Precision medicine informatics

RECEIVED 7 March 2016
ACCEPTED 15 March 2016
PUBLISHED ONLINE FIRST 6 June 2016

Lewis J Frey,^{1,2} Elmer V Bernstam,^{3,4} and Joshua C Denny^{5,6}

This special issue on precision medicine informatics flowed from the AMIA 2015 Translational Bioinformatics Summit theme of “Accelerating Precision Medicine”¹ and President Obama’s 2015 State of the Union call “to give all of us access to the personalized information we need to keep ourselves and our families healthier.”² The goal is to focus on the inherent translational informatics challenges, concerns, and opportunities afforded by precision medicine to provide an accurate, personalized characterization of patient populations based on various characteristics including molecular (eg, genomic, proteomic), clinical (eg, comorbidities), environmental exposures, lifestyle, patient preferences, and other information.^{4–13} Informatics is a necessary component in a comprehensive initiative to tackle precision medicine. Its roles include (1) managing big data,^{15–17} (2) creating learning systems for knowledge generation,^{18–21} (3) providing access for individual involvement,^{22–25} and (4) ultimately supporting optimal delivery of precision treatments derived from translational research.^{26–29} The papers submitted for this special issue cover these 4 areas and expand on the themes of informatics emerging in the field of precision medicine.

BIG DATA

Data become “big” when the scale of the data and analyses challenge our traditional systems in size, velocity, or complexity.⁸ The large scale of the Million Veterans Program is an example, with systems and analysis methods being developed for application in a repository that has established policies and procedures. The Million Veterans Program can serve as a model for involving thousands of individuals in precision medicine research.³⁰ Similarly, the Precision Medicine Initiative Cohort Program seeks to connect electronic health records with participant-provided data, molecular determinants, environment, and lifestyle patterns to deeply impact our knowledge of health and therapies.³¹ The ambitious goal of enrolling 1 million volunteers whose demographics reflect the diversity of the US population can only be accomplished with robust and scalable informatics. Tenenbaum et al.^{3,32} present an informatics perspective and describe key informatics innovations required to advance precision medicine.

Big data can also be characterized by the velocity of data, such as in real-time data collection. We have come to expect nearly instant access and tracking of information that improves our quality of life.^{33–35} With precision medicine, the informatics research community is responding to the challenge and is providing applications relevant to the health of individuals. The Substitutable Medical Applications, Reusable Technologies (SMART) application by Warner et al.³⁶ provides the ability to visualize genomic information for oncology treatment using mobile technology. This can facilitate individualized communication about cancer treatment options in the context of genomic information. Personalizing treatment using the vast volume of available molecular data requires tools that reduce cognitive load.^{14,37} The tools described by Xu et al.³⁸ create a knowledge base of 2024 genomically informed clinical trials and treatments to support the delivery of personalized cancer therapy. Fathiamini et al.,³⁹ Hintzsche et al.,⁴⁰ and Cheng et al.⁴¹ describe resources that identify variants relevant to therapeutic treatments, variant calls, and natural language processing approaches to create structured information resources.

LEARNING SYSTEMS

Informatics has the potential to enable the collection, analysis, and reuse of data to facilitate learning knowledge representations for a wide range of diseases using phenotypes, genotypes, and predictive analytics.^{42–51} Halpern et al.⁵² present an approach to derive computable phenotypes using machine learning techniques that reduce the need to manually create phenotypes, which can be expensive and time-consuming. The use of machine learning techniques to establish disease-mutation relationships is described by Singhal et al.⁵³ Rieth et al.⁵⁴ describe automatic parsing and categorizing of molecular profiles gathered in routine care into a database that can inform research and clinical practice. Hoffman et al.⁵⁵ describe guidelines for incorporating pharmacogenetic tests into clinical decision support systems. The integration of epidemiological evidence into knowledge representations that can inform care is explored by Torosyan et al.⁵⁶ The integration of information from a wide variety of sources combined with learning systems promises to speed discovery.

With patients and physicians making decisions based on genetic tests that identify actionable risks, a patient’s individual attitudes can influence how the information is accepted and used. Strategies and frameworks that address engagement and disparities will need to be developed.^{57–62} Dye et al.⁶³ show that some minority populations are less likely to want genetic testing or participate in genetic research, which has the potential to exacerbate existing disparities. Precision medicine is focused on the individual patient. Given that molecular measurements are increasingly guiding care, studies will be needed to understand how to improve acceptance and participation by minorities. Adams and Petersen⁶⁴ discuss ethical, legal, and social issues that can arise when large numbers of individuals participate in genetic research. A framework for addressing ethical, legal, and social challenges can facilitate trust while protecting the individual and advancing research. Ni et al.⁶⁵ describe a learning framework to predict patient involvement in clinical trials with the goal of improving the effectiveness of recruitment.

PRECISION TREATMENT

The evaluation of methods to match genomics to therapeutics is an active area of research.^{66–74} Eubank et al.⁷⁵ developed an informatics solution to manage genetic/genomic and clinical data to expedite clinical trials of targeted cancer therapies at Memorial Sloan Kettering Cancer Center. As of August 2015, the system contained data on 159 893 patients, with 64 473 being tracked across 134 research cohorts with 51 192 patients in a genotype-matched eligible pool. The system can serve as a model for cohort programs in cancer clinical trials. Patients will want to know the outcomes for other patients in similar situations, and Warner et al.⁷⁶ provide a system that tracks clinical outcomes over time. Access to such information is vital to informed decision makers who are involved in keeping themselves and their families healthy. The direct relationship between clinical trials and knowledge representations is demonstrated through a knowledge base for proper cancer drug selection and a The Cancer Genome Atlas (TCGA) analysis of triple-negative breast cancer, where 71.7% of 85 cancer patients had Food and Drug Administration (FDA)-approved “drug-able” genomic targets.⁷⁷

Informatics tools will be required for precision medicine to both catalyze research on huge populations and enable implementation of precision care now done for isolated cases in routine care. With the powerful response of the translational informatics community to the request for papers for this precision medicine special focus issue, we can chart the main themes of the course ahead. The initial response is the beginning of a transition in health care that is supported through informatics to propel participating individuals into the center of research and care. The needs of such an interactive model of care are broad and will need to be further defined as the initial systems are tested and measure how, when, and where they improve the health and outcomes of individuals and families.

FUNDING

The work of L.J.F. was supported in part by National Institutes of Health (NIH) grants 1R01GM108346-01 and U54-GM104941, Health Equity and Rural Outreach Innovation Center grant CIN 13-418, and funding from the Hollings Cancer Center's support grant P30 CA138313 at the Medical University of South Carolina. The work of E.V.B. was supported by National Cancer Institute (NCI) U01 CA180964, the Sheikh Bin Zayed Al Nahyan Foundation, the Cancer Prevention Research Institute of Texas Precision Oncology Decision Support Core RP150535, National Center for Advancing Translational Sciences (NCATS) grant UL1 TR000371 (Center for Clinical and Translational Sciences), the Bosarge Foundation, and an MD Anderson Cancer Center support grant (NCI P30 CA016672). The work of J.C.D. was supported by R01 LM 010685, R01 GM 103859, and R01 GM 105688 from the NIH.

REFERENCES

- Translational Bioinformatics Summit 2015, American Medical Informatics Association [Internet]. <https://www.amia.org/jointsummits>. Accessed March 5, 2016.
- Remarks by the president in his State of the Union address, January 20, 2015 [Internet]. <https://www.whitehouse.gov/the-press-office/2015/01/20/remarks-president-state-union-address-january-20-2015>. Accessed March 5, 2016.
- AMIA Genomics and Translational Bioinformatics Working Group (Gen-TBI, formerly the Genomics Working Group) [Internet]. <https://www.amia.org/programs/working-groups/genomics-and-translational-bioinformatics>. Accessed March 5, 2016.
- Stead WW, Kelly BJ, Kolodner RM. Achievable steps toward building a National Health Information Infrastructure in the United States. *J Am Med Inform Assoc*. 2005;12(2):113–120.
- Collins FS. The case for a US prospective cohort study of genes and environment. *Nature*. 2004;429(6990):475–477.
- National Research Council (US) Committee on a Framework for Developing a New Taxonomy of Disease. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease* [Internet]. Washington, DC: National Academies Press; 2011. <http://www.ncbi.nlm.nih.gov/books/NBK91503/>. Accessed March 5, 2016.
- Roychowdhury S, Chinnaiyan AM. Translating genomics for precision cancer medicine. *Annu Rev Genomics Hum Genet*. 2014;15:395–415.
- Frey LJ, Lenert L, Lopez-Campos G. EHR big data deep phenotyping. *IMIA Yearbook*. 2014;9(1):206–211.
- Robinson PN. Deep phenotyping for precision medicine. *Hum Mutat*. 2012;33(5):777–780.
- Hripscak G, Albers DJ. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc*. 2013;20(1):117–121.
- Richesson RL, Hammond WE, Nahm M, et al. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. *J Am Med Inform Assoc*. 2013;20(e2):e226–e231.
- Ritchie MD, Denny JC, Crawford DC, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet*. 2010;86(4):560–572.
- Newton KM, Peissig PL, Kho AN, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *J Am Med Inform Assoc*. 2013;20(e1):e147–e154.
- Stead WW, Starmer JM. Beyond expert based practice. In McClellan M, et al., eds. *Evidence-Based Medicine and the Changing Nature of Healthcare: 2007 IOM Annual Meeting Summary*. Washington, DC: National Academies Press; 2008. Institute of Medicine (US). <http://www.ncbi.nlm.nih.gov/books/NBK52820/>. Accessed March 5, 2016.
- PMI Working Group. *Precision Medicine Initiative—National Institutes of Health (NIH)* [Internet]. <http://www.nih.gov/precisionmedicine/working-group.htm>. Accessed March 5, 2016.
- Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61–70.
- UK Biobank. *UK Biobank: Protocol for a Large-scale Prospective Epidemiological Resource* [Internet]. <http://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf>. Accessed March 5, 2016.
- Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Transl Med*. 2010;2(57):57cm29.
- Kohane IS. Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet*. 2011;12(6):417–428.
- Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc*. 2010;17(2):124–130.
- Frey LJ, Sward KA, Newth C, et al. Virtualization of open-source secure web services to support data exchange in a pediatric critical care research network. *J Am Med Inform Assoc*. 2015;22(6):1271–1276. <http://jamia.oxfordjournals.org/content/22/6/1271.abstract>. Accessed March 5, 2016.
- The White House. *Precision Medicine Initiative: Proposed Privacy and Trust Principles* [Internet]. https://www.whitehouse.gov/sites/default/files/docs/pmi_privacy_and_trust_principles_july_2015.pdf. Accessed March 5, 2016.
- Wallerstein N, Duran B. Community-based participatory research contributions to intervention research: the intersection of science and practice to improve health equity. *Am J Public Health*. 2010;100 (Suppl 1):S40–S46.
- Jagosh J, Macaulay AC, Pluye P, et al. Uncovering the benefits of participatory research: implications of a realist review for health research and practice. *Milbank Q*. 2012;90(2):311–346.
- OHRP. *Federal Policy for the Protection of Human Subjects ("Common Rule")* [Internet]. <http://www.hhs.gov/ohrp/humansubjects/commonrule/>. Accessed March 5, 2016.
- Ashley EA, Butte AJ, Wheeler MT, et al. Clinical assessment incorporating a personal genome. *Lancet*. 2010;375(9725):1525–1535.
- Chapman PB, Hauschild A, Robert C, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med*. 2011;364(26):2507–2516.
- Ellis PM, Coakley N, Feld R, et al. Use of the epidermal growth factor receptor inhibitors gefitinib, erlotinib, afatinib, dacomitinib, and icotinib in the treatment of non-small-cell lung cancer: a systematic review. *Curr Oncol Tor Ont*. 2015;22(3):e183–e215.
- Starren J, Williams MS, Bottinger EP. Crossing the omic chasm: a time for omic ancillary systems. *JAMA*. 2013;309(12):1237–1238.
- Gaziano JM, Concato J, Brophy M, et al. Million Veterans Program: a mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016;70:214–223.
- PMI Working Group—Precision Medicine Initiative—National Institutes of Health (NIH). *The Precision Medicine Initiative Cohort Program—Building a Research Foundation for 21st Century Medicine* [Internet]. <http://www.nih.gov/sites/default/files/research-training/initiatives/pmi/pmi-working-group-report-20150917-2.pdf>. Accessed March 5, 2016.
- Tenenbaum J, Avillach P, Benham-Hutchins M, et al. An informatics research agenda to support precision medicine: 7 key areas. *J Am Med Inform Assoc*. 2016;23(4):791–795.
- Duggan M. Cell Phone Activities 2013 [Internet]. *Pew Res Cent Internet Sci Tech*. <http://www.pewinternet.org/2013/09/19/cell-phone-activities-2013/>. Accessed March 5, 2016.
- Smith A. U.S. Smartphone Use in 2015 [Internet]. *Pew Res Cent Internet Sci Tech*. <http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/>. Accessed March 5, 2016.
- Rappaport SM. Implications of the exposome for exposure science. *J Expo Sci Environ Epidemiol*. 2011;21(1):5–9.
- Warner JL, Rieth MJ, Mandl KD, et al. SMART precision cancer medicine: a FHIR-based app to provide genomic information at the point of care. *J Am Med Inform Assoc*. 2016;23(4):701–710.

37. Bainbridge MN, Wiszniewski W, Murdock DR, et al. Whole-genome sequencing for optimized patient management. *Sci Transl Med*. 2011;3(87):1–6.
38. Xu J, Lee HJ, Zeng J, et al. Extracting genetic alteration information for personalized cancer therapy from ClinicalTrials.gov. *J Am Med Inform Assoc*. 2016;23(4):750–757.
39. Fathiadini S, Johnson AM, Zeng J, et al. Automated identification of molecular effects of drugs (AIMED). *J Am Med Inform Assoc*. 2016;23(4):758–765.
40. Hintzsche J, Kim J, Yadav V, et al. IMPACT: Whole-exome sequencing analysis pipeline of integrating molecular profiles with actionable therapeutics in clinical samples. *J Am Med Inform Assoc*. 2016;23(4):721–730.
41. Cheng F, Zhao J, Fooksa M, Zhao Z. A network-based drug repositioning infrastructure for precision cancer medicine through targeting significantly mutated genes in human cancer genomes. *J Am Med Inform Assoc*. 2016;23(4):681–691.
42. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):1–10.
43. Ahmad T, Pencina MJ, Schulte PJ, et al. Clinical implications of chronic heart failure phenotypes defined by cluster analysis. *J Am Coll Cardiol*. 2014;64(17):1765–1774.
44. Doshi-Velez F, Ge Y, Kohane I. Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis. *Pediatrics*. 2013;132(3):e2013–0819.
45. Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH. Sequence variations in PCSK9, Low LDL, and protection against coronary heart disease. *N Engl J Med*. 2006;354(12):1264–1272.
46. Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013;31(12):1102–1111.
47. Kho AN, Hayes MG, Rasmussen-Torvik L, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *J Am Med Inform Assoc*. 2012;19(2):212–218.
48. Kho AN, Pacheco JA, Peissig PL, et al. Electronic medical records for genetic research: results of the eMERGE Consortium. *Sci Transl Med*. 2011;3(79):79re1.
49. Crawford DC, Crosslin DR, Tromp G, et al. eMERGEing progress in genomics—the first seven years. *Front Genet*. 2014;5:184.
50. Collins FS, Hudson KL, Briggs JP, Lauer MS. PCORnet: turning a dream into reality. *J Am Med Inform Assoc*. 2014;21(4):576–577.
51. Pulley JM, Denny JC, Peterson JF, et al. Operational implementation of prospective genotyping for personalized medicine: The design of the Vanderbilt PREDICT project. *Clin Pharmacol Ther*. 2012;92(1):87–95.
52. Halpern Y, Horng S, Choi Y, Sontag D. Electronic medical record phenotyping using the anchor & learn framework. *J Am Med Inform Assoc*. 2016;23(e1):e20–e27.
53. Singhal A, Simmons M, Lu Z, et al. Text mining for precision medicine: automating disease mutation relationship extraction from biomedical literature. *J Am Med Inform Assoc*. 2016;23(4):766–772.
54. Rieth MJ, Thota R, Staggs D, et al. Pragmatic precision oncology: the secondary uses of clinical tumor molecular profiling. *J Am Med Inform Assoc*. 2016;23(4):773–776.
55. Hoffman JM, Dunnenberger HM, Hicks JK, et al. Developing knowledge resources to support precision medicine: principles from the clinical Pharmacogenetics Implementation Consortium. *J Am Med Inform Assoc*. 2016;23(4):796–801.
56. Torosyan Y, Hu Y, Hoffman S, et al. An in silico framework for integrating epidemiologic and genetic evidence with healthcare applications: ventilation-related pneumothorax as a case illustration. *J Am Med Inform Assoc*. 2016;23(4):711–720.
57. NIH. Request for Information: NIH Precision Medicine Cohort—Strategies to Address Community Engagement and Health Disparities. 2015. <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-107.html>. Accessed March 5, 2016.
58. Braveman PA, Kumanyika S, Fielding J, et al. Health disparities and health equity: the issue is justice. *Am J Public Health*. 2011;101(S1):S149–S155.
59. Kaufman D, Murphy J, Scott J, Hudson K. Subjects matter: a survey of public opinions about a large genetic cohort study. *Genet Med*. 2008;10(11):831–839.
60. Murphy J, Scott J, Kaufman D, et al. Public expectations for return of results from large-cohort genetic research. *Am J Bioeth AJOB*. 2008;8(11):36–43.
61. Trinidad SB, Fullerton SM, Ludman EJ, et al. Research practice and participant preferences: the growing gulf. *Science*. 2011;331(6015):287–288.
62. Beskow LM, Linney KN, Radtke RA, et al. Ethical challenges in genotype-driven research recruitment. *Genome Res*. 2010;20(6):705–709.
63. Dye, et al. Sociocultural variation in attitudes toward use of genetic information and participation in genetic research by race in the United States: implications for precision medicine. *J Am Med Inform Assoc*. 2016;23(4):782–786.
64. Adams SA, Petersen C. Precision medicine: opportunities, possibilities, and challenges for patients and providers. *J Am Med Inform Assoc*. 2016;23(4):787–790.
65. Ni Y, Beck AF, Taylor R, et al. Will they participate? Predicting patients' response to clinical trial invitations in a pediatric emergency department. *J Am Med Inform Assoc*. 2016;23(4):671–670.
66. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov*. 2013;12(8):581–594.
67. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014;506(7488):376–381.
68. Sanseau P, Agarwal P, Barnes MR, et al. Use of genome-wide association studies for drug repositioning. *Nat Biotechnol*. 2012;30(4):317–320.
69. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47(8):856–860.
70. DePeralta DK, Boland GM. Melanoma: advances in targeted therapy and molecular markers. *Ann Surg Oncol*. 2015;22(11):3451–3458.
71. Relling MV, Gardner EE, Sandborn WJ, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing. *Clin Pharmacol Ther*. 2011;89(3):387–391.
72. Shuldiner AR, O'Connell JR, Bliden KP, et al. Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA*. 2009;302(8):849–857.
73. The International Warfarin Pharmacogenetics Consortium. Estimation of the Warfarin Dose with Clinical and Pharmacogenetic Data. *N Engl J Med*. 2009;360(8):753–764.
74. Kimmel SE, French B, Kasner SE, et al. A pharmacogenetic versus a clinical algorithm for warfarin dosing. *N Engl J Med*. 2013;369(24):2283–2293.
75. Eubank MH, Hyman DM, Kanakamedala AD, et al. Automated eligibility screening and monitoring for genotype driven precision oncology trials. *J Am Med Inform Assoc*. 2016;23(4):777–781.
76. Warner JL, Wang L, Pao W, et al. CUSTOM-SEQ: A Prototype for Oncology Rapid Learning in a Comprehensive EHR Environment. *J Am Med Inform Assoc*. 2016;23(4):692–700.
77. Cheng L, Schneider B, Li L. A bioinformatics approach for the precision medicine off-label drug selection among triple negative breast cancer patients. *J Am Med Inform Assoc*. 2016;23(4):741–749.

AUTHOR AFFILIATIONS

¹Biomedical Informatics Center, Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

²Health Equity and Rural Outreach Innovation Center, Ralph H. Johnson Department of Veterans Affairs Medical Center, Charleston, SC, USA

³School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, USA

⁴Division of General Internal Medicine, Department of Internal Medicine, Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA

⁵Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA

⁶Department of Medicine, Vanderbilt University, Nashville, TN, USA