

## REGULATORY SCIENCE

# A research roadmap for next-generation sequencing informatics

Russ B. Altman,<sup>1\*</sup> Snehit Prabhu,<sup>2</sup> Arend Sidow,<sup>3</sup> Justin M. Zook,<sup>4,11</sup> Rachel Goldfeder,<sup>5</sup> David Litwack,<sup>6</sup> Euan Ashley,<sup>7</sup> George Asimenos,<sup>8</sup> Carlos D. Bustamante,<sup>2</sup> Katherine Donigan,<sup>6</sup> Kathleen M. Giacomini,<sup>9</sup> Elaine Johansen,<sup>6</sup> Natalia Khuri,<sup>10</sup> Eunice Lee,<sup>6</sup> Xueying Sharon Liang,<sup>6</sup> Marc Salit,<sup>4,10,11</sup> Omar Serang,<sup>8</sup> Zivana Tezak,<sup>6</sup> Dennis P. Wall,<sup>12</sup> Elizabeth Mansfield,<sup>6</sup> Taha Kass-Hout<sup>6</sup>

Next-generation sequencing technologies are fueling a wave of new diagnostic tests. Progress on a key set of nine research challenge areas will help generate the knowledge required to advance effectively these diagnostics to the clinic.

The Precision Medicine Initiative (PMI) is a U.S. national effort “to enable a new era of medicine through research, technology, and policies that empower patients, researchers, and providers to work together toward development of individualized care” (1). One goal is to bring about the routine use of next-generation precision diagnostics to benefit individuals and public health. Central to the introduction of safe and effective new precision diagnostic technologies is an adequate understanding of how well they perform. Through the PMI, the U.S. Food and Drug Administration (FDA) is seeking to address this issue by providing dynamic, flexible, and well-balanced regulation of precision diagnostics. Because these complex technologies pose new challenges in understanding their likely benefits and their limits in terms of accuracy, precision, and clinical validity, FDA is advancing a robust research agenda in regulatory science. New knowledge gained from this agenda will inform the next generation of regulation for precision medicine.

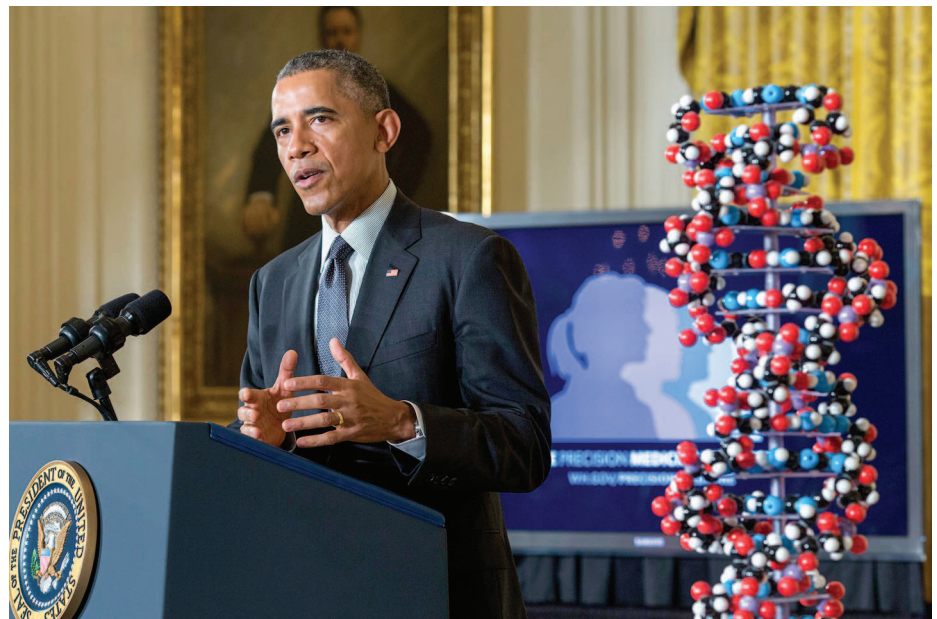
The UCSF-Stanford Center for Excellence in Regulatory Science & Innovation (CERSI) hosted a series of meetings in September 2015

that included a public workshop and discussions on identifying key activities needed to evaluate the clinical implications of next-generation nucleic acid sequencing (NGS). Here we summarize the ideas and directions that were proposed and put forth a working “roadmap”

Eventually, these developments are likely to culminate in the routine sequencing of patients’ genomes. In the meantime, there will be several years during which the process of DNA sequence determination remains challenging and in which cost-, quality-, and goal-driven trade-offs result in a large diversity of testing strategies. In this Perspective, we lay out the technological challenges that are slowing the routine clinical use of a new generation of genetic tests and propose questions that regulatory science should address to arrive at a flexible yet robust regulatory framework that results in maximum benefit for patients.

As part of its PMI effort, FDA seeks to undertake and support regulatory science research that will enhance our understanding of NGS test products and their development and validation, as well as how the results of such tests are best communicated in an evolving health care environment.

A centerpiece of this effort is precisionFDA, a research and development portal that



Sharing the stage with a DNA double helix, U.S. President Barack Obama discusses the Precision Medicine Initiative.

for NGS evaluation, as a possible exemplar of how many other new next-generation diagnostics may be understood.

## DEFINING THE TASK

Technological breakthroughs have recently led to DNA sequencing methods that can generate the raw data necessary for determining nearly the entire genome sequence of any individual.

will allow community members to better understand, develop, and improve existing and new bioinformatics approaches for processing the vast amount of genomic data that is collected using NGS technology. precisionFDA is a public, cloud-based platform developed by FDA and its contractor DNAnexus that hosts shared tools, crowdsourced testing, and community challenges, to improve and share

<sup>1</sup>Bioengineering, Genetics, and Medicine, Stanford University, Stanford, CA 94305, USA. <sup>2</sup>Biomedical Data Science and Genetics, Stanford University, Stanford, CA 94305, USA. <sup>3</sup>Pathology and Genetics, Stanford University, Stanford, CA 94305, USA. <sup>4</sup>Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA. <sup>5</sup>Biomedical Informatics, Stanford University, Stanford, CA 94305, USA. <sup>6</sup>Food and Drug Administration, Silver Spring, MD 20993, USA. <sup>7</sup>Medicine, Genetics, and Pathology, Stanford University, Stanford, CA 94305, USA. <sup>8</sup>DNAnexus, Mountain View, CA 94040, USA. <sup>9</sup>Bioengineering and Therapeutic Sciences, University of California at San Francisco, San Francisco, CA 94143, USA. <sup>10</sup>Bioengineering, Stanford University, Stanford, CA 94305, USA. <sup>11</sup>Material Measurement Laboratory, National Institute of Standards and Technology, Stanford University, Stanford, CA 94305, USA. <sup>12</sup>Systems Medicine and Psychiatry, Stanford University, Stanford, CA 94305, USA.

\*Corresponding author. E-mail: russ.altman@stanford.edu

knowledge and methods for evaluating NGS bioinformatics pipelines.

This is currently a preregulatory platform that can host research-grade software. With time, we expect best practices to emerge for the evaluation and use of NGS pipelines that may allow NGS test developers and FDA to rely on precisionFDA-based analyses to build standards and communicate the technical performance of NGS tests.

In addition, FDA seeks to answer practical regulatory science questions, such as which reference sequences and data sets will be optimal for supporting development and validation of NGS bioinformatics tools, and how providers and patients want to receive genetic information.

**INTERROGATED REGIONS, DETECTABLE VARIATION, AND INTENDED USE**

NGS is similar to traditional DNA-based genetic tests, in that it begins with specimen collection and DNA extraction and requires interpretation of the detected genetic variants, and variant findings are reported as test results to clinicians and patients. However, NGS differs from traditional genetic tests in many ways, including its ability to assess large segments of the genome and, perhaps more importantly, its ability to detect variants in an untargeted way. These differences pose singular challenges to evaluation of the quality of an NGS test, which this Perspective seeks to highlight.

For the purposes of this discussion, we will refer to the sequence or sequences of the ge-

nome that are being interrogated by an NGS test as the **interrogated region**, which includes the DNA segment(s) that are intended to be measured and whether the intent of the test is to measure a particular base in the genome, an entire gene, a locus, a chromosome, an exome, or a complete genome. We will refer to the types of potential variants of interest in this region as the **detectable variants**. This could include single-nucleotide polymorphisms (SNPs), insertions/deletions (indels), and copy-number or other types of variants. The **expected use** defines the anticipated clinical (or research) use of genomic information and will factor in minimal performance characteristics that make the test useful to researchers, providers, or patients. Both the interrogated region and detectable variants depend critically on the expected use of the test results. A range of expected uses envisioned today are shown in Table 1.

In the context of different expected uses, bioinformatic pipelines might be distinct in design and heterogeneous in implementation. This, in turn, suggests that performance characteristics and evaluation metrics for different applications of NGS-based tests would benefit from use-specific development and should be evaluated in accordance with their individual benefits and potential for harm. The evaluation of pipelines and their differences, as well as the assessment of how the pipelines affect test performance, is the vision behind the precisionFDA infrastructure: To accommodate a diverse set of use-cases for NGS testing and to measure performance levels and

understand tradeoffs as appropriate for each application.

**REGULATORY SCIENCE ROADMAP**

In order to organize precisionFDA as a community platform and to ensure an effective means of developing robust methods for evaluating NGS-based diagnostics, we propose a research roadmap that highlights nine areas of potential regulatory science investigation for FDA (including precisionFDA) accompanied by a discussion of aims, research milestones, software tools, and data services. This roadmap focuses on questions that we believe should be asked, not on the exact research methodology that should be employed. Methodological discussions are left to individual stakeholders pursuing these questions and should (as for all research) be nimble and responsive to the reality of rapidly changing knowledge and technological advances in the field.

**(i) Address secure storage, sharing, and maintenance of genomic data and software tools for regulatory science and research.** PrecisionFDA must be able to accept, store, and manage data from authorized users. Ideally, it will include the ability to create, test, and promote adoption of methods for storing and accessing large numbers of genomes. Therefore, it is important to create methods to ensure security and confidentiality of information and access controls tied to specifics of the informed consent obtained from those who provide samples. Of course, these methods should respect intellectual property and ownership. When sharing genomes, technical challenges include (but are not limited to) the creation of unique genome identifiers, methods for searching genomes based on specific features (for example, disease status), and adopting and promulgating standard formats of sequence representation and variant calls.

In a similar vein, software developed by the precisionFDA community might be for private use, for group sharing, or open to the public. Therefore, precisionFDA needs a set of rules for access to software and mechanisms for enforcing them. Robust systems for version control of software and data are also needed, so that experimental results can be effectively tracked and audited. Conditions under which software updates should trigger reevaluation of an entire pipeline need to be identified, to ensure continued integrity of analyses performed. The precisionFDA team is now building the first generation of these capabilities.

**(ii) Create reference data sets based on diverse patterns of expected use.** In order to

**Table 1. Range of currently envisioned uses of NGS.** “Expected use” has some similarities to the FDA regulatory term “intended use,” but because there are differences in the details, we simplify here to “expected use.”

Test	Expected use
Noninvasive fetal testing	To report chromosomal abnormalities known to be associated with harmful or lethal phenotypes
Targeted gene panels to test a patient’s tumor	To report any somatic mutations in known cancer driver genes
Carrier tests for individuals of reproductive age	To report whether an individual carries known or new pathogenic mutations in known heritable disease genes
Whole-genome sequencing tests for children with congenital diseases or conditions	To aid in discovery and mechanistic elucidation of likely causal mutations
Whole-genome sequencing tests for patients with end-stage cancers	To detect neoantigens that can be used for immune-therapy applications
Whole-genome sequencing tests for individuals to calculate risk of common disease	To detect variants associated with risk for a wide range of phenotypes, including rare diseases, common diseases, and drug responses

conduct rigorous tests of NGS pipelines, it is critical to have “gold standard” data sets (also called reference data sets) that contain “known” validated genetic sequences and variants to be used as benchmarks. The National Institute of Standards and Technology (NIST) has led an effort to create reference materials and data sets with associated known sequences by creating the Genome in a Bottle consortium (2), whose output includes several high-quality genome data sets established through sequencing on a variety of platforms.

Ideally, a large suite of data sets would be available to provide assurance that different types of variants in different contexts are adequately represented in pipeline testing and that specific platform biases are not driving the availability of reference data sets. Some of the data sets may be generated from sequencing human samples and others could be created using genome “synthesizers,” such as VarSim and HIVE Insilico, that can create *de novo* genomes with specific targeted variants. Synthesizers, if perfected, could represent a way to generate data sets for rigorous pipeline evaluation without bias toward performance on “known” samples. FDA’s regulatory science effort will benefit from community sharing of new synthesizer tools to aid in generating suitable data sets for evaluation of bioinformatics pipelines.

**(iii) Understand error models of NGS technologies, how these errors inform characterization, and how combinations of technologies may complement one another.** The various existing sequencing platforms demonstrate different biases and errors; these differences will only increase as new platforms emerge. Developing an error profile for each technology will help guide decisions surrounding the types of interrogated genomic regions for which the technology is best suited and, by extension, the range of expected uses for which it might be deployed. For example, if optimal clinical impact could be achieved by combining platforms and technologies (for example, 90% short read, 10% long read, in a mixture), then principles to evaluate tradeoffs and to design hybrid tests will need to be developed. These decisions may be guided by “error model” abstraction software tools [such as ART (3)], which could be made available through precisionFDA in a documented and version-controlled manner. The suitability of these existing error-model abstractions will need to be assessed, and additional research into representative error models for each platform will need to be pursued. The availability of gold-standard genome sequence data sequenced by multiple vendors on the

precisionFDA portal could encourage experimentation with such combinatorial tests.

**(iv) Develop competitions for systematic, summary statistic–based comparison of NGS pipelines.** In order to optimize NGS performance over time, a comprehensive suite of metrics to evaluate how well different NGS platforms perform in the context of a variety of expected uses is an important goal for FDA regulatory science. A series of precisionFDA competitions could be organized to build communal knowledge of high-quality pipelines and best practices. These competitions would benefit from comprehensive gold-standard data sets as well as software to compare the performance of candidate submissions (usually through metrics such as sensitivity, specificity, positive and negative predictive value, and other widely used measures of reliability and accuracy).

Competition success metrics could reflect performance focused around specific uses as well as overall performance of candidate platforms in the contexts of the type of variation, interrogated regions, and intended use. A key challenge would be to identify sources of variability and systematic bias, if any, and encourage the community to address them. New or optimized informatics tools built through this effort could be shared on the precisionFDA platform, allowing researchers and, eventually, regulatory applicants—those submitting new applications to FDA—to evaluate their own pipelines.

**(v) Understand strengths and limitations of different benchmarking strategies using a variety of data types.** Benchmarking methods are likely to vary in their ability to evaluate the different wet-lab and informatics stages of a pipeline. Whereas entirely synthetic data are clearly defined and characterized, they may not reflect all the features of natural human DNA sequences. Conversely, natural data that capture these features might not be perfectly characterized and may contain undefined sequence elements. Hybrid methods that inject synthetic variation into natural sequences have strengths and weaknesses, too. Therefore, an important research goal is to compare natural data with different strategies for creating synthetic test sequences to understand the utility of various synthetic strategies.

**(vi) Understand the clinical relevance of population genetic information on the detection, characterization, and interpretation of variants.** A core principle of the PMI is the inclusion of diverse, underrepresented populations. A critical challenge for clinical NGS is to accurately identify medically relevant variation in the context of an ethnically and

geographically diverse and admixed target population. The issues of causal versus linked variants, baseline variation in each population, the creation of ethnicity-specific reference genomes for performance characterization, and methods for analyzing admixed genomes (genomes that have several contributing ancestries) are all relevant to FDA regulatory science, to test developers, and, more generally, to precision medicine. Collection of high-quality samples representing many population groups through the PMI and other efforts will enable their characterization and contribute to the creation of gold standard reference data sets for specific ethnicities and geographically defined groups. The precisionFDA community might help in determining the proper role of population-specific reference genomes in benchmarking clinical tests. Preliminarily, it seems reasonable to suggest that tools that will help investigators generate realistic genomes (by synthesis, injection, or new methods) should incorporate principles of population genetics, and that computational analysis pipelines should be tested on such genomes. Development and dissemination of practices for detecting and incorporating linkage disequilibrium–based inference into pipelines, an understanding of when inferences are robust or brittle, and the relevance of such information will also be important for many clinical NGS tests.

**(vii) Understand costs and performance tradeoffs of NGS strategies in nuanced clinical contexts.** Not all genetic variation is equally important under every circumstance. Ideally, clinically important variation would be easy to identify. Some important variation is, and will continue to be, challenging [for example, human leukocyte antigen (HLA) typing and *CYP2D6* genotyping]. PrecisionFDA may play a role in catalyzing research into methods for recognizing medically important genomic regions and promoting the performance assessment of single and combinatorial technologies at effectively interrogating variants of known and unknown significance. These regions may be identified collaboratively with genetic data resources that focus on particular genes, diseases, or drug responses, while the overall characterization of NGS platforms for clinical use would emphasize performance in these critical areas. Appropriate performance stratification will not only allow a better understanding of the tradeoffs of different test designs, it might provide information useful for the design of new reference data sets.

**(viii) Understand how to use databases with clinically validated variants in the**

**assessment of individual technologies and their error rates.** The field of genetics is fortunate to have a number of public databases that catalog functionally critical variants alongside the evidence supporting each, providing focus on regions that are important for clinical applications of NGS. Key projects currently categorizing genetic variation of importance to human health include ClinVar, ClinGen, PharmGKB, LOVD [Leiden open (source) variation database], the Human Gene Mutation Database, and OMIM (Online Mendelian Inheritance in Man) as well as many locus- and disease-specific genetic databases. The value in these third-party resources could be leveraged in FDA regulatory science, which could seek to develop ways to evaluate their content and recognize them (and their standard operating procedures) as resources for test developers and clinicians to use in many of the activities described in the previous sections. These databases serve several useful functions: (i) They can provide evaluation of levels of evidence associated with genotype-phenotype correlations; (ii) they allow test developers and FDA to focus on loci of medical importance when evaluating the performance of informatics pipelines; and (iii) they provide a valuable longitudinal source of information about medically important variation that will inform, over time, many of the activities described above, without requiring FDA to mount parallel ef-

forts in genetic surveillance and capture of new knowledge.

**(ix) Understand how patients and practitioners comprehend and use genetic testing.** The goal of clinical genetic testing is to inform clinicians and patients about diagnoses, disease risks, adverse drug responses, therapeutic interventions, and other medically relevant issues. Numerous groups have emphasized that genetic test results should be presented to physicians and patients in a way that is understandable and informative for making rational choices about health care. The ability to understand the implications of genetic test results for health care decisions without always requiring the involvement of a genetics expert is critical if genetic testing is to become widely and effectively used in current and future health care settings. Regulatory science research could work with a broadly drawn cross section of both health care providers and the public to understand provider and patient preferences and limitations are adequately communicated within the label. Useful starting points could include discussions involving patients with diverse genetic disease diagnoses, to help articulate the kinds of information patients find most relevant, the level of certainty they find acceptable, and the support structures and additional information resources needed to support the recipients of NGS test results.

## MAKING PRECISION MEDICINE A REALITY

NGS is a transformative technology for clinical medicine and is poised to propel precision medicine into real-world clinical use. This Perspective reflects a number of regulatory science issues for NGS tests and identifies activities that will contribute to a robust understanding of NGS tests, all with the goal of enabling better test development and validation. Some of these activities are already in their initial stages, and some are, as yet, unaddressed. We present these ideas to spur advancements in NGS testing that allow this technology to reach its full potential in providing important health care information in a timely, safe, and effective manner.

## REFERENCES

1. PMI, <https://www.whitehouse.gov/precision-medicine>.
2. J. M. Zook, B. Chapman, J. Wang, D. Mittelman, O. Hofmann, W. Hide, M. Salit, Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat. Biotechnol.* **32**, 246–251 (2014).
3. W. Huang, L. Li, J. R. Myers, G. T. Marth, ART: A next-generation sequencing read simulator. *Bioinformatics* **28**, 593–594 (2012).

10.1126/scitranslmed.aaf7314

**Citation:** R. B. Altman, S. Prabhu, A. Sidow, J. M. Zook, R. Goldfeder, D. Litwack, E. Ashley, G. Asiminos, C. D. Bustamante, K. Donigan, K. M. Giacomini, E. Johansen, N. Khuri, E. Lee, X. S. Liang, M. Salit, O. Serang, Z. Tezak, D. P. Wall, E. Mansfield, T. Kass-Hout, A research roadmap for next-generation sequencing informatics. *Sci. Transl. Med.* **8**, 335ps10 (2016).



### A research roadmap for next-generation sequencing informatics

Russ B. Altman, Snehit Prabhu, Arend Sidow, Justin M. Zook, Rachel Goldfeder, David Litwack, Euan Ashley, George Asimenos, Carlos D. Bustamante, Katherine Donigan, Kathleen M. Giacomini, Elaine Johansen, Natalia Khuri, Eunice Lee, Xueying Sharon Liang, Marc Salit, Omar Serang, Zivana Tezak, Dennis P. Wall, Elizabeth Mansfield and Taha Kass-Hout (April 20, 2016)  
*Science Translational Medicine* **8** (335), 335ps10. [doi: 10.1126/scitranslmed.aaf7314]

Editor's Summary

---

The following resources related to this article are available online at <http://stm.sciencemag.org>.  
This information is current as of May 12, 2016.

---

**Article Tools** Visit the online version of this article to access the personalization and article tools:  
<http://stm.sciencemag.org/content/8/335/335ps10>

**Related Content** The editors suggest related resources on *Science's* sites:  
<http://stm.sciencemag.org/content/scitransmed/7/315/315ps22.full>  
<http://stm.sciencemag.org/content/scitransmed/6/252/252cm7.full>  
<http://stm.sciencemag.org/content/scitransmed/8/333/333ps8.full>  
<http://stm.sciencemag.org/content/scitransmed/7/296/296ed9.full>  
<http://stm.sciencemag.org/content/scitransmed/4/163/163fs43.full>  
<http://stm.sciencemag.org/content/scitransmed/4/150/150fs35.full>  
<http://stm.sciencemag.org/content/scitransmed/8/338/338ed6.full>

**Permissions** Obtain information about reproducing this article:  
<http://www.sciencemag.org/about/permissions.dtl>

*Science Translational Medicine* (print ISSN 1946-6234; online ISSN 1946-6242) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science Translational Medicine* is a registered trademark of AAAS.