Applied & Translational Genomics 2 (2013) 34–40

Contents lists available at ScienceDirect

Applied & Translational Genomics

journal homepage: www.elsevier.com/locate/atg

Review

Practical, ethical and regulatory considerations for the evolving medical and research genomics landscape☆

Gholson J. Lyon [a,b,*], Jeremy P. Segal [c,**]

[a] Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, NY, United States
[b] Utah Foundation for Biomedical Research, Salt Lake City, UT, United States
[c] New York Genome Center, New York City, NY, United States

ARTICLE INFO

Article history:
Received 11 December 2012
Received in revised form 13 February 2013
Accepted 13 February 2013

Keywords:
Genomics
Whole genome sequencing
Ethics
Regulation
CLIA
Exome
Genetic testing

ABSTRACT

Recent advances in sequencing technology are making possible the application of large-scale genomic analyses to individualized care, both in wellness and disease. However, a number of obstacles remain before genomic sequencing can become a routine part of clinical practice. One of the more significant and underappreciated is the lack of consensus regarding the proper environment and regulatory structure under which clinical genome sequencing and interpretation should be performed. The continued reliance on pure research vs. pure clinical models leads to problems for both research participants and patients in an era in which the lines between research and clinical practice are becoming increasingly blurred. Here, we discuss some of the ethical, regulatory and practical considerations that are emerging in the field of genomic medicine. We also propose that many of the cost and safety issues we are facing can be mitigated through expanded reliance on existing clinical regulatory frameworks and the implementation of distributive work-sharing strategies designed to leverage the strengths of our genomics centers and clinical interpretive teams.

Contents

Abbreviations: CLIA, Clinical Laboratory Improvements Amendments; NGS, Next generation sequencing; WGS, whole-genome sequencing.

* Correspondence to: G.J. Lyon, Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, NY, United States. Tel.: +1 1446468721219.
** Correspondence to: J.P. Segal, New York Genome Center, New York, NY. Tel.: +1 888 415 6942.
E-mail addresses: GholsonJLyon@gmail.com (G.J. Lyon), jeremypsegal@gmail.com (J.P. Segal).

1. Introduction

We are entering a fascinating and uncertain period of medical history, as today's DNA sequencing technology has the potential to help each of us direct our care and predict our future based on knowledge of our own individual inherited and acquired genetics. However, from a global and local economic perspective, these are lean years, and this adds a significant degree of uncertainty to the immediate future of this enterprise. It is therefore incumbent upon us to show that the personalized medical

2212-0661/$ – see front matter
http://dx.doi.org/10.1016/j.atg.2013.02.001

application of large-scale genomic analysis will not just be a luxury or a burdensome cost center, but that it truly has the potential to save both lives and health care expenses via data-driven management, early disease detection/screening and more efficacious pharmaceutical delivery. To this end, we need to determine how to move forward towards expanded clinical use of this technology in a manner both rapid and economical, while ensuring the integrity of the process and the safety and well-being of patients and research participants. This will require careful thought and consideration regarding the proper environment and regulatory structure surrounding genomics, as well as the development of consensus regarding what exactly constitutes a genetic test in the age of large-scale genomics and informatics.

2. Paving the way for the broad implementation of clinical genomic medicine

A report published in 2011 by the National Research Council for the National Academy of Sciences elegantly described the major divisions between the clinical and research worlds, including in regards to large-scale genomic analyses, such as whole genome (WGS) sequencing. The report went on to offer suggestions for how to help merge these two worlds, including articulating the need for a "Knowledge Network" and "New Taxonomy", with the recommendation that pilot studies along such lines should be conducted (Anon., 2011). However, the report did not address a critical issue related to genetic testing, namely the rules that should govern genomic research and clinical care as we move into the coming era of individualized medicine. The United States federal government mandates that any laboratory performing tests on human specimens "for the purpose of providing information for the diagnosis, prevention, or treatment of any disease" must satisfy the conditions set forth in the Clinical Laboratory Improvement Amendments (CLIA) of 1988 (Group®, 2012). Research laboratories performing investigative analyses of human samples that are not meant to provide clinically actionable results are currently considered exempt, and it is a simple fact that most research laboratories do not have sufficient standards in place to qualify them for CLIA approval (Lyon, 2012a,b). At the time CLIA was enacted, the separation of the clinical and research worlds seemed a fairly straightforward proposition. But today, the issues we face from a regulatory and ethical standpoint around genomics stem from the simple question: what do we do when it becomes difficult to draw a clear line of distinction between these two types of laboratory practices, particularly when researchers are working directly with families? Next generation sequencing (NGS) technology has fundamentally transformed what it means to perform research on human participants, as a direct consequence of its power, speed and efficacy (Lyon and Wang, 2012). Rare disease discovery research is nothing new: we have been analyzing samples from rare disorders in a scientific, hypothesis-driven way for more than a century. What has changed are the prospects for meaningful findings in a clinically actionable time frame, as well as the relative standardization of these research practices. Families afflicted with rare genetic disorders now have a reasonable expectation of definitive and potentially actionable results on the order of weeks to months, and all such families regardless of diagnosis are candidates for a relatively standardized genomic (rather than disease-specific mechanistic) analysis. The situation is similar for cancer patients, as standardized tumor-agnostic genomic analyses have a high likelihood of uncovering plausible drug targets during the lifetimes of patients, even those with late-stage disease.

The impact of altered expectations and standardized methods should not be underappreciated. Most critically, the implementation of NGS has resulted in a rapid evolution of research practices towards a process whereby suitable patients (usually with either cancer or a rare genetic disease) are referred by clinicians into exome (Ng et al., 2008, 2009, 2010a, 2010b; Musunuru et al., 2010; Choi et al., 2009; Albert et al., 2007; Hodges et al., 2007; Okou et al., 2007; Need et al., 2012) or whole genome sequencing (WGS) (Lam et al., 2012) research programs. However, rather than being driven mainly by a hypothesis-driven basic discovery motive, the goal is increasingly patient-focused and intended to find information of clinical benefit to the participant, indeed blurring the lines between "patient" and "research participant". Perhaps the key difference is ultimately with intent, but as the shift is made from basic discovery to participant-focused analysis, it becomes necessary to admit that, for these participants, the process is becoming a clinical one. Underscoring this trend is the release of a growing number of CLIA-certified exome sequencing tests at some of the larger genetics laboratories, including Baylor, Ambry Genetics, and elsewhere. However, there persists a healthy amount of disagreement regarding the proper environment and regulatory framework for human genome sequencing.

We believe that the observed shift from research to the clinic is natural and positive, a clear sign of the growing prospects for individualized or precision medicine (Anon., 2011). But, at the same time, this changing landscape begs an in-depth discussion of the ethical and regulatory issues that face genomics research and genomic medicine. Ultimately, the goal of such a discussion is to lead the field towards increased standardization, improved confidence in both individual results and shared datasets, and the elimination, or at least reduction, of redundant sequencing and confirmatory re-testing. It is our belief that we as a field need to re-think the boundaries between clinical and research practice. A recent report from the Presidential Commission for the Study of Bioethical Issues (Anon., 2012a) acknowledged this, and among their many recommendations we would like to highlight the following:

> **Recommendation 4.1**
> Funders of whole genome sequencing research, relevant clinical entities, and the commercial sector should facilitate explicit exchange of information between genomic researchers and clinicians, while maintaining robust data protection safeguards, so that whole genome sequence and health data can be shared to advance genomic medicine.
> Performing all whole genome sequencing in CLIA-approved laboratories would remove one of the barriers to data sharing. It would help ensure that whole genome sequencing generates high-quality data that clinicians and researchers can use to draw clinically relevant conclusions. It would also ensure that individuals who obtain their whole genome sequence data could share them more confidently in patient-driven research initiatives, producing more meaningful data.

3. Feasibility of clinical sequencing?

Some in the field might argue for the continued suitability of the "research first, clinical follow-up" model. Economics are a common area of concern, as the average cost for a CLIA-certified exome is currently about 2–3-fold higher than the price for a typical research exome at the same sequencing depth. As of November 2012, for example, one company appears to be charging $7900 for a trio of clinical grade exomes, including both parents and one child, for a cost of ~$2600/exome (Davies, 2012). In contrast, the typical cost for exomes in research laboratories is ~$1000–1500, depending on many variables. There are a few main reasons for this discrepancy between "research-grade" and "clinical-grade" sequencing, of which some are largely matters of accounting. The largest driver of price discrepancy between the research and clinical exomes is the hours of professional interpretation time built into the cost of the CLIA exome. On the research side, when a patient undergoes exome sequencing, typically the same level of interpretive effort is spent on their case, but it is performed both by teams of researchers as a basic component of their duties as well as by physicians on a non-billed basis. Regardless,

these experts spend their time, and that time has a real cost which is ultimately reimbursed, just not via straightforward accounting.

Another cost issue is that of result confirmation. Researchers are barred by CLIA from releasing non-CLIA certified results to participants or physicians that will impact diagnosis or management. Currently when an actionable result is found, a new sample can be sent to certain CLIA-approved laboratories specializing in the confirmation of that individual variant, at a cost of approximately $300 per variant (Anon., 2012b). However, until recently, it was not well known by researchers which companies will provide such testing for new genes, as one of us discovered and discussed in a recent commentary (Lyon, 2012b) and blog posting (Lyon, 2012a). This cost is not factored into the research exome, but is an included and currently essential component of all currently marketed CLIA exome offerings. Looking forward, as we learn more and more from the genome, and depending upon what types of data individuals want to receive from their genomes, the number of variants requiring confirmation will only increase, perhaps substantially. At $300 per variant (which is Sanger-based and unlikely to change anytime soon), the research model may rapidly become prohibitively expensive when used to manage care. Eventually the technology and informatics might allow for clinical-grade confidence in primary next generation sequencing data such that Sanger re-sequencing will not be required, but this could only legally be achieved via CLIA-certified NGS.

Other significant factors leading to higher CLIA prices are amortized validation/development costs and volume-related costs. Both of these factors reduce to issues of volume, as up-front validation costs would diminish in significance as volume increases. The typically low volume of samples processed by CLIA genomics centers affects price as well, as this either necessitates less frequent runs (with greater instrument depreciation per run) or under-multiplexing. For example, a larger CLIA genomics operation may run on the order of a few hundred clinical exomes per year, whereas a research center may sequence 20,000 or more with resultant economies of scale in sample and library preparation as well as sequencing. In the subsequent sections we will discuss mechanisms for performing more sequencing within CLIA operations, which may largely serve to obviate this problem.

4. Benefits of sequencing in CLIA-certified laboratories

Before delving into strategies for expanding the role of CLIA in genomic sequencing, it is worth exploring the potential benefits that increased oversight could produce for both clinical and research applications. There would clearly be some costs associated with expanded oversight that the benefits would have to outweigh. It is important to note that CLIA accreditation is much more than just a rubber stamp on a laboratory, and merely performing sequencing in a CLIA-certified facility does not sufficiently satisfy the requirements of the CLIA statutes, as we will discuss.

Prior to the enactment of CLIA, there were widespread abuses throughout the clinical laboratory community. For example, in the 1960s in New York State, a sweeping investigation of laboratories uncovered a startling array of deficiencies, including labs that were run by untrained personnel (including insurance salesmen), labs that routinely performed testing on spoiled specimens or used inadequate equipment, and labs (like one lab that reported every pregnancy test as positive) that simply returned fraudulent results (Dales, 1964; Godbout, 1960). We are certainly not suggesting that today's research-grade facilities in any way compare to these pre-CLIA labs, but we do believe that the straightforward practices dictated by CLIA can provide significant protections. Clearly history has shown that the quality of unregulated diagnostics is susceptible to perverse market forces, as all of the poor practices above can be tied directly to economic conflicts of interest. At a minimum, CLIA erects a wall protecting laboratories from these forces and provides an enforcement structure. The field would be prudent to embrace this protection, in light of the ongoing commoditization of sequencing (Mardis, 2011; Anon., 2010) and its associated potential for severe price competition.

As part of its requirements, CLIA mandates oversight of physical plant and staffing, quality control, quality assurance and quality improvement practices, laboratory documentation and specimen handling policies, individual test validation and performance procedures, interpretation and reporting practices, and proficiency testing programs. With respect to individual assays, any laboratory performing high-complexity non-FDA approved genetic tests must perform validation studies to establish (as appropriate) accuracy, precision, analytical sensitivity, analytical specificity, reportable range, reference intervals, and any other important performance characteristics, and to determine and document all calibration and control procedures. Bi-annual proficiency testing is mandated, involving a comparison with another CLIA facility using matched specimens, which is a critical process for uncovering outlier laboratories and un-noticed technical issues. These are all important and straightforward controls that provide the type of assurance that anyone should want before using genomic information to influence healthcare choices.

Though the “research first, clinical follow-up” model does provide CLIA assurance for select positive findings deemed clinically significant, this model has a number of drawbacks. Most fundamentally, the problem is not in how to handle positive results, but in how to understand and interpret negative or absent results. Without formal validation (accuracy, expected and actual coverage details, inter-lab validation, etc.) or clinical-grade specimen handling, an individual receiving a negative result might be overly confident in that result or be left wondering about the laboratory or specific test factors like coverage, informatics, and interpretation and ultimately be left unsure about how to proceed.

As mentioned previously, the number of clinically significant secondary findings from research exomes or genomes will only increase over time, and the only way to move towards a future without required Sanger confirmation is in the CLIA setting. An additional problem, addressed above, is that these findings are only currently permitted to break the research/clinical barrier if they pose a significant health risk (Wolf et al., 2012). All other information, including but not limited to pharmacogenomics information, carrier status, and disease risk associations is currently not returned. As the vast majority of exome and genome sequencing conducted today is performed in the research setting, this represents an enormous waste of potentially life-altering information. Moreover, the transference of that much data to participants and the resulting personal and clinical follow-through would be its own rich source of learning about the real-life practice of personalized medicine. Thus, we as a population are doubly deprived when that information remains behind the research firewall. Primary CLIA sequencing could circumvent this issue and potentially allow clinical analysis and use of all genomic data from many research projects. The ultimate goal, translating genomic information into care algorithms for well individuals and families, is a monumental challenge, and we will not get there unless we can move towards unleashing this information in the safest and most standardized way possible, and carefully examining and cataloging downstream ramifications.

Another overlooked benefit of performing more standardized genome sequencing would be its effect on resultant datasets, and the community's resultant confidence therein. Research genomics datasets routinely feed into a variety of databases, such as dbSNP, DECIPHER (Swaminathan et al., 2012), the NHLBI Exome Sequencing Project (Norton et al., 2012), and the 1000Genomes Project (Abecasis et al., 2012), to name just a few. These databases have become critical tools for physicians and researchers attempting to assign clinical meaning to variant sets. Thus, non-clinical data can have a dramatic impact on even the most optimally validated clinical assays. However, it is well

known that these data sets are imperfect, both with respect to primary results and annotation, to the extent that some authors are even arguing against adoption of increased clinical standards due to their unreliability (Ledoux, 2012). We would argue that regulatory reform will actually result in additional positive feedback, as CLIA results would help to improve both the depth and quality of clinical variant databases that would in turn boost our clinical interpretive power. When viewed in this light, sequencing in CLIA-certified laboratories could produce significant benefits for nearly any person undergoing genome sequencing, whether for research or clinical purposes.

5. Applying CLIA to genomics

While in theory it sounds reasonable to advocate for all exome or genome sequencing to be performed to CLIA standards, it begs the question: wouldn't this be a practical nightmare? There are many rational worries about this model, for while the actual sequencing processes may be, or at least could be, standardized, the informatics and interpretive processes are often designed to be quite different, both across centers but also across the myriad of individual clinical and translational research projects performed by all of the various groups and collaborators at each institution.

With respect to sequencing, while not all exomes and genomes sequenced would qualify (because many require modified library preps, read lengths or read depths), the majority of constitutional exomes and genomes run today (and a growing number of cancer exomes and genomes) are assayed by one of a small number of standardized library preparation and sequencing techniques. Today, most whole genome sequencing involves either fragmentation/ligation or Nextera-based library preparation and 100 base paired-end 30× sequencing (Illumina) or circularized "nanoball" preparation with 35 base paired-end sequencing (Complete Genomics). Exome sequencing is almost entirely performed on Illumina equipment using one of a few library preparation methods, with 100 base paired-end sequencing performed in the major research sequencing centers to an average depth of 70–100× to achieve >80% of the target region covered by 20 or more reads. Others have made suggestions for standardizing exome sequencing (Klein et al., 2012), and we believe it is high-time to establish such standards, at least for exomes being sequenced from live human beings, so that results can be returned to participants.

However, while sequencing is relatively standardizable, it is true that many of the downstream processes are not, as bioinformatics analyses and interpretive schemes can be extremely variable. While the desired informatics and interpretive analysis for healthy individuals might focus on alleles relevant for future disease risk, carrier status and pharmacogenomics, genomic analyses for rare diseases might instead focus on de novo, homozygous or X-linked disease variants, possibly in the context of a parent–child trio or preferably in the context of even larger families, including grandparents. Certain findings seen in one patient may escape detection in another patient simply due to differences in the basic strategy of analysis or the phenotype of the individuals. With respect to population studies, the analytical variation can be tremendous, with focuses ranging from ethnicity-specific variation to variation associated with complex disease, basic human phenotypes and evolutionary processes. The number of different performable analyses is limited only by the imagination. While the informed consent process for each individual study would be required to include a discussion of the analysis details, the process can be confusing for participants and easily leave them at the end unclear whether or not particular findings were investigated and frustrated by an inability to access the data. This being the case, it would be beneficial to move towards a system whereby a straightforward clinical analysis of data from research projects could be subsequently performed at a later time, within a proper regulatory framework.

This downstream variation in informatics and interpretation raises an important question: from the clinical standpoint, what exactly constitutes a genetic laboratory test? Is it simply the analytics (the sequencing), or is it a combination of analytics and interpretation, or is it the entire process from sample receipt through to the generation and return of a report? Here, the legal definition is really quite clear, as CLIA specifically states that a medical laboratory test is an all-encompassing process (Anon., 2013a). The introduction to CLIA subpart K states that "each laboratory that performs nonwaived testing must establish and maintain written policies and procedures that implement and monitor quality systems for all phases of the total testing process (that is, preanalytic, analytic, and postanalytic) as well as general laboratory systems" (see Table 1 for a summary of the analytic systems).

Table 1
Processes involved in a CLIA-certified genetic test.

Preanalytic system
1) Test request and specimen collection criteria
2) Specimen submission, handling and referral procedures
3) Preanalytic systems assessment
Analytic system
1) A detailed step-by-step procedure manual
2) Test systems, equipment, instruments, reagents, materials and supplies
3) Establishment and verification of performance specifications
4) Maintenance and function checks
5) Calibration and calibration verification procedures
6) Control procedures, test records, and corrective actions
7) Analytic systems assessment
Post-analytic system
1) Test report, including (among other things):
a) interpretation
b) reference ranges and normal values
2) Post-analytic systems assessment

It is noteworthy that test interpretation and reporting are specifically covered by the CLIA statutes and included as part of the regulated test process. This is important because, as the community has discovered, the actual sequencing has become increasingly straightforward, whereas the true difficulties and pitfalls lie in the informatics, interpretation and reporting. Any meaningful regulatory framework for NGS-based diagnostics must include oversight of informatics pathways and interpretive criteria, as there are simply too many ways to do informatics incorrectly, with resultant possibilities for harm to patients and participants.

This issue is beginning to get the attention of the agencies responsible for overseeing clinical laboratories, now that a large number of clinical laboratories have begun developing a variety of tests on NGS instruments. The College of American Pathologists (CAP) has recently released a new checklist for molecular pathology laboratories that includes both general laboratory and test development guidelines covering NGS wet lab practices, bioinformatics processing and data storage and transfer practices. Additionally, the New York State Department of Health Clinical Laboratory Evaluation Program (CLEP) has issued detailed guidelines for the development and validation of NGS cancer genomics assays (Anon., 2013b). New York is one of two CLIA-exempt states as a result of its own state licensure regulations being deemed "equal to, or more stringent than" CLIA by CMS per CLIA subpart E, thus clinical laboratories in New York receive their CLIA license through the state following successful state certification. The CLEP NGS oncology guidelines are quite thorough, including requirements for quality scores, control procedures, acceptable numbers of specimens for validation studies and guidelines for establishing read depth, accuracy, sensitivity, etc., focusing on actual performance rather than the details of bioinformatics pipelines. Overall, the regulatory framework for NGS on the pure clinical side is coming together, with certain aspects such as reporting criteria hopefully being sorted out in the near future.

However, if a clinical NGS test is defined by both the sequencing and downstream informatics, and the informatics possibilities for a standard sequence are essentially limitless, how could CLIA

supervision be applied to combined research and clinical genomics operations without placing an extreme regulatory burden on the sequencing laboratory? Would every analysis type need to be certified, or would a time-consuming standardized analysis be required even if it were not needed for each particular operation?

6. The distributive model: an analytical-interpretive split across genomics

Any ideal solution would allow sequencing centers to focus on their strengths and to leverage their economies of scale, without requiring them to devote their time to unnecessary informatics and interpretation. How can that be achieved in keeping with the spirit of proper CLIA oversight? As a solution, we would propose an analytic-interpretive split (or a so-called "distributive model") across both clinical and research genomics. This split model simply means that one laboratory performs analytics and then a second laboratory performs the interpretation and reporting. Thus, together, the two laboratories perform all the functions that make up a laboratory test. This should be a straightforward arrangement, but while some precedent and guidance policies exist, the regulatory structure that would govern such a system is still evolving, as we will discuss.

The benefits of enacting such a split model could be substantial, and we believe they could be gained without significantly burdening our sequencing centers with undue excess costs. Under this type of system, the basic sample processing and sequencing operation could be standardized across clinical patients and the majority of new genomics research participants. The practical effect of this split would be to turn an exome or genome sequence into a discrete deliverable unit that could be used for multiple downstream purposes by multiple downstream labs. For each patient or participant, the same validated sequencing would be performed, and that raw data, if individually desired, could be merged with their electronic health record. Then, the data could be directly used per consent agreement with researchers for any type of downstream investigative analysis. It could also be used for any type of downstream clinical analysis for guidance of healthcare decisions. As we move towards higher-quality primary data, depending on the type of analysis, this CLIA-certified raw data could be interpreted and used to guide healthcare choices with less requirements for Sanger validation, obviating the need for secondary finding confirmation and resulting in large cost savings.

In turn, we believe the resources required to install CLIA oversight of genomics centers would not be prohibitive. Genomics labs would need to hire personnel and set up the physical plant commensurate with CLIA and state regulations. Most genomics centers, being relatively new and well equipped, would not require major modifications. In terms of test development, a modest number of assays and confirmatory tests would need to be run to establish performance parameters, as discussed below, but compared to the high sample volumes run by these centers it should not be overly burdensome. Of course, we fully realize that there are many already biobanked samples in the research system, many of which have been deidentified and/or not ideally consented for return of results, and it will therefore not be possible to apply the above standards to such samples without considerable expense and effort to re-consent prior participants. Thus, we are suggesting the above mainly for new patients and research participants.

7. Regulatory considerations

The Federal CLIA statutes clearly state that all phases of testing, from test submission up to interpretation and reporting procedures, are subject to oversight by the Centers for Medicare and Medicaid Services (CMS). Assuming that participation in any one or more of these testing phases equates to involvement in medical testing then, practically speaking, if two laboratories agree to split the processes of a medical test, then both should require CLIA accreditation. This is a point of critical importance, because in the coming years it is likely that economic forces will lead to more whole genome sequencing being performed in larger centers, at least until sequencing becomes so inexpensive and ubiquitous that it diffuses back into smaller labs and even office settings. But until then, we are likely to live through a time when sequencing will be standardized and centralized, and yet many different downstream clinical analytic processes may be desired. Enacting the type of split described herein allows organizations to leverage their strengths in sequencing or bioinformatics and lower costs across the entire process. It may also help spawn a generation of laboratories devoted to clinical genomic data mining. But if so, it is essential that we as a community think of them and treat them in that way, as laboratories, even if they may be separated from the wet lab. Their post-analytic processing would simply be an extension of a prior wet-lab process, and a core component of a medical test. We must come to accept the in silico manipulation and mining of DNA information as not different from its physical manipulation in aqueous solutions.

If we lose sight of informatics and interpretation as core components of medical testing, we will take on huge risk as a community in the form of unregulated and unsupervised genomic interpretation. In the clinical genomics field, the informatics and interpretation are the portions of the process most fraught with complexity and that require the most in-depth knowledge. Every year, sequencing becomes more straightforward, but analytics and especially medical interpretation are becoming more and more complex. These processes require strict oversight if the clinical genomics field is going to be capable of maintaining high standards of information and recommendations for the individuals we serve. CLIA is one available avenue to ensure on a constant, ongoing basis that these processes are being applied correctly, by demanding transparency of practices, disclosure of validation data, review of personnel, operating procedures and patient protection practices, and requiring a system of constant re-evaluation and improvement.

It is our sense that the various regulatory agencies are realizing the possibility of additional upcoming test cases for genomics distributive models and are giving the matter serious thought. From discussions with individuals at CMS and CAP, their view is that in almost all cases an organization performing post-analytic processes would require proper CLIA certification, but they do not yet have published guidelines on the matter. CLIA subpart K, section 493.1242 dealing with specimen submission, handling, and referral, states that a CLIA certified lab "must refer a specimen for testing only to [another] CLIA certified laboratory or a laboratory meeting equivalent requirements as determined by CMS". While not explicitly covering a split, this at least offers a framework for inter-lab cooperation. The only governing body we are aware of that has made a clear statement on the issue of distributive models is the New York State CLEP program, which released business practice guidelines in 2008 covering this model (http://www.wadsworth.org/labcert/clep/Administrative/NYSBusinessPracticeGuidelines.pdf). CLEP requires both parties in a distributive model (analytic and interpretive) to be fully licensed, unless the interpretation is performed by a solo practitioner or small physician group exclusively for their own patients. The guidelines were written in this way so as not to infringe on the right of a physician to practice medicine in the manner they see fit, which is another significant though perhaps less important gray area that should be examined in the future.

8. Practical application and suggested standards

In the largest operating genome sequencing centers in America (Broad, Baylor, WashU-St. Louis), most, if not all, research exomes are sequenced to a depth of coverage of 20 reads or more per base pair in >80% of the target region. In general, the clinical exomes on

the market include sequencing to a somewhat higher average depth, 100–150× compared with 70–100× for the research exomes. The reason for this emphasis on deeper sequencing for the clinical exomes is the degree of variability across exome capture baits with respect to pull-down efficacy. This is common across all exome capture technologies and results in a skewed depth distribution (Sulonen et al., 2011), thus requiring high average coverage in order to cover a desired proportion of the exome at a sufficient depth to make accurate calls. From a clinical perspective, as depth increases, the sensitivity and specificity of variant calls improve dramatically, minimizing unnecessary confirmation.

WGS presents a different picture, as sequencing of unselected fragments results in a far more even read depth distribution profile. The current CLIA-certified WGS at Illumina is performed with an average sequencing depth of >30×, with >95% of all calls made at a depth of greater than 10 reads or more, which is essentially the same as the research gold standard, making the establishment of crossover procedures more straightforward compared with exome sequencing. It should be noted that the basic Illumina CLIA genome does not include a confirmation methodology and thus the report recommends separate confirmation for concerning findings prior to patient management (Illumina, 2012). Though Illumina is moving towards a full-process interpreted genome, their original uninterpreted genome is actually an ideal example of the split model we are advocating for, as it is essentially a technical product that requires downstream analysis, interpretation and confirmation in a CLIA-approved setting. We would encourage Illumina to consider expanding this basic CLIA process, currently a low-volume and higher cost separate sample pathway, and to move towards making it a standard offering available to both clinical and research specimens processed in their sequencing core.

The issue of clinical sequencing depth is tricky, particularly with respect to exomes, because it is difficult to advocate for raising depth standards across sequencing applications when the sequencing is already quite expensive. Though there is a firm rationale for deeper sequencing for clinical exomes, depending on the application, lower depth may provide sufficient data for research applications. In the long term, we believe that the advantages of somewhat higher depth clinical-grade data would outweigh the associated costs, but the issue of payers and resources is by no means trivial. Perhaps the recommendations for split models are best suited for WGS applications, as the current basic research and clinical standards are quite similar. Here, for both applications, the largest barrier to higher depth is cost, and even though 30× average depth is the current standard, there is a general understanding that this depth opens the door to significant numbers of false positive and negative findings, and represents a barrier to sensitive detection of larger structural rearrangements.

So, in the future it would be desirable for the research and clinical field to move towards higher quality/depth sequencing standards. It should be noted that these recommendations are not for overnight adoption. In this field, we have seen repeatedly that the impossible can seem simple with the passage of mere months. Even today, with the HiSeq 2500, we are capable of performing WGS in just a day or two for approximately 4–5 thousand dollars. This machine supports paired reads of 150 bp, 50% higher than HiSeq 2000 reads, and already certain groups have reported success with dramatically longer read lengths, which directly produces greater depth. Thus, already this machine should produce genomes significantly in excess of 30×, and as prices continue to drop, the costs associated with increased depth will seem more and more reasonable. This is particularly true when these costs are weighed against the potential benefits, which include full clinical utility of the primary sequence, better quality data for projects and resulting databases, reduced prices for clinical sequencing tests, and reduced costs associated with confirming variants, which eventually will approach zero as sequencing power and quality improves over time.

9. Ethical rationale for more broadly applied clinical sequencing

The topic of analytical variation touches on the critical related ethical issue of return of results, expectations and therapeutic misconception, about which much has already been written (Zawati and Knoppers, 2012; Tabor et al., 2011a, 2011b, 2012; Fisher, 2012; Lyon, 2012c; Angrist, 2011, 2013; Lunshof et al., 2010; Angrist and Cook-Deegan, 2006). Individuals, whether they are well or ill or participating in a research study, need to be informed during the consent process about the exact specifics of the proposed analysis. This is the interpretive corollary of the research sequencing dilemma that arises when clinicians and participants are left uninformed regarding expected data quality, coverage, and to what degree the data actually meet those standards. Of course, we fully realize that our society is rapidly moving toward more and more online activity, with younger generations more willing to share their data freely among each other. At some point, we hope that whole genome sequencing will occur at birth, with all data uploaded to one's own secure profile, to be used to guide individual health decisions into adulthood and old age and also as a baseline reference for particular disease states (cancer, etc.). This is in line with the idea that all medical records and information should be made available to each respective person. As recently articulated by Maynard Olson at the Cold Spring Harbor Laboratory Personal Genomes and Medical Genomic conference, such an arrangement could allow certain people to opt into a “medical information donor network”, whereby they would donate their longitudinal phenotype and medical laboratory data, including not only genomic data but also imaging, blood testing, and many other things. One can already see that such a thing is possible with the recent move by the company Facebook to allow people to register themselves as medical organ donors on their Facebook page (Cohen, 2012). Also, the Personal Genome Project is helping to break down some of the fear and mythologies surrounding the public dissemination of genomic data (Ball et al., 2012).

10. Conclusions

NGS-based sequencing has revolutionized how we pursue genetics research, and it is starting to produce a similar effect in the clinical genetics space. But it is not just our knowledge, scientific direction and expectations for future medical applications that are being transformed. As we move forward, this technology is beginning to push against our basic historic attitudes regarding what constitutes research vs. medical practice, and is creating a host of resultant ethical and regulatory dilemmas. Clinical assays are gaining a significant discovery component, while our research activities are producing an explosion of potentially medically relevant data. As a community, it is essential to explore and intelligently navigate these issues to ensure a rapid and smooth transition to a time when genomic analysis is a standard component of individualized care.

Current cost/benefit considerations are largely responsible for impeding wider adoption of clinical genome sequencing. Some argue that the information gleaned for most individuals does not outweigh the sequencing cost. Merging clinical and research data collection via a distributive model would help this enormously, as many clinical patients and research participants have legitimate sequencing needs and one validated sequencing operation would then produce multiple benefits. Higher volume clinical-grade genome sequencing would also drive down clinical sequencing costs, which are currently a 2–3× multiple of the research costs. Analysis possibilities would then be left open for any sample: research samples, even if not initially analyzed for clinical benefit, could be re-processed clinically at any time. This is a huge potential cost saver, enabling increased clinical utility to flow from a huge array of genomics projects, along with improved quality assurance for research samples (Lyon, 2012b; Poste, 2012) and the elimination of duplicative sequencing. The benefits will

grow exponentially as data quality and general comfort level increases to the point where base-calling confidence statistics obviate the need for most confirmatory testing. Improved genomics databases will also enable more powerful and higher quality future genome interpretation. Therefore, clinical-grade results will enable all information from a genome/exome to be used for the benefit of the individual, unlike the current system where many results are not allowed to pass through the research-clinical wall. All of these benefits will serve to provide positive feedback and shift the cost/benefit ratio in favor of further expanded sequencing.

Next-generation sequencing (NGS) is clearly a disruptive technology (Christensen, 2003; Christensen et al., 2009; Wu, 2010; Topol, 2012). However, it is important to keep in mind that each of the disruptive technologies of the past (the automobile, airplane, computer, etc.) were certainly not free of regulatory or ethical dilemmas at the time of their introduction, or even to this day, yet each has led to unmistakable improvements in our quality of life. This will ultimately be the case with the genomics revolution, but to realize this promise we need to plan carefully as a community while remaining cognizant of ethical and regulatory considerations. We believe that greater reliance on existing CLIA regulations represents the most viable way forward, as CLIA provides clear guidance for ethical issues related to patient safety and clinical/research demarcation and points us towards practical and efficient solutions for bringing about true clinical genomics on a large scale.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

G.J.L. is supported by funds from the Stanley Institute for Cognitive Genomics at Cold Spring Harbor Laboratory. We thank Jason O'Rawe, Kai Wang, and Filipe Ribeiro for comments on an earlier version of the manuscript.

References

Abecasis, G.R., et al., 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491 (7422), 56–65.
Albert, T.J., et al., 2007. Direct selection of human genomic loci by microarray hybridization. Nature Methods 4 (11), 903–905.
Angrist, M., 2011. You never call, you never write: why return of 'omic' results to research participants is both a good idea and a moral imperative. Personalized Medicine 8 (6), 651–657.
Angrist, M., 2013. Genetic privacy needs a more nuanced approach. Nature 494 (7435), 7.
Angrist, M., Cook-Deegan, R.M., 2006. Who owns the genome? New Atlantis 11, 87–96.
Anon., 2010. Human genome at ten: the sequence explosion. Nature 464 (7289), 670–671.
Anon., 2011. Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease. New York Academy of Sciences, Washington DC.
Anon., 2012a. Privacy and Progress in Whole Genome Sequencing. Available from: http://bioethics.gov/cms/node/764.
Anon., 2012b. Carrier/Mutation-specific Testing. Available from: http://www.genedx.com/test-catalog/mutation-specific-testing/.
Anon., 2013a. HHS website with CLIA regulations. Available from: http://wwwn.cdc.gov/clia/regs/toc.aspx.
Anon., 2013b. New York State Guidelines for Development and Validation of NGS Cancer Genomics Assays. Available from: http://www.wadsworth.org/labcert/TestApproval/forms/NextGenSeq_ONCO_Guidelines.pdf.
Ball, M.P., et al., 2012. A public resource facilitating clinical use of genomes. Proceedings of the National Academy of Sciences of the United States of America 109 (30), 11920–11927.
Choi, M., et al., 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proceedings of the National Academy of Sciences of the United States of America 106 (45), 19096–19101.
Christensen, C.M., 2003. The Innovator's Dilemma: the Revolutionary Book that will Change the Way You Do Business. HarperCollins, New York (xxxii, 286 pp.).
Christensen, C.M., Grossman, J.H., Hwang, J., 2009. The Innovator's Prescription: a Disruptive Solution for Health Care (p. li, 441 pp.).
Cohen, D., 2012. Facebook Launches Organ-donation Tool. Available from: http://allfacebook.com/organ-donation_b87500.
Dales, D., 1964. Governor Signs Laboratory Bill: Measure sets up Licensing and Testing Procedures. New York Times, New York City.
Davies, K., 2012. Ambry Genetics Catches the Clinical Sequencing Wave. (Available from: http://www.bio-itworld.com/news/06/13/12/Ambry-Genetics-catches-clinical-sequencing-wave.html).
Fisher, R., 2012. A closer look revisited: are we subjects or are we donors? Genetics in Medicine 14 (4), 458–460.
Godbout, O., 1960. Unethical Medical Laboratories Face Curbs in City Crackdown. New York Times, New York City.
Group® U., 2012. Personalized Medicine: Trends and Prospects for the New Science of Genetic Testing and Molecular Diagnostics. (Available from: http://www.unitedhealthgroup.com/newsroom/news.aspx?id=e1e22479-3600-4834-9633-a30afa84f972).
Hodges, E., et al., 2007. Genome-wide in situ exon capture for selective resequencing. Nature Genetics 39 (12), 1522–1527.
Illumina, 2012. Available from: http://www.illumina.com/clinical/illumina_clinical_laboratory.ilmn.
Klein, C., Konig, I.R., Lohmann, K., 2012. Exome sequencing for gene discovery: time to set standard criteria. Annals of Neurology 72 (4), 627–628.
Lam, H.Y., et al., 2012. Performance comparison of whole-genome sequencing platforms. Nature Biotechnology 30 (1), 78–82.
Ledoux, M.S., 2012. Exome sequencing for gene discovery: time does not stand still. Annals of Neurology 72 (4), 628–629.
Lunshof, J.E., et al., 2010. Personal genomes in progress: from the human genome project to the personal genome project. Dialogues in Clinical Neuroscience 12 (1), 47–60.
Lyon, G.J., 2012a. Guest Post: Time to Bring Human Genome Sequencing into the Clinic. (Available from: http://www.genomesunzipped.org/2012/02/guest-post-time-to-bring-human-genome-sequencing-into-the-clinic.php).
Lyon, G.J., 2012b. Personalized medicine: bring clinical standards to human-genetics research. Nature 482 (7385), 300–301.
Lyon, G.J., 2012c. There is nothing "incidental" about unrelated findings. Personalized Medicine 9 (2), 163–166.
Lyon, G.J., Wang, K., 2012. Identifying disease mutations in genomic medicine settings: current challenges and how to accelerate progress. Genome Medicine 4 (7), 58.
Mardis, E.R., 2011. A decade's perspective on DNA sequencing technology. Nature 470 (7333), 198–203.
Musunuru, K., et al., 2010. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. The New England Journal of Medicine 363 (23), 2220–2227.
Need, A.C., et al., 2012. Clinical application of exome sequencing in undiagnosed genetic conditions. Journal of Medical Genetics 49 (6), 353–361.
Ng, P.C., et al., 2008. Genetic variation in an individual human exome. PLoS Genetics 4 (8), e1000160.
Ng, S.B., et al., 2009. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461 (7261), 272–276.
Ng, S.B., et al., 2010a. Exome sequencing identifies the cause of a Mendelian disorder. Nature Genetics 42 (1), 30–35.
Ng, S.B., et al., 2010b. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature Genetics 42 (9), 790–793.
Norton, N., et al., 2012. Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. Circulation Cardiovascular Genetics 5 (2), 167–174.
Okou, D.T., et al., 2007. Microarray-based genomic selection for high-throughput resequencing. Nature Methods 4 (11), 907–909.
Poste, G., 2012. Biospecimens, biomarkers, and burgeoning data: the imperative for more rigorous research standards. Trends in Molecular Medicine 18 (12), 717–722.
Sulonen, A.M., et al., 2011. Comparison of solution-based exome capture methods for next generation sequencing. Genome Biology 12 (9), R94.
Swaminathan, G.J., et al., 2012. DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. Human Molecular Genetics 21 (R1), R37–R44.
Tabor, H.K., et al., 2011a. Genomics really gets personal: how exome and whole genome sequencing challenge the ethical framework of human genetics research. American Journal of Medical Genetics. Part A 155A (12), 2916–2924.
Tabor, H.K., et al., 2011b. Parent perspectives on pediatric genetic research and implications for genotype-driven research recruitment. Journal of Empirical Research on Human Research Ethics: JERHRE 6 (4), 41–52.
Tabor, H.K., et al., 2012. Informed consent for whole genome sequencing: a qualitative analysis of participant expectations and perceptions of risks, benefits, and harms. American Journal of Medical Genetics. Part A 158A (6), 1310–1319.
Topol, E.J., 2012. The Creative Destruction of Medicine: How the Digital Revolution Will Create Better Health Care. Basic Books, New York (xi, 303 pp.).
Wolf, S.M., et al., 2012. Managing incidental findings and research results in genomic research involving biobanks and archived data sets. Genetics in Medicine: Official Journal of the American College of Medical Genetics 14 (4), 361–384.
Wu, T., 2010. The Master Switch: the Rise and Fall of Information Empires, 1st ed. Alfred A. Knopf, New York (x, 366 pp.).
Zawati, M.H., Knoppers, B.M., 2012. International normative perspectives on the return of individual research results and incidental findings in genomic biobanks. Genetics in Medicine: Official Journal of the American College of Medical Genetics 14 (4), 484–489.