Pacific Symposium on Biocomputing 25:547-550(2020)

Precision Medicine: Addressing the Challenges of Sharing, Analysis, and Privacy at Scale

Steven E. Brenner,[1] Martha L. Bulyk,[2] Dana C. Crawford,[3] Alexander A. Morgan,[4]
Predrag Radivojac,[5] Nicholas P. Tatonetti[6]

[1] *University of California, Berkeley;* [2] *Brigham & Women's Hospital and Harvard Medical School;*
[3] *Case Western Reserve University;* [4] *Khosla Ventures;*
[5] *Northeastern University;* [6] *Columbia University*

Rapid technological advances, along with developments in medicine are opening new horizons in the field of personalized and precision medicine. Yet to bridge molecular measurement, clinical observation, social interaction and clinical practice, substantial gains must be made in methods for data integration, analysis, model development, and interpretation. These present daunting challenges of how to ethically share and analyze data, as we aim to ensure privacy, personal initiative, and mitigate disparities. Towards these goals, the Precision Medicine session at Pacific Symposium on Biocomputing is presenting fifteen papers covering a broad range of biomedical topics and collectively advancing state of the art in the field.

1. Introduction

Precision medicine encompasses an emerging medical practice concerned with maximizing the benefits of care based on an individual's genetic, molecular, environmental, and psychosocial factors.[1–7] Although a personalized approach to treatment has always been integral to medicine,[8] the recent advances in biotechnology, the rapid transition to electronic health records (EHR),[9,10] the sophistication of mobile health tracking devices, and increasingly pervasive Internet connections including social media, have dramatically changed the nature and increased the volume of health-related data, creating a vision of high-resolution clinical decision that enhance health.[11]

The affordability and broad availability of genetic information alongside potential of regularly interrogating molecular and lifestyle determinants at a relatively low cost has simultaneously raised expectations not only of rapidly achieving optimal personalized treatments, but also of exploring the utility of data in prevention and post-treatment. However, the path to realising the goals of precision medicine presents systemic obstacles that deserve national-scale focus and support, and requiring mathematical, computational, statistical, biological, legal, and ethical solutions.[11–13] First, a large number of constantly developing technologies coupled with partially observable molecular data (both static and dynamic), inherently complex EHR data, and incomplete lifestyle data place significant burdens on data integration strategies and all aspects of predictive modeling, ranging from advanced incorporation of domain knowledge in modeling to interpretation of outputs of machine learning tools. Second,

the scale and complexity of data further contribute to fundamental computational demands, such as the provenance, movement, sharing, and computation on big data.[14,15] Third, there are extant disparities in access to disease prevention/treatment as well as unequal trust in the healthcare system based on regional, ethnic and economic factors, collectively requiring customized solutions, outreach, and new regulations.[16–20]

Finally, and importantly, the same technologies and data that are enabling powerful precision modeling are threatening patient privacy and have the potential to erode the trust between scientist and patient.[21,22] Moreover, any new medical technology or intervention raises concerns of growing disparities; prior to its deployment, nobody had access to it, while subsequently only some will. In the case of many types of data, our ability to effectively interpret them are closely tied to the individual's ancestry and socioeconomic background. Some of the challenges of sharing models of internet browsing behavior while preserving privacy are being addressed in the computer science field. For example, differential privacy is a relatively mature model that allows for sharing models with provable levels of protections on individual privacy.[23] However, these techniques have yet to be meaningfully applied to most biomedical data. At least one reason is that biomedical data are much more complex and much more consequential than clicks on a website. Beyond this, differential privacy is primarily attractive for routine analyses, while our current state of knowledge means we currently need support for novel approaches. There is a need for innovative approaches that enable sharing while at the same time protect and empower patients and research participants.

As current discoveries and methodologies are absorbed in medical practice, new technologies, algorithmic and systems solutions will be proposed to improve the quality of care and explore the boundaries of automation. Pacific Symposium on Biocomputing (PSB), with its long successful tradition of advancing core computational areas of biomedicine, serves as an ideal forum for presenting, evaluating and auditing these solutions and understanding their value for precision medicine.

2. Session Papers

We solicited papers that analyze genomic sequence, genotypes, protein sequence, 'omics data, electronic health records, mobile health data, lifestyle measurements (e.g., via wearable technologies), and other individual or patient data as well as methods that empower individuals, safeguard privacy, and encourage sharing. We believe that the focus on a broad repertoire of topics has led to an overall high quality of submissions and accepted papers.

We received twenty submissions covering a wide range of topics in precision medicine, including algorithm development for and the analysis of molecular, genetic, clinical and lifestyle data. Nine papers were accepted for oral presentations and six for poster presentations. These papers appear in the PSB 2020 Proceedings and are listed below alphabetically.

- Arslanturk S, Draghici S, Nguyen T. "Integrated cancer subtyping using heterogeneous genome-scale molecular datasets."
- Bae H, Jung D, Choi HS, Yoon S. "AnomiGAN: Generative adversarial networks for anonymizing private medical data."
- Crawford DC, Lin J, Cooke Bailey JL, Kinzy T, Sedor JR, O'Toole JF, Bush WS.

"Frequency of ClinVar pathogenic variants in chronic kidney disease patients surveyed for return of research results at a Cleveland public hospital."

- Kong SW, Hernandez-Ferrer C. "Assessment of coverage for endogenous metabolites and exogenous chemical compounds using an untargeted metabolomics platform."
- Larson NB, Larson MC, Na J, Sosa C, Wang C, Kocher JP, Rowsey R. "Coverage profile correction of shallow-depth circulating cell-free DNA sequencing via multi-distance learning."
- Lever J, Barbarino JM, Gong L, Huddart R, Sangkuhl K, Whaley R, Whirl-Carrillo M, Woon M, Klein TE, Altman RB. "PGxMine: Text mining for curation of PharmGKB."
- Liu Q, Ha MJ, Bhattacharyya R, Garmire L, Baladandayuthapani V. "Network-based matching of patients and targeted therapies for precision oncology."
- Liu S, Hachen D, Lizardo O, Poellabauer C, Striegel A, MilenkoviÄ T. "The power of dynamic social networks to predict individualsâ mental health."
- Luthria G, Wang Q. "Implementing a cloud based method for protected clinical trial data sharing."
- Passero K, He X, Zhou J, Mueller-Myhsok B, Kleber ME, Maerz W, Hall MA. "Two phenome-wide association studies on cardiovascular health and fatty acids considering phenotype quality control practices for epidemiological data."
- Pershad Y, Guo M, Altman RB. "Pathway and network embedding methods for prioritizing psychiatric drugs."
- Pietras CM, Power L, Slonim DK. "aTEMPO: Pathway-specific temporal anomalies for precision therapeutics."
- Tong J, Duan R, Li R, Scheuemie MJ, Moore JH, Chen Y. "Learning from heterogeneous health systems without sharing patient-level data."
- Washington P, Paskov KM, Kalantarian H, Stockham N, Voss C, Kline A, Patnaik R, Chrisman B, Varma M, Tariq Q, Dunlap K, Schwartz J, Haber N, Wall DP. "Feature selection and dimension reduction of social autism data."
- Wolf JM, Barnard M, Xia X, Ryder N, Westra J, Tintle N. "Computationally efficient, exact covariate-adjusted multivariate methods for genetic analysis leveraging summary statistics from large biobanks."

Acknowledgements

National Institutes of Health award U41 HG007346 and Tata Consultancy Services.

References

1. G. H. Fernald, E. Capriotti, R. Daneshjou, K. J. Karczewski and R. B. Altman, Bioinformatics challenges for personalized medicine, *Bioinformatics* **27**, 1741 (2011).
2. E. D. Green, M. S. Guyer and National Human Genome Research Institute, Charting a course for genomic medicine from base pairs to bedside, *Nature* **470**, 204 (2011).
3. C. Carlsten, M. Brauer, F. Brinkman, J. Brook, D. Daley, K. McNagny, M. Pui, D. Royce, T. Takaro and J. Denburg, Genes, the environment and personalized medicine: we need to harness both environmental and genetic data to maximize personal and population health, *EMBO Rep* **15**, 736 (2014).

4. F. S. Collins and H. Varmus, A new initiative on precision medicine, *N Engl J Med* **372**, 793 (2015).
5. D. J. Duffy, Problems, challenges and promises: perspectives on precision medicine, *Brief Bioinform* **17**, 494 (2016).
6. B. Rost, P. Radivojac and Y. Bromberg, Protein function in precision medicine: deep understanding with machine learning, *FEBS Lett* **590**, 2327 (2016).
7. M. A. Hall, J. H. Moore and M. D. Ritchie, Embracing complex associations in common traits: critical considerations for precision medicine, *Trends Genet* **32**, 470 (2016).
8. J. F. Murray, Personalized medicine: been there, done that, always needs work!, *Am J Respir Crit Care Med* **185**, 1251 (2012).
9. W. G. Feero, A. E. Guttmacher and F. S. Collins, The genome gets personal–almost, *JAMA* **299**, 1351 (2008).
10. D. Blumenthal, Launching HITECH, *N Engl J Med* **362**, 382 (2010).
11. All of Us Research Program Investigators, J. C. Denny, J. L. Rutter, D. B. Goldstein, A. Philippakis, J. Smoller, G. Jenkins and E. Dishman, The "All of Us" research program, *N Engl J Med* **381**, 668 (2019).
12. I. S. Kohane, Ten things we have to do to achieve precision medicine, *Science* **349**, 37 (2015).
13. P. L. Sankar and L. S. Parker, The Precision Medicine Initiative's All of Us research program: an agenda for research on its ethical, legal, and social issues, *Genet Med* **19**, 743 (2017).
14. K. J. van Nimwegen, R. A. van Soest, J. A. Veltman, M. R. Nelen, G. J. van der Wilt, L. E. Vissers and J. P. Grutters, Is the $1000 genome as near as we think? A cost analysis of next-generation sequencing, *Clin Chem* **62**, 1458 (2016).
15. J. A. Lynch, B. Berse, W. D. Dotson, M. J. Khoury, N. Coomer and J. Kautter, Utilization of genetic tests: analysis of gene-specific billing in Medicare claims data, *Genet Med* **19**, 890 (2017).
16. C. D. Bustamante, E. G. Burchard and F. M. De la Vega, Genomics for the world, *Nature* **475**, 163 (2011).
17. A. B. Popejoy and S. M. Fullerton, Genomics is failing on diversity, *Nature* **538**, 161 (2016).
18. B. Thompson and J. Boiani, The legal environment for precision medicine, *Clin Pharmacol Ther* **99**, 167 (2016).
19. B. M. Hollister, N. A. Restrepo, E. Farber-Eger, D. C. Crawford, M. C. Aldrich and A. Non, Development and performance of text-mining algorithms to extract socioeconomic status from de-identified electronic health records, *Pac Symp Biocomput* **22**, 230 (2017).
20. G. Sirugo, S. Williams and S. Tishkoff, The missing diversity in human genetic studies, *Cell* **177**, p. 1080 (2019).
21. S. E. Brenner, Be prepared for the big genome leak, *Nature* **498**, p. 139 (2013).
22. S. Wang, X. Jiang, S. Singh, R. Marmor, L. Bonomi, D. Fox, M. Dow and L. Ohno-Machado, Genome privacy: challenges, technical approaches to mitigate risk, and ethical considerations in the United States, *Ann N Y Acad Sci* **1387**, 73 (2017).
23. C. Dwork and A. Roth, The algorithmic foundations of differential privacy, *Found Trends Theor Comput Sci* **9**, 211 (2014).