Author Manuscript

Accepted for publication in a peer-reviewed journal

National Institute of Standards and Technology • U.S. Department of Commerce

Published in final edited form as:

Sci Transl Med. 2016 April 20; 8(335): 335ps10. doi:10.1126/scitranslmed.aaf7314.

A Roadmap for Regulatory Science Research for Next Generation Sequencing Informatics

Russ B Altman,
Stanford

Snehit Prabhu,
Stanford

Arend Sidow,
Stanford

Justin Zook,
NIST

Rachel Goldfeder,
Stanford

David Litwack,
FDA/CDRH

Euan Ashley,
Stanford

George Asimenos,
DNANexus

Carlos Bustamante,
Stanford

Katherine Donigan,
FDA/CDRH

Kathy Giacomini,
UCSF

Elaine Johansen,
FDA

Natalia Khuri,
Stanford

Eunice Lee,
FDA/CDRH

Sharon Liang,
FDA/CDRH

NIST Author Manuscript

NIST Author Manuscript

NIST Author Manuscript

Carol Linden,
FDA

Marc Salit,
NIST

Omar Serang,
DNANexus

Zivana Tezak,
FDA/CDRH

Dennis Wall,
Stanford

Elizabeth Mansfield, and
FDA/CDRH

Taha Kass-Hout
FDA

Introduction

The Precision Medicine Initiative (PMI) is a national effort in the United States “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 of goals of this effort is to bring about the routine use of the next generation of precision diagnostics to benefit individuals and the public health. Central to the introduction of safe and effective new precision diagnostic technologies is an adequate understanding of how they perform, both in terms of their likely benefits and their limitations. Through the PMI, FDA is seeking to address this by providing dynamic, flexible, and well-balanced regulation of precision diagnostics.. Because these complex technologies pose new challenges in understanding with respect to 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.

A series of meeting hosted by the UCSF-Stanford Center for Excellence in Regulatory Science & Innovation (CERSI) in September, 2015 included a public workshop and discussions on identifying key activities needed to evaluate and understand the clinical implications of NGS. This perspective summarizes the ideas and directions that were proposed and puts forth a working “roadmap” for NGS evaluation, as a possible exemplar of how multiple other new next generation diagnostics may be understood.

Background

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. Eventually, these developments are likely to culminate in the routine determination of patients' genome sequences. In the meantime there will be several years

during which the process of DNA sequence determination remains challenging, and where cost-, quality-, and goal-driven tradeoffs result in a very large diversity of testing strategies. This perspective is intended to lay out the technological challenges slowing the routine clinical use of the new generation of genetic testing, and asks what questions in regulatory science ought to be addressed to arrive at a flexible yet robust regulatory framework that results in maximum benefit for patients.

As part of its PMI effort, FDA is seeking to undertake and to support regulatory science research that will enhance 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 healthcare environment.

A centerpiece of this effort is precisionFDA, a research and development portal that 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 next generation sequencing (NGS) technology. [3] precisionFDA is a public, open-source, cloud-based platform developed by FDA and its contractor DNANexus that hosts shared tools, crowdsourced testing, and community challenges, to improve and share knowledge and methods for evaluating NGS bioinformatics pipelines.

This is currently a pre-regulatory 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 the FDA to rely on precisionFDA-based analyses to build standards and to communicate the technical performance of NGS tests.

In addition, FDA seeks to answer practical regulatory science questions such as which reference sequences and datasets 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 unique challenges to evaluation of the quality of an NGS test, which this document seeks to highlight.

For purposes of this paper, we will refer to the sequence or sequences of the genome 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), copy number variants or other types of variants. The **expected**

use¹ defines the anticipated clinical (or research) use of genomic information and will factor into the minimal performance characteristics that will make the test useful to researchers, providers and/or patients. Both the interrogated region and detectable variation depend critically on the **expected use** of the test results. For example, “low resolution” sequencing results might be sufficient to power tests looking for duplications and deletions, but higher resolution sequencing might be required to reliably detect smaller aberrations.

The range of expected uses envisioned today may include:

1. Non-invasive prenatal tests of a fetus. This type of test would be expected to report chromosomal abnormalities known to be associated with harmful or lethal phenotype.
2. Targeted gene panels to test a patient’s tumor. This type of test would be expected to report any somatic mutations in 'known' cancer driver genes.
3. Carrier tests for individuals of reproductive age. This type of test would be expected to report whether the individual carries known or novel pathogenic mutations in known heritable disease genes.
4. Whole genome sequencing tests for children with congenital diseases or conditions. This type of test is expected to aid in discovery and mechanistic elucidation of likely causal mutations.
5. Whole genome sequencing tests for patients with end-stage cancers. This type of test would be expected to detect neo-antigens that can be used for immune-therapy applications.
6. Whole genome sequencing tests for individuals to calculate risk of common disease. This type of test would be expected to detect variants associated with risk for a wide range of phenotypes including rare disease, common disease and drug response.

In the context of different expected uses, bioinformatic pipelines may 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, and the assessment of how they affect test performance is the vision behind the precisionFDA infrastructure: to accommodate a diverse set of use-cases for NGS test and to measure performance levels and understand tradeoffs as appropriate for each application.

A Regulatory Science Research Roadmap for NGS

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 the following research roadmap. We note that this roadmap focuses on questions that we believe

¹“Expected use” has some similarities to the FDA regulatory term “intended use” but because there are differences in details, we simplify here to “expected use”

should be asked and 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. Below, we highlight 9 areas of potential regulatory science investigation for FDA, including precisionFDA, accompanied by a discussion of aims, research milestones, software tools and data services.

1. 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 a large numbers of genomes. Therefore, it is important that precisionFDA offer security and confidentiality of information and access controls tied to specifics of the informed consent obtained from those who provide samples. precisionFDA also must be respectful of issues relating to intellectual property and ownership. When sharing genomes, precisionFDA will need to address issues including (but not limited to) creation of unique genome identifiers, searching genomes based on specific features (e.g. disease status), and adopting and promulgating standard formats of sequence representation and variant calls.

In similar vein, software developed by the precisionFDA community might be for private use, group sharing, or open to the public. Therefore precisionFDA needs a set of rules for access to software and will need to enforce 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 re-evaluation of an entire pipeline need to be identified, to ensure continued integrity of analyses performed.

The precisionFDA team, together with its contractor DNANexus, is building the first generation of these capabilities.

2. Create reference datasets based on diverse patterns of expected use

In order to conduct rigorous tests of NGS pipelines, it is critical to have “gold standard” datasets (also called reference datasets) that contain “known” validated genetic sequence and variants to be used as benchmarks. The National Institute of Standards and Technology (NIST) has led an effort to create reference materials and datasets with associated known sequence by creating the “Genome in a Bottle” consortium [2], whose output includes several high quality genome datasets established through sequencing on a variety of platforms. Ideally, a large suite of datasets 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 datasets. Some of the datasets may be generated from sequencing real samples and others could be created using genome “synthesizers” such as [names] that can create *de novo* genomes with specific targeted variants. Synthesizers, if perfected, could represent a way to generate datasets for rigorous pipeline evaluation without bias towards performance on “known” samples. FDA’s regulatory science effort will benefit from community sharing of new synthesizer tools to aid in generating suitable datasets for evaluation of bioinformatics pipelines.

3. 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 continue to emerge.. Developing an error profile for each technology will help guide decisions surrounding the types of interrogated regions for which the technology is best suited, and by extension, the range of expected uses for which it might be deployed. If optimal clinical impact could be achieved by combining platforms and technologies (e.g. 90% short read, 10% long read, in a mixture), then principles to evaluate tradeoffs and to design hybrid tests will need to be developed. Existing software tools could be made available through precisionFDA in a documented and version-controlled manner. The suitability of these existing “error model” abstractions (such as ART [3]) 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 precisionFDA could encourage experimentation with such combinatorial tests.

4. 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 a context of variety of expected uses is an important goal for FDA regulatory science. A series of precisionFDA competitions and “bake-offs” could be organized to build communal knowledge of high-quality pipelines and best practices. These competitions would benefit from both comprehensive gold-standard datasets as well as software to compare the performance of candidate submissions (usually through metrics like sensitivity, specificity, precision, 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 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 the FDA--to evaluate their own pipelines.

5. 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. While entirely synthetic data gives full control of the ground truth, it may not reflect all the natural features of DNA sequence. Conversely, natural data that captures these features may not be perfectly measured to provide absolute ground truth. Hybrid methods that inject synthetic variation into natural sequences have strengths and weaknesses too. Understanding the appropriate use of each method will require careful analysis, research and community engagement.

6. 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, under-represented 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 vs. 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, 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 datasets for specific ethnicities and geographically-defined groups. The precisionFDA community may 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 novel methods) should incorporate principles of population genetics, and pipelines should be tested on such genomes. Development and dissemination of practices of 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.

7. Understand costs and performance trade-offs 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, and less important variation may not require easy identification. Some important variation is, and will continue to be, challenging (e.g., 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 in these regions. These regions may be identified collaboratively with genetic data resources that focus on particular genes, diseases or drug response, 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, but may provide information useful for the design of new reference datasets.

8. 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, HGMD, OMIM as well as many locus-specific and disease-specific genetic databases. The value in these 3rd 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 provide several useful functions: (1) they can provide evaluation of levels of evidence associated with genotype/phenotype correlations, (2) they allow test developers and FDA to focus on loci of medical importance when evaluating the performance of informatics pipelines, (3) they provide a valuable longitudinal source of information about medically important variation that will inform many of the activities described above over time, without requiring the FDA to mount parallel efforts in genetic surveillance and capture of new knowledge.

9. Understand how patients and practitioners comprehend and use genetic testing

The goal of clinical genetic testing is to inform clinicians and patients about diagnosis, disease risks, adverse drug response, 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 to them for making rational choices about healthcare. The ability to understand the implications of genetic test results for healthcare decisions without always requiring the involvement of a genetics expert is critical if genetic testing is to become widely used and to scale effectively in the current and future healthcare setting. Regulatory science research could work with a broadly drawn cross section of both healthcare providers and the public, to understand provider and patient preferences for reporting and how test risks, benefits, and limitations are adequately communicated. Useful starting points could include discussions involving patients with diverse genetic disease diagnoses, to help articulate the kinds of information they find most relevant, the level of certainty they find acceptable, and the support structures and additional information resources that need to exist to support the recipients of NGS test results.

Summary

NGS is a transformative technology for clinical medicine and is poised to propel “precision medicine” into reality. This perspective reflects a number of regulatory science issues for NGS tests that were recognized at a series of meetings hosted by the UCSF-Stanford Center for Excellence in Regulatory Science & Innovation (CERSI). It identifies a number of activities that will contribute to a robust understanding of NGS tests and should enable better test development and validation. Some of these activities are already in their initial stages, and some are as yet unaddressed. We are presenting these ideas to advance the goal of enabling NGS tests to reach their full potential in providing important healthcare information in a timely, safe and effective manner.

Acknowledgments

Certain commercial equipment, instruments, or materials are identified in this paper only to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

References

1. [Accessed 16 March, 2016] <https://www.whitehouse.gov/precision-medicine>
1. Thatcher SA. DNA/RNA preparation for molecular detection. Clin Chem. 2015 Jan; 61(1):89–99. Epub 2014 Dec 1. Review. PubMed PMID: 25451869. DOI: 10.1373/clinchem.2014.221374 [PubMed: 25451869]
2. Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W, Salit M. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. Nat Biotechnol. 2014 Mar; 32(3):246–51. Epub 2014 Feb 16. PubMed PMID: 24531798. DOI: 10.1038/nbt.2835 [PubMed: 24531798]
3. Huang W, Li L, Myers JR, Marth GT. ART: a next-generation sequencing read simulator. Bioinformatics. 2012 Feb 15; 28(4):593–4. Epub 2011 Dec 23. PubMed PMID: 22199392; PubMed Central PMCID: PMC3278762. DOI: 10.1093/bioinformatics/btr708 [PubMed: 22199392]
4. Dewey FE, Chen R, Cordero SP, Ormond KE, Caleshu C, Karczewski KJ, Whirl-Carrillo M, Wheeler MT, Dudley JT, Byrnes JK, Cornejo OE, Knowles JW, Woon M, Sangkuhl K, Gong L, Thorn CF, Hebert JM, Capriotti E, David SP, Pavlovic A, West A, Thakuria JV, Ball MP, Zaranek AW, Rehm HL, Church GM, West JS, Bustamante CD, Snyder M, Altman RB, Klein TE, Butte AJ, Ashley EA. Phased whole-genome genetic risk in a family quartet using a major allele reference sequence. PLoS Genet. 2011 Sep.7(9):e1002280. Epub 2011 Sep 15. PubMed PMID: 21935354; PubMed Central PMCID: PMC3174201. doi: 10.1371/journal.pgen.1002280 [PubMed: 21935354]