need for deployment-centered benchmarking and validation on multi-study cohorts.

3.2. *Automatic reading of radiographic images*

Developments in computer vision, particularly in deep convolutional neural networks, have enabled a range of applications in medical imaging. Expert-level predictive models have been developed and published descriptions of algorithms capable of diagnosing skin cancer, brain cancer, lung lesions, or osteoarthritis progression from RGB camera photos, MRI sequences, X-ray, or CT scan input data.

Expert-level results can be achieved in a wide variety of use cases; however, applicability of these algorithms in practice is inhibited by any differences between the new real-world clinical data never seen by the model, and the datasets used for training. Additionally, even though demographics or ethnicity of patients are not explicitly expressed in radiographic images, there can be a bias in diagnostics, due to underrepresentation of certain groups or biased labels provided by clinicians.

In order to investigate the behavior of machine learning models as a function of demographics, ethnicity, or other patient data, one can look at a model's performance and analyze the True Positive Rate statistic of a model in different groups of interest. To that end, Sayyed-Kalantari et al. (2021) built a deep learning model for classifying chest X-rays, using multiple public chest X-ray datasets[15]. They trained a model with close to state-of-the-art performance and found that its accuracy depends on the patient's demographics, ethnicity, and insurance type. This discovery implied that validating the quality of the model across different populations should be one of the key quality checks for practitioners deploying a machine learning model in clinics. Without these kinds of quality checks, deep learning models may end up perpetuating biases rather than alleviating them.

3.3. *Quality control of images in telemedicine*

Other elements of clinical workflows can be addressed much more immediately than algorithmic imaging diagnostics, such as automating quality control. This is particularly important whenever images are collected by patients themselves (as in telemedicine) rather than by trained personnel under standardized hospital conditions. Systems for addressing quality control issues can bring immediate benefits to clinics and present low risk to patients while improving their care. Confident image assessment for clinicians, such as dermatologists, who are using telemedicine is challenging. Low quality images require extra time to read, causing additional delays related to retakes and extra reads. Vodrahalli et al. (2021) proposed a machine learning system for identifying low quality images automatically using a deep convolutional neural network classifier[16]. Their proof-of-concept algorithm could identify 50% of poor-quality images at a cost of only mislabeling 20% of the good quality images. Given a massive upsurge in telemedicine visits during the COVID-19 pandemic, this fraction could lead to significant time savings for hospitals and patients, as well as improved outcomes for time sensitive cases, such as malignant skin cancers. Moreover, these preliminary results could be further improved with better data and more thorough machine learning modelling.

4. Conclusion and future directions

Submissions to "Computational Challenges and Artificial Intelligence in Precision Medicine" session at Pacific Symposium on Biocomputing 2021 have revealed the growing importance and interest in decomposing components of black-box machine learning models, particularly for causality and for finding biases in data. Moreover, while automating the work of clinicians has always been a holy grail of artificial intelligence in medicine, papers in this session highlighted that there are more direct benefits of machine learning methods. Based on submissions to this session we expect further developments in interpretability of computational methods for precision medicine and more low-risk clinical applications, such as those motivated by COVID and post-COVID healthcare requirements.

5. Author contributions

All authors contributed equally to the PSB session. SEB was unable to review this manuscript for medical reasons.

References

1. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 1080 (2019).
2. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
3. Dias, R. & Torkamani, A. Artificial intelligence in clinical and genomic diagnostics. *Genome Med.* **11**, 70 (2019).
4. Singh, A., Zhong, Y., Nahlawi, L., Park, C.S., De, T., Alarcon, C., & Perera, M.A.. Incorporation of DNA methylation into eQTL mapping in African Americans. in *Pac Symp Biocomput* (2021).
5. Guanlan Dong, Michael C. Wendl, Bin Zhang, Li Ding, and Kuan-lin Huang. AeQTL: eQTL analysis using region-based aggregation of rare genomic variants. in *Pac Symp Biocomput* (2021).
6. McInnes, G., & Altman, R.B.. Drug Response Pharmacogenetics for 200,000 UK Biobank Participants. in *Pac Symp Biocomput* (2021).
7. Klein, M. E., Parvez, M. M. & Shin, J.-G. Clinical Implementation of Pharmacogenomics for Personalized Precision Medicine: Barriers and Solutions. *J. Pharm. Sci.* **106**, 2368–2379 (2017).
8. Aoki, R., & Ester, M. ParKCa: Causal Inference with Partially Known Causes. in *Pac Symp Biocomput* (2021).
9. Tomczak, K., Czerwińska, P. & Wiznerowicz, M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol.* **19**, A68–77 (2015).
10. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
11. Chau, A. S. *et al.* The Longitudinal Immune Response to Coronavirus Disease 2019: Chasing the Cytokine Storm. *Arthritis Rheumatol* (2020) doi:10.1002/art.41526.
12. Gupta, S. *et al.* Factors Associated With Death in Critically Ill Patients With Coronavirus Disease 2019 in the US. *JAMA Intern. Med.* (2020) doi:10.1001/jamainternmed.2020.3596.
13. Alexander, G. C. *et al.* Use and Content of Primary Care Office-Based vs Telemedicine Care Visits During the COVID-19 Pandemic in the US. *JAMA Netw Open* **3**, e2021476 (2020).
14. Michael B. Mayhew, Elizabeth Tran, Kirindi Choi, Uros Midic, Roland Luethy, Nandita Damaraju, and Ljubomir Buturovic. Optimization of Genomic Classifiers for Clinical Deployment: Evaluation of Bayesian Optimization to Select Predictive Models of Acute Infection and In-Hospital Mortality. in *Pac Symp Biocomput* (2021).
15. Laleh Seyyed-Kalantari, Guanxiong Liu, Matthew McDermott, Irene Y. Chen, and Maryzeh Ghassemi. CheXclusion: Fairness gaps in deep chest X-ray classifiers. in *Pac Symp Biocomput* (2021).
16. Kailas Vodrahalli, Roxana Daneshjou, and James Zou. TeleQC: A Machine Learning Algorithm to Improve the Quality of Telehealth Photos. in *Pac Symp Biocomput* (2021).