Author Manuscript

HHS Public Access
Author manuscript
J Mol Diagn. Author manuscript; available in PMC 2019 May 07.

Published in final edited form as:
J Mol Diagn. 2012 November ; 14(6): 525–540. doi:10.1016/j.jmoldx.2012.04.006.

Opportunities and Challenges Associated with Clinical Diagnostic Genome Sequencing:

A Report of the Association for Molecular Pathology

Iris Schrijver[*,†], **Nazneen Aziz**[*,‡], **Daniel H. Farkas**[*,§], **Manohar Furtado**[*,¶], **Andrea Ferreira Gonzalez**[*,‖], **Timothy C. Greiner**[*,**], **Wayne W. Grody**[*,††], **Tina Hambuch**[*,‡‡], **Lisa Kalman**[*,§§], **Jeffrey A. Kant**[*,¶¶], **Roger D. Klein**[*,‖‖], **Debra G. B. Leonard**[*,***], **Ira M. Lubin**[*,§§], **Rong Mao**[*,†††], **Narasimhan Nagan**[*,‡‡‡], **Victoria M. Pratt**[*,§§§], **Mark E. Sobel**[*,¶¶¶], **Karl V. Voelkerding**[*,‖‖‖,****], and **Jane S. Gibson**[*,††††]

[*]Whole Genome Analysis Working Group, Association for Molecular Pathology Clinical Practice Committee, Bethesda, Maryland [†]Departments of Pathology and Pediatrics, Stanford University School of Medicine, Stanford, California [‡]Department of Transformation Program Office, College of American Pathologists, Northfield, Illinois [§]Department of Clinical Diagnostics, Sequenom Center for Molecular Medicine, Grand Rapids, Michigan [¶]Department of Molecular Medicine, Life Technologies, Foster City, California [‖]Department of Pathology, Virginia Commonwealth University, Richmond, Virginia [**]Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska [††]Department of Pathology and Laboratory Medicine, University of California, Los Angeles School of Medicine, Los Angeles, California [‡‡]Illumina Clinical Services Lab, San Diego, California [§§]Division of Laboratory Science and Standards, Centers for Disease Control and Prevention, Atlanta, Georgia [¶¶]Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania [‖‖]Molecular Oncology, Blood Center of Wisconsin, and the Department of Pathology, Medical College of Wisconsin, Milwaukee, Wisconsin [***]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, New York [†††]Department of Pathology, University of Utah/ARUP Laboratories, Salt Lake City, Utah [‡‡‡]Esoterix Genetic Laboratories, LLC, Westborough,

Address correspondence to: Iris Schrijver, M.D., Department of Pathology, Stanford University School of Medicine, 300 Pasteur Dr, Room L235, Stanford, CA 94305. ischrijver@stanfordmed.org. Address reprint requests to: The Association for Molecular Pathology, 9650 Rockville Pike, Bethesda, MD 20814-3993. amp@amp.org.
Current address of R.K.: Department of Molecular Pathology, Cleveland Clinic, Highland Heights, KY.

Disclosures: V.M.P. is employed by and holds stock in Quest Diagnostics, M.F. is employed by and holds stock in Life Tech, J.A.K. is a member of the advisory board at Complete Genomics, T.H. is employed by and holds stock in Illumina, and N.N. is employed by Esoterix.

The Whole Genome Analysis Working Group is a working group of the AMP Clinical Practice Committee. The 2010–2011 Whole Genome Analysis Working Group consisted of Nazneen Aziz, Pinar Bayrak-Toydemir, Philip Cotter, Daniel H. Farkas, Andrea Ferreira-Gonzalez, Manohar Furtado, Amy Gargas, Jane S. Gibson (Chair 2010–2011), Timothy C. Greiner, Tina Hambuch, Lisa Kalman, Roger D. Klein, Jennifer Leib, Debra G. B. Leonard, Ira M. Lubin, Elaine Lyon, Karen Mann, Rong Mao, Narasimhan Nagan, Victoria M. Pratt, Iris Schrijver, Mark E. Sobel, Patrik Vitazka, and Karl V. Voelkerding.

The 2010–2011 AMP Clinical Practice Committee consisted of M. Fernanda Sabato Charreun, Christine Curtis, Michelle Dolan, William Funkhouser, Jane Gibson, Cyrus Hedvat, Loren Joseph, Neal Lindeman, Janina Longtine (Chair 2011), Ira Lubin, Kathleen Montone, Federico Monzon, Narasimhan Nagan, Joseph Pulliam, Iris Schrijver (Chair 2010), Siby Sebastian, Laura Tafe, Patrik Vitazka, Jeffrey Wisotzkey, and Donna Wolk.

Standard of practice is not being defined, and there may be alternatives. See *Disclaimer* for further details.

The findings, conclusions, and opinions expressed in this report do not necessarily reflect those of any individual committee member or his or her employer, including the Centers for Disease Control and Prevention or the Agency for Toxic Substances and Disease Registry.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Massachusetts [§§§]Molecular Genetics, Quest Diagnostics, Nichols Institute, Chantilly, Virginia [¶¶¶]American Society of Investigative Pathology, Bethesda, Maryland [‖‖‖]Department of Pathology, University of Utah School of Medicine, Salt Lake City, Utah [****]ARUP Institute for Experimental and Clinical Pathology, Salt Lake City, Utah [††††]Department of Clinical Sciences–Pathology, University of Central Florida College of Medicine, Orlando, Florida

Abstract

This report of the Whole Genome Analysis group of the Association for Molecular Pathology illuminates the opportunities and challenges associated with clinical diagnostic genome sequencing. With the reality of clinical application of next-generation sequencing, technical aspects of molecular testing can be accomplished at greater speed and with higher volume, while much information is obtained. Although this testing is a next logical step for molecular pathology laboratories, the potential impact on the diagnostic process and clinical correlations is extraordinary and clinical interpretation will be challenging. We review the rapidly evolving technologies; provide application examples; discuss aspects of clinical utility, ethics, and consent; and address the analytic, postanalytic, and professional implications.

The introduction of high-throughput, next-generation sequencing (NGS) in 2005 heralded a critical and transformative step in the history of DNA sequencing.[1] NGS has been broadly adopted by the biomedical research community as the method of choice for complex genomic analyses, and NGS-based diagnostic testing for several inherited disorders is available in a handful of clinical laboratories. Before describing the conceptual essence and utility of NGS, a brief reflection on the origins and growth of sequencing is warranted. Two seminal articles, describing the first methods for DNA sequencing, were published in 1977. One was by Allan Maxam and Walter Gilbert,[2] who described an approach by which radioactively end-labeled DNA fragments were subjected to base-specific chemical cleavage, reaction products were separated by gel electrophoresis, and subsequent visualization was accomplished using film autoradiography. In a second approach, Frederick Sanger and colleagues[3] reported the use of chain-terminating dideoxynucleotides that caused base-specific termination of primed DNA synthesis. Although both approaches yielded DNA sequences, the method developed by Sanger's group ultimately was more robust and amenable to technical innovations. These first approaches included supplanting radioactivity by fluorescently labeled primers, followed by the development of fluorescently labeled dideoxynucleotide chain terminators.[4, 5] In conjunction, the conversion from slab gel to capillary electrophoresis (CE) for resolving chain termination products provided an additional foundational element that accelerated commercialization and dissemination of Sanger sequencing and resulted in its gold standard status.[6] The deployment of Sanger sequencing in a factory-style configuration culminated, in 2003, with the sequencing of the first human genome under the auspices of the parallel private efforts by Craig Ventor and the public Human Genome Project at an estimated cost of $2.7 billion.[7, 8] This momentous accomplishment and historic milestone must be interpreted in the context of our current sequencing capacity: with current NGS instrumentation, a human genome can be sequenced in days for a reagent cost that may soon be comparable to some of the more conventional

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

molecular diagnostic assays. This precipitous decrease in cost and time to sequence at the genome scale level has been technically realized by the convergence of innovations in the chemistry, optics, fluidics, computational hardware, and bioinformatics solutions that underlie NGS. Although NGS platforms differ in design and specific chemistries, they are fundamentally related by a paradigm in which sequencing of spatially separated, clonally amplified DNA templates or single DNA molecules is performed in a massively parallel manner. A key feature is the sequencing of each clonal template or single molecule, permitting enumeration of individual sequences among all of the sequences generated. The ability of NGS to generate high-throughput qualitative and quantitative sequence information has enabled investigations that were previously technically infeasible or cost prohibitive. The diverse applications using NGS are extensive and growing and include, among others, genome, transcriptome, and methylome sequencing, metagenomics, characterization of protein-nucleic acid interactions, and targeted resequencing of multiple genes or genomic regions identified during linkage and genomewide association studies (GWAS).

The power of NGS to generate hundreds of millions to multigigabase levels of sequence in a single instrument run, while having opened a diversity of research and diagnostic avenues, is concomitantly stretching our ability to process data. This unprecedented amount of sequencing information poses bottlenecks that vary, depending on application, at the level of data extraction, analysis, and interpretation. These challenges have become part and parcel of the biomedical research community where investigators have increasingly needed to incorporate bioinformatics and biostatistics into their armamentarium. In a similar manner, clinical laboratories developing and adopting NGS need to build bioinformatics infrastructures composed of expertise and computational hardware. Indeed, the balance of time and effort required for NGS-based research or diagnostics is substantially shifted toward data analysis, as opposed to the technical component required to generate the data.

We can anticipate for the foreseeable future that the current NGS paradigm will continue to evolve with improvements in performance, accuracy, and instrumentation options that will further facilitate clinical translation. Also emerging in this rapidly evolving field are physical methods for sequencing based on biological or inorganic nanopore methods. In the context of this dynamic evolution, it is timely to review and highlight the opportunities and challenges that are experienced by clinical laboratories as they incorporate and leverage NGS for an ever-increasing array of complex genomic analyses. Although some laboratories are developing or using NGS for targeted resequencing of multiple genes, there are also pioneering laboratories that are performing clinical diagnostic exome and genome sequencing. This report describes current technical, bioinformatic, and clinical implementation considerations, as well as medical applications, clinical utility, and ethical, legal, and education issues presented by genome- level diagnostic testing.

NGS Technology

The NGS technologies are characterized by impressive throughput; all offer the ability to simultaneously sequence thousands to millions of relatively short nucleic acid sequences in parallel. The appeal is that they can provide orders of magnitude more information, at

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

competitive costs, when large regions of the genome are sequenced. For clinical applications, this becomes attractive because individual sequencing assays based on CE tend to be expensive, laborious, and less comprehensive, often necessitating serial gene-by-gene testing to identify the causative mutation(s).

Although the various NGS technologies use different types of chemistries, some of the platforms in use today bear several similarities to Sanger-based CE. For example, both methods begin with extracted DNA that is amplified. For CE methods, however, the PCR generally occurs in a 96- (or a 384-) well microtiter plate or tubes and is aimed at amplifying a limited number of regions. NGS methods take advantage of the high-throughput capabilities by targeting many more regions or a whole genome. Targeted sequencing also can be accomplished through the use of multiplex (Fluidigm) or droplet (Raindance) PCR or through pull-down methods in which probes bind to specific regions, which are then amplified and sequenced. For genome sequencing, the genomic DNA (gDNA) is randomly sheared and fragments are size selected by gel purification or solid-phase reversible immobilization bead fractionation. Adaptors are attached to the targeted primers and incorporated during PCR amplification or ligated to the DNA fragments themselves. These adaptors are then physically bound in space to beads, balls, or glass flow cells that have lawns of complementary adaptor oligonucleotides. After a PCR-based amplification, the individual clusters of amplified fragments are simultaneously sequenced. These fragments are typically shorter than usual CE fragments, although length capabilities continue to increase. Although the various NGS technologies use different types of chemistries, in most cases, nucleotides used for the sequencing reaction are fluorescently labeled and excited by a laser, as with CE. A digital image is then captured for data analysis. However, unlike CE, which uses dye-terminated sequencing, the same labeled fragment can typically have the fluorescent dye and blocker removed so that the next nucleotide can be added. Because there are thousands to millions of these reactions being performed next to each other, this is referred to as massively parallel sequencing. Platforms that have different chemistries or methods include those that use native, fluorescently unlabeled nucleotides and those in which sequence data are obtained by either pyrophosphate-dependent chemiluminescence signal detection by a charge-coupled device camera or direct detection of ions produced by template-directed synthesis on a massively parallel nonoptical semiconductor-sensing device (ion chip). Some platforms skip the amplification step altogether and perform single-molecule sequencing in real time.

Different NGS technologies and their technological and performance differences can be reviewed in articles by Shendure and Ji,[9] Voelkerding et al,[10] Anderson and Schrijver,[11] Metzker,[12] and Rothberg et al.[13] However, technology development occurs at a fast pace and additional platforms are expected to emerge and add to the options for application in clinical diagnostic laboratories.

Application Examples

Inherited Conditions

Monogenic conditions are caused by mutations that are infrequent across the population. Before the availability of a human genome reference sequence, disease genes were sought

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

by linkage analysis of large families with a specific disease using markers mapped across the human chromosomes. This approach proved to be highly successful for discovering monogenic conditions[14–17] but not for complex disorders because of the difference in strength of genotype-phenotype associations (effect size). Being less direct than NGS, linkage analysis was limited by the need for large pedigrees, densely mapped markers, and clear clinical records, as well as by complexities such as incomplete penetrance and modifier genes. Triad analysis by sequencing the patient and parental genomes or exomes using NGS will, in many cases, be able to uncover the underlying pathogenic mutation(s).

Since the completion of the Human Genome Project, there has been a steady increase in the number of singlegene disorders that have been elucidated. Some efforts have focused on disease gene discovery, which inform development of an appropriate clinical assay for the condition. The use of NGS as a discovery tool can be illustrated by the identification of the *MLL* gene as a major cause for autosomal-dominant Kabuki syndrome.[18] In other cases, a condition may be caused by any of a large number of genes, and NGS approaches are able to quickly hone in on the responsible gene, compared with the one-by-one approach of traditional sequencing. In such cases, the testing is less of a gene discovery effort; rather, it is focused on finding the proverbial needle in the haystack of previously identified genes.[19] Genome sequencing is already being used as a clinical test for patients with unidentified monogenic conditions in several clinical centers, and reimbursement has been provided in at least one such instance. Until costs further decrease, only those patients who are most likely to benefit from genome or exome sequencing will typically be offered this testing. Diagnostic challenges remain, however, especially when monogenic disorders arise because of novel missense mutations, structural aberrations, germ-line or embryonic mosaicism, or imprinting and other epigenetic factors. A major application of NGS in testing for inherited disorders is in conditions such as congenital disorders of glycosylation, hypertrophic cardiomyopathy, developmental delay, neuromuscular disorders, retinitis pigmentosa, and seizure disorders in which the overlapping symptoms of multiple possible syndromes and many potentially affected genes can make an NGS approach cost-effective.

In contrast to the mutations in monogenic conditions, risk loci for common, multifactorial diseases are more frequently located in noncoding or intergene regions, and the pathogenic mechanism of these variations is not well understood.[20] The availability of a reference human genome sequence led to cataloging of variants through the International HAPMAP Project (http://hapmap.ncbi.nlm.nih.gov, last accessed August 24, 2012), which identified the linkage disequilibrium of 3.5 million single-nucleotide polymorphisms (SNPs) across several populations. SNPs and other variants are listed in the dbSNP Short Genetic Variations database (http://www.ncbi.nlm.nih.gov/projects/SNP, last accessed February 29, 2012), which contains 11 million SNPs and three million indels. The extensive characterization of SNPs, repeats, and linkage disequilibriums across chromosomes has enabled association studies for the identification of disease genes in complex common diseases, for which a population-based, rather than a family-based, approach must be used to elucidate genotype-phenotype correlations.[21, 22] In the past decade, research on common complex diseases would focus on genotyping of a few candidate genes that were hypothesized to be involved in the disease using a casecontrol approach. There were a few success stories, most notably the identification of apolipoprotein-ε4 allele association in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Alzheimer disease, but most results could not be reproduced in subsequent analyses or in different populations.[23, 24] The technological advance of dense SNP chips has facilitated GWAS with many cases and controls (>1000 per group), and associations with candidate disease genes can be confirmed with subsequent GWAS.[20, 22, 25] Before the emergence of NGS technologies, multifactorial conditions were part of the diagnostic test spectrum offered, but to a much lesser extent than was the case for monogenic conditions. Given the complexity of the clinical interpretation, many such conditions were studied only on a research basis, rather than offered in a clinical molecular diagnostic setting. Examples of routine application, however, include the testing of sequence changes known to increase the risk of deep vein thrombosis. Deep vein thrombosis is thought to have a multifactorial etiology and is influenced by factors such as ethnicity, sequence changes that confer genetic risk, and socioeconomic differences.[26]

The Web site genome.gov[22] catalogues >1200 polymorphisms that confer disease risk in >165 common human diseases and traits. However, these variants typically confer low relative risk and predictive power. Even when multiple risk markers are combined, the cumulative risk scores are often comparable to traditional risk factors, such as family history and individual clinical profiles (eg, cholesterol levels, age, sex, and obesity).[27, 28] These markers were ascertained using population-based associations that may only apply to limited populations and may lack clear biological meaning for any given individual. Moreover, in almost all common diseases, multiple risk alleles only explain approximately 5% of the variance in the population, even when the disease is considered highly heritable, leading to the concept of missing heritability.[29] The gene expression of these multiple risk alleles with low predictive power may be influenced by diet, environment, lifestyle, and epigenetic factors. Nevertheless, even in patients with common diseases, NGS will permit the detection of rare variants that would not be identified with SNP arrays, which represent only the common variants used in GWAS or with limited genotyping.

In the immediate future, the uncovering of three million SNPs within an individual's genome is only as useful as our knowledge of what these SNPs signify. The full potential of a DNA-based understanding of the common complex diseases so prominent in the practice of medicine can only be realized when the genome sequences of thousands of individuals are available and linked to corresponding phenotypes. Ultimately, data mining studies, such as the U.K. 10,000 Genomes Project (Wellcome Trust Sanger Institute; http://www.sanger.ac.uk/about/press/2010/100624-uk10k.html, last accessed February 29, 2012), may lead to a refinement of disease pathways and a better understanding of how specific alleles (rare or common) present themselves as protective in health and deleterious in disease.

Cancer: Diseases of Acquired, Somatic, and Genetic Changes

During the past few years, knowledge of the molecular mechanisms of cancer has markedly expanded as a result of the identification of specific driver mutations and other somatic changes, consequent alterations of gene expression, and the discovery of the role of microRNAs in carcinogenesis.[30] As a consequence, the traditional histological and immunohistochemical characterizations are enhanced with respect to the classification of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

cancer subtypes and by predictions of therapeutic response, prognosis, and patient survival. Understanding the importance of cancer target gene analysis has led to considerable advances in drug therapies and has made possible more personalized cancer treatment. Moreover, the insight that the molecular profile of cancers is heterogeneous, even within a single patient, may change during the course of a patient's disease, and affects the acquired resistance to drugs, has transformed the field. Multiple markers for specific cancers are being used to stratify patients in clinical trials of new investigational drugs, which are intended to be used for targeted inhibition of cell proliferation.[31, 32] Response to treatment and relapse are also being monitored in serial samples from the patient through the detection of tumor-specific chromosomal rearrangements and other genetic abnormalities. Investigation of germ-line or constitutional cancer predisposition markers that increase an individual's risk for developing cancer has contributed to screening and patient management.[33, 34] The applications of NGS in oncology have ranged from mutation panel screening in gene pathways to exome or transcriptome sequencing and ultimately genome sequencing, with comparison to the patient's normal tissue or other reference samples. [35–37] The concomitant availability of a complete reference sequence of the human genome has helped spur massive coordinated initiatives, such as the Cancer Commons (http://cancercommons.org, last accessed August 24, 2012), the International Cancer Genome Consortium (http://www.icgc.org, last accessed February 29, 2012), and the Cancer Genome Atlas (International Cancer Genome Consortium; http://cancergenome.nih.gov, last accessed February 29, 2012).[38, 39] These efforts have led to extensive sequencing and cataloging of somatic mutations and gene expression analyses in thousands of genomes from different classes of solid and hematological cancers.[40] As the number of known cancer markers increases, it will likely be more cost-effective to profile a patient's tumor DNA using NGS than to conduct genetic analysis for a growing set of mutations or to assay gene expression profiles from fresh frozen or formalin-fixed tissues. The quantitative nature of NGS technology is another improvement over chip-based gene expression technologies. Recent use of gene expression analysis for prognostic algorithms, such as Mammaprint (70 genes), OncotypeDX (21 genes), and Rotterdam signature (76 genes) will likely be replaced with next-generation sequencing of signature transcripts. [41–43] Such information may be incorporated in novel algorithms that use a patient's somatic mutation profile in addition to cancer susceptibility gene findings. NGS will be transformative for diseases resulting from somatic genetic changes because it will allow speedy, cost-effective, and complete profiling of a patient's tumor and germ-line DNA, transcriptome, and DNA methylation profile in a single platform. However, it is important to appreciate that different NGS platforms have distinct capabilities of detecting DNA or RNA sequence changes, DNA copy number changes, DNA rearrangements, RNA expression changes, and DNA methylation. The specific NGS method determines which combination of these changes can be detected in either germ-line or tumor DNA. One method will usually provide only part of this variety of information. Therefore, a decision has to be made and this may be based on cost, specimen availability, and specimen quality, as well as the specific questions to be answered.

Because of the significant false-positive rate associated with NGS,[44] mutation confirmation in both inherited diseases and neoplastic conditions is typically performed using Sanger sequencing. However, confirmation on a different NGS platform may also be considered. In

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the future, independent confirmation may become increasingly redundant, considering further evolution of the technology will undoubtedly facilitate a more complete characterization of mutations. Compared with germ-line sequencing, however, there are several specific challenges regarding somatic mutations. The percentage of tumor cells within a sample varies, and the percentage of tumor tissue with a mutation may vary, especially if it is a (presumably) secondary mutation. A low percentage of a secondary mutation in a tumor may overlap with the false-positive rate identified by NGS. Low-level mutations, however, cannot be easily confirmed by Sanger sequencing, which has a higher threshold of detection (approximately 15% to 20%) than NGS, without using a more sensitive screening technique (eg, denaturing gradient gel electrophoresis or denaturing high-performance liquid chromatography), an alternative approach (eg, pyrosequencing), or co-amplification at lower denaturation temperature–PCR to enrich for the mutation.[45–48] Another challenge is the remarkable cancer genome complexity reflected by the many different aberrations detected in some tumors. Such heterogeneity within and between tumors adds complexity to analytical validation and clinical confirmation of identified mutations and personalized therapeutic approaches. Thus, approaches and algorithms different from those used in germ-line testing will need to be developed for the applications in oncology.

Clinical Utility

Aided by the rapidly decreasing cost of data generation, large-scale personal genome sequencing will be applied to a variety of settings and applications. With proper interpretation, NGS will permit a deeper understanding of disease mechanisms, allowing for more evidence-based medical interventions. It will influence, to a greater degree than possible with conventional molecular testing, medical decisions, including personalized treatment of disease, monitoring of disease progression, and assessments of disease risk, prognosis, and reproductive matters; it is likely to affect the way in which individuals approach lifestyle questions.[49] Although much of this is likely to remain in the medical realm in which physicians guide their patients to the testing best suited for their medical management and to the appropriate interpretation, some genome testing will inevitably be offered directly to consumers.[50] In such a setting, consumers will decide independently whether they want to learn about topics ranging from ancestry to disease risks. Thus, NGS will be relevant to both patients and general consumers.

The personalized nature of testing at the genome level includes a subjective assessment of the overall value of the generated information. Apart from personal utility, which could be defined as the perceived usefulness for an individual given his or her interests (whether these are medical or not), there is the issue of clinical utility. Clinical utility is a measure of the net health benefits, reflected by the balance of benefit versus harm. For an accurate evaluation of clinical utility, factors such as test rationale, patient population, and clinical scenarios have to be considered. In addition, the principles of comparative effectiveness should be followed. This requires an individualized evidence-based approach for each patient, ultimately accomplishing the separation of information with demonstrated medical utility from information with less clear or missing utility evidence.[51, 52] Some questions that are explored through NGS can be answered with a high level of confidence, whereas others are

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

complex and require additional research. One of the complexities is the assignment of pathogenicity of variants. Even under the straightforward assumption of a singlegene autosomal-recessive condition approached with Sanger sequencing, the relationship of sequence variants to disease causation can be difficult to ascertain with certainty. Given the data density of NGS, the issue of variants of unknown significance is encountered many times over, and our understanding of the biological, pathophysiological, and functional impact of certain genomic regions and sequence changes is still limited. Thus, the development of evidence-based, validated scientific standards to evaluate the clinical utility of genomic results in different populations with accurate genotypebased risk estimation is a considerable challenge. Furthermore, there is a possibility of potential overinterpretation of results based on a limited understanding of contextual information that could lead to unnecessary medical action and cause unwarranted psychological distress.[53] Careful selection of patients for genome sequencing and concomitant genetic counseling are crucial.[54]

Informed Consent and Ethical Considerations

Informed consent is important to make the patient or client aware of the benefits and harms that may occur as a consequence of the testing process. For example, incidental findings may suggest that the patient is at risk for a disease that was not suspected. The capacity to perform large-scale sequencing on the human genome presents unique challenges regarding the provision of informed consent, particularly in deciding on the level of detail that needs to be shared. No specific guidance exists, and each institution offering such testing is deriving its own policies. Recently, concerns about genetic discrimination have been focused mostly on predictive testing for defined single-gene conditions, such as Huntington disease and familial cancer syndromes. Genetic nondiscrimination laws have been passed at both the state and federal levels, offering some protection to individuals who carry deleterious sequence changes, in an effort to alleviate concerns about potential abuse and discrimination and to reduce societal obstacles to genetic testing.

With genome testing by NGS, the perceived and real potential risks are magnified compared with genetic tests that target only one gene at a time.[55] A patient who undergoes genome sequencing will likely learn about sequence variants of unknown clinical significance or about other sequence changes that indicate disease risk unrelated to the disease for which the testing was requested. The potential for anxiety and uncertainty, resulting from results of unknown clinical significance, is associated with any NGS method, as well as with less comprehensive techniques, such as high-density oligonucleotide microarrays.[56] With NGS, however, the entire sequence is read, and those changes that are not clinically understood cannot be blocked out. One question that is raised with any large-scale sequencing of the human genome is the consideration for what to interpret and report. For well-defined cases, such as the search for a therapeutic target in a set of known cancer genes, the levels of analysis and reporting are reasonably clear. In diagnosing a rare disease through genome sequencing, the presence of incidental findings becomes a more significant topic. These issues remain under discussion, but their solutions will no doubt involve an emphasis on counseling and education before testing is performed, informed consent with a clear explanation of the current limits of testing and interpretation, maintenance of privacy and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

confidentiality, and sensitivity to culture within families, their heritage, and their communities. As long as a balance is struck between the promise and potential of NGS testing and its current limitations, it can transition from the research realm to the clinical setting, similar to the many molecular genetic innovations that have preceded it.

Analytical NGS Considerations

Regulation, Assay Validation, and Reference Materials

Before making any new laboratory-developed test available for patient care, a clinical laboratory undertakes the process of analytical validation. For Food and Drug Administration– approved/cleared tests, the similar, but more limited, required evaluation is termed verification. However, Food and Drug Administration–approved/ cleared tests using NGS do not exist. In fact, there is considerable uncertainty about the regulatory pathway for NGS testing and resolution could take years. Thus, laboratories will, for the foreseeable future, continue to rely on regulations under the Clinical Laboratory Improvement Amendment. The purpose of assay validation for a laboratory-developed test is to document that the targeted analyte(s) can be detected in a robust and consistent manner. In accordance with Clinical Laboratory Improvement Amendment regulations [Code of Federal Regulations §493.1253(b)(2)], the following analytical characteristics must be documented for this purpose: i) accuracy, ii) precision, iii) analytical sensitivity, iv) analytical specificity, v) reportable range, vi) reference intervals (normal values), and vii) any other performance characteristic required for test performance (eg, carryover, dilutions, and calculations).

NGS ranges from a relatively limited scope to genome sequencing. With exome sequencing, the coding regions of the genome are interrogated, whereas with more comprehensive approaches to genome sequencing, intervening regions are also included. In either case, however, the sequences are not truly complete because of, for example, gaps in the obtained sequences, issues with GC-rich regions, and bioinformatic limitations regarding the calling of indel variants. These challenges can be brought to light by a comparison with Sanger sequencing, but NGS presents a challenge to the notion of proving and documenting that the tested genome sequences have been identified correctly. When validating a new platform, conventional wisdom requires the laboratory to compare the new approach with a gold standard. For nucleic acid sequencing, however, that standard is the traditional method of Sanger sequencing, the technical capabilities of which are dwarfed by NGS. This recognition presents obvious dilemmas. It is practically not feasible for a laboratory to perform Sanger sequencing of whole genomes for reasons of reagent costs, labor, and efficiency of both testing and analysis. Nevertheless, all steps of NGS need to be evaluated, including sample library preparation, clonal fragment amplification, the sequencing itself, and data analysis. For each of these steps, quality control metrics need to be determined. The question then arises whether the sequencing and documentation of representative portions of a reference genome would be satisfactory, or whether comparison to online reference sequences could suffice. In addition, it needs to be established what reference materials (RMs) could be developed to make the validation process both efficient and meaningful. An RM (commonly referred to as a control sample) is a material or substance, one or more of whose properties and/or values are sufficiently homogeneous and well established to be used

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

for the calibration of a measuring system, for the assessment of a measurement procedure, or for assigning values to materials (International Organization for Standardization 15195).[57] RMs are tested alongside patient samples and allow laboratories to detect errors due to test system failure or operator error. In addition, RMs are needed for test development and validation, lot testing of new reagent batches, and proficiency testing and/or external quality assessment programs. There is a lack of RMs for NGS-based tests. The development of RMs for NGS of panels of genes, exomes, or genomes will require a new paradigm because it will no longer suffice to develop or use a limited set of characterized gDNA samples that represent most or all mutations in a single gene that commonly cause a particular disorder.

There are several efforts underway to develop guidance for laboratories regarding the application of NGS in the clinical setting. The Division of Laboratory Science and Standards at the Centers for Disease Control and Prevention is working with stakeholders to develop useful metrics to help define analytic and clinical performance specifications of NGS assays. This project is addressing test validation, quality control, and proficiency testing to derive a set of principles and guidelines for the use of NGS in the diagnosis of heritable conditions. The selection and use of appropriate reference materials is central to all these themes and, therefore, the Centers for Disease Control and Prevention's Genetic Testing Reference Material Coordination Program (GeT-RM; http://wwwn.cdc.gov/dls/genetics/rmmaterials/default.aspx, last accessed February 29, 2012) is engaging to generate renewable and publicly available characterized gDNA RMs that can be used for clinical NGS testing. This need is urgent because the pace of technology development in NGS is so fast that conventional analytical validation methods may not be feasible or realistic. It is envisioned that gDNA from publicly available cell lines will be sequenced on multiple NGS platforms, after which the data generated will be assembled into a consensus genotype for each cell line. This will be posted in a publicly accessible database and updated to reflect corrections to the original sequence, should these be necessary. All Genetic Testing Reference Material Coordination Program efforts, including decisions about RM priorities, specimen collection, cell line development, and mutation characterization studies, occur through voluntary collaborations with clinical laboratories and RM providers. The Clinical Laboratory Standards Institute is in the process of revising its guideline, *MM9: Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine*.[58] This document will be substantially revised to address the use of NGS in the clinical laboratory environment. The American College of Medical Genetics has also established working groups to develop recommendations relevant to the clinical use of NGS. Finally, the College of American Pathologists is developing checklist questions specific to NGS that will be used in its accreditation program. It is expected that the principles and recommendations developed from each of these groups will be in alignment, where there is overlap, owing to an open information exchange and many of the participants contributing to multiple efforts.

Bioinformatics Requirements

Only three major computations are performed with NGS data: i) data assembly with base calling at the level of individual reads, ii) alignment of the assembled sequence to a reference sequence, and iii) and variant calling. Within variant calling, there may be different computational analyses devoted to SNPs, small indels, structural variants or large

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

indels, and copy number variants. Despite the few major computations, NGS experiments generate an unprecedented amount of medical data that result in special informatics needs and require tools for data management, storage, analysis, and archiving, both to manage the large data set without error and to ensure proper quality and documentation. Considering these steps in greater detail, they can be delineated as follows: i) image processing, ii) base calling and quality scoring, iii) purity filtering, iv) quality control checks, v) alignment to a reference sequence or *de novo* assembly, vi) SNP-finding or application-specific steps (eg, chromatin immunoprecipitation or RNA sequencing), vii) packaging, and viii) archiving.

Typically, analysis begins with a large set of tiled fluorescence or luminescence images of a flow-cell surface, which are recorded after each iterative round of sequencing. The next step is the most computationally intensive part of the basic bioinformatics pipeline and comprises the conversion of the image data into sequence reads, termed base calling. During base calling, individual beads or clusters are identified and localized in an image series and parameters (eg, intensity, background, and noise) are then used by platform-specific algorithms to generate read sequences and quality scores for each base.[12] There is a continuing need to reduce error rates, especially as platforms are pushed to generate longer reads. As a result of the base calling process, FASTQ files are generated. FASTQ files are usually platform specific and contain sequencing read data that combine both the sequence and an associated per base quality score. The sequence reads then are aligned to a known reference sequence or assembled *de novo*. Mapping many reads to the reference genome requires highly efficient and accurate algorithms. The appropriate alignment method will depend on the sequencing platform, data type, and computational resources.

One limitation of short–reads alignment and assembly is the inability to uniquely align much of a read set. Similarly, the number of reads is reduced when aligning to larger, more complex genomes or reference sequences because of the higher probability of repetitive sequences.[59] Interestingly, error rates for individual NGS reads are higher than for Sanger sequencing, which is a reflection of the difference in methods. Even among NGS platforms in use, error profiles vary. Nevertheless, the accuracy of NGS is achieved by sequencing a given region numerous times, which is facilitated by the massively parallel process. Assembly, alignment, and analysis of NGS data require an adequate number of overlapping reads for each nucleotide, or coverage depth. A consensus sequence is derived through this process. In practice, however, coverage across a sequenced region is variable, and factors (eg, Poisson-like randomness and amplification bias, differential ligation of adapters to template sequences, and disparate amplification during clonal template generation) may contribute to the observed variability.[60] Beyond sequence errors, inadequate coverage can cause failure to accurately detect nucleotide variation, leading to false-positive and false-negative results. Depending on the sequencing application, the next step may involve quantification of reads, mutation calling, copy number analysis, and/or analysis of structural variations.

The NGS data deluge presents opportunities and hurdles for data analysis and management. One current obstacle toward clinical translation is that most algorithms in use need some programming expertise, together with specialized servers to handle and store all of the data. The generation of user-friendly informatics tools to effectively analyze NGS data will be

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

essential to the successful clinical application of NGS technology. Today, most users align the NGS reads to a known reference, such as the most current build of the genome, using SNP databases to filter out alterations that are known germ-line polymorphisms. In the analysis of mendelian conditions, use of unaffected family members provides a useful background in which to identify novel disease- causing mutations. However, the calling of somatic mutations in malignancies is much more complicated because of the variance in ploidy and purity of the tumor DNA, and is strongly dependent on the allelic fraction. However, comparison to nontumor tissue from the same patient provides an important tool for identification of somatic mutations, although differentiation of cancer-causing from passenger mutations remains challenging. Elimination of false positives and negatives is critical for both inherited and acquired diseases; however, although elimination of false positives can largely be achieved through verification by conventional PCR and Sanger sequencing, minimizing false negatives is less straightforward. And, even though the challenges in assigning reliable mutation calls, insertions/deletions, and structural variations are considerable, the main challenge ultimately may well lie in distinguishing clinically relevant mutations from benign variants. With respect to data storage, typically tens or hundreds of giga bp of short reads can be generated during one NGS run. As a result, the average NGS experiment generates terabytes of raw data, requiring a vast data storage and management solution. The image files, in fact, require the most storage capacity, and many researchers opt to delete these files once the base calling has been completed. Thus, only the FASTQ or just the postalignment files are stored. Clinical laboratories are required to maintain certain data from the tests that are performed to verify their test results, but the extent of data that need to be retained from NGS assays has not been established. The challenge to laboratories is that the retention of the complete data is costly and, in many instances, cost prohibitive. Possibly, one way of overcoming the need to obtain expensive servers may be the use of cloud computing.[54]

From the perspective of the electronic health record (EHR), NGS test results reporting raises several issues. Still, it is important to clarify that several aspects of genetic or genomic testing, regardless of the method, can be supported by EHRs that provide the architectural framework. These aspects include the ordering of the test, the receiving of a document that summarizes the clinical interpretation, and storage of the interpretation.[61] Although there are many potential ways to integrate NGS data into the EHR, three issues seem to consistently emerge within this context. First, there is debate about whether the EHR should store the large, raw data files, given that the EHR is the legally binding medical record. In general, the information in the EHR has been interpreted by the appropriate diagnostic expert. Similar storage issues arise when considering magnetic resonance images; typically, these images are stored in archives that are linked to, but separate from, the EHR. Second, it is questioned to what extent the EHR should assist diagnosticians in interpreting NGS results. The software involved in interpreting raw data (eg, calling bp) is generally tightly coupled to the instrumentation platform, not the EHR. A variety of software capabilities can be used to compare a patient-generated sequence with a reference sequence and to identify variations with a likely functional impact. Current and future variant databases may help the diagnostician determine whether a variant is novel. The capability of linkage to such variant databases is the most probable extent of direct access in the context of the EHR. Finally,

there is discussion of the role of the EHR in sharing new knowledge when it becomes available after the formal interpretative report is completed. The EHR is intended to enable compliance with the operational regimen of the end users, and each organization using an EHR is obligated to reach its own conclusions about potential incorporation of new knowledge. The technology required to enable dynamic interpretation of NGS results is feasible but will take significant development before it is ready for implementation. Perhaps the most likely context in which dynamic interpretation will occur is outside the context of the EHR in the less-regulated framework of the personal health record. Already, some direct-to-consumer companies regularly adjust their content based on newly emerging knowledge, although these reports are not considered to fall within the legally binding diagnostic model. In the clinical laboratory, it also may actually be more economically feasible to resequence (subsets of) genomes than to regularly update the EHR or to store large data files indefinitely, which would obviate some of these bioinformatics concerns. Resequencing also allows for the benefits from NGS technology improvements, which may enhance the data generated, such as the extent of the genome sequenced or the depth of coverage for each nucleotide position.

Bioinformatics Quality Assessment

Similar to CE-based methods, the quality of NGS images is digitally evaluated using components such as signal strength relative to background, detected signals and their ratios, signal strength relative to neighboring calls, and other similar metrics. Based on a combination of these parameters and image-processing adjustments (eg, dye cross talk corrections and/or phasing corrections), the particular image is assigned a base call and a quality score that indicate the likelihood that the base call is correct. In CE analyses, this is typically called a Phred score and is generated on a logarithmic scale ($Q = -10 \log E$, and E = probability of error), so that a Phred of 10 means a 1/10 probability of error, whereas a Phred of 20 indicates a 1/100 probability that the call would be in error. Thus, the higher the score, the more confidence can be placed in the call. An obvious way to evaluate scoring accuracy is to take a known reference sequence and to sequence it using NGS technology at great depth. In the course of sequencing, the same positions will be called at various Phred (or Q) scores in multiple, independent sampling events, after which the data can be plotted so that the score assigned to any call based on the nucleotide calling software can be compared with what was actually observed based on the known call at that position. Once confident that a score reflects the quality of the data accurately, thresholds can be determined and data can be filtered accordingly. For sequencing information, the Clinical Laboratory Standards Institute has recommended filtering on a Phred score of 20 to 30,[58] resulting in an error rate of 1/100 to 1/1000. When interrogating variants in a population-level situation, such as microbes or tumor samples, higher thresholds may be necessary.

Each platform is characterized by its own error profile. Some platforms have insertion or deletion errors during homopolymer runs, whereas others have an increased likelihood of sequence errors toward the end of the read.[9–13] Data production and output also differentiate platforms. An example of this is color space (a color represents a class of dinucleotides, such as AA, CC, GG, or TT), as opposed to base space (a series of bases represented by A, C, G,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and T calls). These platform differences must be addressed at the algorithmic level. When analysis tools take these aspects into account, some sequence errors are correctable.[62]

Once a subset of individual calls has passed the base-calling filter, the specific sequence fragments must be assigned to the reference genome (in the case of whole genome sequencing, this is typically NCBI 37). The reference genome in National Center for Biotechnology Information is derived from a few individuals and is a small sampling of human genetic variation. It also contains rare and common disease risk variants complicating the detection of these risk alleles. There are efforts to build major allele reference sequences that should be considered for accurate ethnically concordant variant calling.[63] For large data sets, such as whole genome reconstruction, speed becomes critical, and a tiered approach to sequence alignment is taken. First, the sequences that align based on some subset of the total read are assigned a quality score based on how well they align to the reference, taking into account such components as the number of mismatches relative to the reference. The expectation is that most sequences will have only a limited amount of variation relative to the reference sequence. Second, for the subset of samples that were not successfully placed using the first round of alignment criteria, a multispeed gapped alignment approach is taken. For this subset of sequences, gaps are introduced to evaluate whether this will allow for improved alignment and assignment of the sequence; if on introducing a gap somewhere in the sequence both regions align with high fidelity to the reference, the gap is retained. Details of this process differ among alignment algorithms. Therefore, it is important to consider the appropriate performance parameters and the user manuals when choosing alignment algorithms. After this step, probability values are assigned based on all possible alignment solutions, and the best alignment is kept and assigned a score based on the Q scoring scale. Various alignment algorithms exist with various penalties or thresholds for the fidelity of sequence alignment to the reference. Also, algorithms vary to their effectiveness for detecting specific types of sequence variations, such as SNPs and indels. This may require use of more than one aligner to derive optimal results when calling sequence variants. For clinical applications, conservative approaches are most appropriate because high confidence should exist when making a call.

When all of the sequences that passed the initial base calling and sequencing filters have been aligned and stacked, a consensus call is made. This step represents a significant departure from CE sequencing methods, because in those cases, a PCR is sequenced and both alleles are typically imaged simultaneously. Although this can cause a heterozygous signal to appear with diminished quality (and, at times, the calling software is unable to call these positions without human intervention), in NGS methods, each molecule is sequenced independently and, therefore, there is not a mixed signal. However, because of the fact that each molecule is sequenced independently, the probability of detecting a second allele in a heterozygous position is dictated by the sampling probability:

$$\mathrm{P(x, p, N)} = \sum_{K = x}^{N - X} \frac{\mathrm{N!}}{\mathrm{(X!)(N - X)!}} \mathrm{p^{X} q^{(N - X)}}$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript

In essence, the sampling question being asked is the one encountered in school: if given a bag of marbles and told that the bag contains either all blue marbles or half blue and half yellow, then how many marbles do you need to draw to have 99% confidence that you would detect both colors if they were both present? In this case, we assume that, in a heterozygote, approximately half the alleles should be one nucleotide and half the other. However, the second allele should be observed more than once; thus, if, for example, second allele needs to be present in at least 20% of the calls to make a heterozygous consensus call, then by applying these numbers into the previously described formula, that position needed to be sampled at least 18 times to be at least 99% confident of detecting both alleles, if present. K establishes the lower bound of the interval over which the probabilities need to be summed. In this case, we set this as 20% of the total, so if we sampled 10 times, then we would need to see each allele at least twice and not more than eight times, so $k = 2$ and we would sum the probabilities of observing two, three, four, five, six, seven, and eight times. The consensus call is typically made based on a quality score that represents a combination of the following: i) the number of times that the position has been sampled in independent sequencing events, ii) the quality of each of the base calls made, and iii) the quality of the alignment. However, again, the score assigned represents an opportunity to apply a quality filter so that only calls with high raw data quality, high alignment confidence, and a high level of confidence in the variant call will be kept for further analysis and interpretation.

The number of reads that align to a given reference sequence region is referred to as coverage. Operationally, the amount of coverage obtained is dependent on the amount of sequencing performed versus the target size. The human haploid genome is approximately three billion bases, and most investigators aim for 30- to 40-fold average coverage because the cost of sequencing remains a limiting factor. Depending on the capture probe reagents used for exome enrichment, the target size varies from 30 to 50 Mb. Exome coverage will vary depending on the amount of sequencing, but typical average exon coverage would be 50- to 100-fold. Gene panel target sizes vary considerably, depending on the number of genes. Average coverage, however, is in the same range as for exon sequencing when analyzing for constitutional variants. In comparison, a higher average coverage (500-fold) is aimed for when attempting to detect low-frequency alleles in mixed population analyses, including somatic variants in tumor samples. Although average coverage is one metric to determine, it is essential to also define the range of coverage for targets being sequenced, to characterize the adequacy of sequencing for a given region. A balance exists between increasing coverage to improve variant detection and increasing sequencing costs. The lower coverage for genomes can lead to missed variants that are observed in exome sequencing results from the same individual. Conversely, modestly more exons are interrogated by genome compared with exome sequencing, because of the fact that some exons are not enriched by exome capture. Gene panel sequencing offers the ability to achieve high levels of coverage in a defined target set. Although an attractive approach, gene panels are limited in that adding new genes to the panel typically requires a panel redesign.

Once a position has been called according to the bioinformatics quality thresholds imposed, the clinical implications of the variants that were called need to be evaluated. This generally involves querying all of the calls made in the genome sequence against all known relevant databases that associate variants of clinical significance with the published reports. The first

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

challenge is mapping the clinical variants, which are often reported using Human Genome Organisation (International) nomenclature (which is gene centric),[64] to their genomic coordinates relative to the National Center for Biotechnology Information reference genome sequence. Once this has been accomplished, it is fairly straightforward to query all positions against the genome sequence and to report the calls made. When a definitive call carries a clinical implication, the typical interpretive approach used in clinical laboratories for traditional sequencing methods is appropriate. However, compared with targeted genetic testing, interpreting much of an entire genome poses a substantially larger challenge because many more variants will be detected. The lack of any database curated to accepted clinical standards likely presents the most significant challenge in managing and reporting genome sequencing data.

Postanalytical NGS Considerations

Practical Interpretation and Reporting

There is a paucity of standardized best practice guidelines or criteria for the technical and clinical implementation of genome sequencing. However, it is an extension of current genetic testing practices and, as such, can be realized more completely as our understanding of genomic variation becomes more advanced and practice guidelines that address the optimal interpretation (or combinations of) variants in an individual have been defined. Apart from genetic test modalities, genome sequencing requires the use of a sound biocomputational infrastructure. Toward this end, there is a need for standardized bioinformatics pipelines for refinement in base calling and annotation of the variants obtained for their likely functional consequences. Furthermore, high and consistent levels of coverage are required between available platforms to ensure accuracy in nucleotide base calling. Interpretations resulting from any WGA platform must account for the analytical limitations of the specific technology or platform being used. An additional layer of complexity is that genome sequencing must be integrated with complementary testing tools. An expert interpretation of the meaning associated with the vast amounts of data that result from these types of analyses will be essential. These interpretations may require consideration of ancillary factors, such as the patient's age, sex, clinical presentation, and family history. Other relevant variables that also must be considered include the following: i) the incomplete ascertainment of structural variations in the human genome, ii) lack of accurately curated haplotype data, iii) variable expression of the disease phenotype, iv) disease-associated penetrance, v) gene-environment interactions, vi) heritability, vii) identification of rare mutations of unknown significance, and viii8) genetic heterogeneity (locus and allelic).[65, 66] Therefore, the interpretation and reporting of genome sequencing results requires a team approach, whereby pathologists, geneticists, and other laboratory professionals become even more directly involved with the health care team to bridge the inevitable gap in knowledge between those closely involved with genomic data interpretation and the health care providers who would be hard pressed to keep up with the new developments but who need to integrate the information into their medical practice. These clinicians will require competent expert guidance in their patient communications and effective consultation to enhance their patient care. This will accelerate the need for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

development of new counseling paradigms that appropriately balance new information with the uncertainty about the meaning of a test result.[66]

Part of the challenge of providing expert interpretation of genome sequence data is the evolving nature of our understanding of the meaning of differences between individual genomes. Genotype-phenotype uncertainties continue to confront us in the setting of monogenic disorders, and making sense of combinations of sequence variants across genes is even more challenging because of the plethora of variations in biological effects, including those on protein function and interactions within and between pathways. A logical approach to this complex issue is to initially reduce genome sequence data to focus on a subset of sequence variants that are likely to be clinically significant. Even this approach, however, requires the accumulation of a large set of high-quality and ethnicity-specific reference genomes that can be used for genomic comparisons. It also will require the availability of clinically vetted, regularly updated databases of annotated variants that ideally would include population frequencies and referenced clinical relevance for each sequence variant. Thus, there is a need for consolidation of the various genotype-phenotype databases available into a commonly available and perhaps centralized clinical- grade resource that is publically accessible.

Proficiency Testing/Alternative Assessment

Proficiency testing and alternative assessment refer to a periodic laboratory evaluation of characterized specimens to determine assay variability and for the purpose of comparison to a standard measure, when one exists. Proficiency testing, but not necessarily alternative assessment, requires that specimens be provided and the evaluation administered by a third party for the purpose, in part, of providing an interlaboratory comparison. In the United States, laboratories are required to demonstrate proficiency through participation in proficiency testing programs or through otherwise establishing a means for alternative assessment.[67] Proficiency testing is a well-recognized component of an overall quality management system integral to regulatory oversight[68–70] (College of American Pathologists; http://www.cap.org, last accessed February 29, 2012).

Proficiency testing and alternative assessment have become more challenging as mutation panels have expanded. For example, current surveys for the molecular genetic testing of cystic fibrosis interrogate mutations performed by many laboratories (http://cap.org, last accessed February 29, 2012) but not necessarily full gene sequencing. Thus, even for relatively few mutations, proficiency testing challenges cannot fully address all tests. The use of DNA sequencing in the clinical laboratory has raised additional challenges because no proficiency testing or alternative assessment scheme is practically capable of assessing all possible results. This has been addressed through generation of a method-specific, as opposed to an analyte-specific, scheme in Europe.[71, 72] More recently, the College of American Pathologists has added a similar, multitiered sequencing challenge to its proficiency testing program (http://www.cap.org, last accessed February 29, 2012).

Regarding NGS, no proficiency testing is available in the United States. Because of the quantity and variation of data recovered from these assays, schemes that are based on NGS (as a group of methods), rather than on individual genes, are likely to be pursued. It is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

reasonable to envision that a general scheme can be developed in which the proficiency testing/alternative assessment of blinded samples may vary, depending on the actual application. In the near term, laboratories engaging in NGS are likely to use sample exchanges or blinded retesting of previously tested samples as a means of performing alternative assessment.[70] If a method-based scheme emerges, particular consideration should be paid to the kinds of variants that are intended to be detected by a given test. For example, the ability to detect substitutions, insertions, deletions, inversions, and copy number variations should be considered and proficiency testing for each category of variants intended to be detected by the method should be assessed. A further significant challenge is in developing measures for acceptable performance, including considerations of acceptability thresholds for limited versus comprehensive testing, such as for selected gene panel sequencing as opposed to genome sequencing.

Professional Considerations

Reimbursement

The payment (reimbursement) and coverage experience in molecular pathology has been bemoaned by both laboratories and payers over almost two decades. The transparency of services performed by laboratories, particularly in the areas of inherited diseases, oncology, and histocompatibility, has been particularly problematic for payers because of the analytically precise, but clinically opaque, way in which services are coded, using a universal group of current procedural terminology (CPT) codes that describe individual technical steps (eg, identification by DNA sequence analysis, 83904) that compose an assay. That same 83904 code is a component of all sequencing assays, regardless of the purposes for the tests. Moreover, because DNA sequencing assays require amplification of multiple exons for sequencing, 83904 and other technical codes are (appropriately) submitted in multiple units of service. Hence, payers are presented with code stacks that do not indicate the condition for which the sequence analysis was performed.

A special work group of the American Medical Association CPT Editorial Panel, working off a white paper from the Association for Molecular Pathology (AMP) Economic Affairs Committee, has recently released the framework of a solution for this major shortcoming in molecular pathology coding. Two groups of category 1 codes are proposed to replace the stacking codes (83890 to 83914) over a period of roughly 2 years. The first group (tier 1) provides individual CPT codes for specific tests performed in high volume or desired by payers. Taking cystic fibrosis as an example, individual codes would be available for high-volume screening assays for common mutations, as well as for full gene sequencing, deletion/duplication analysis, and even analysis for known familial mutations or the *CFTR* intron 8 poly (TG)T repeat unit. The second group (tier 2) is a series of nine codes that include many of the less commonly performed tests, divided according to increasing amounts of technical and professional resources. For example, level 4 includes tests that require sequencing of a single exon of a gene (as in testing a relative for a known familial mutation), whereas level 9 is for full sequencing of genes with > 50 exons. A not otherwise specified code (ending in 99) will still exist in the 80000 CPT series reserved for new and emerging technologies. An expert group of individuals with process and technical expertise

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

will advise the CPT Editorial Panel on whether new tests should receive their own CPT code or be placed in one of the nine tier 2 codes with other assays of similar complexity. A code proposal typically takes 15 to 24 months for review and approval by the CPT Editorial Panel.

With the transition of NGS to the clinical laboratory, the question of payment for NGS testing arises. This question presupposes that such tests will be widespread and clinically useful, which is a requirement for approval of a category 1 CPT code. Published peer-reviewed evidence of medical utility will be important for the CPT Editorial Panel to approve codes, as well as for payers to agree to pay for tests such as genome, exome, or transcriptome analyses. Payers already cover some multigene sequencing panels performed by Sanger or microarray-based sequencing. For reasons of cost and complexity of data analysis, these types of targeted tests are likely to be the leading edge of NGS applications for coding and reimbursement. With rapidly decreasing technical costs, it is not unrealistic to consider that exome or ultimately genome sequencing may become the default testing method for a wide range of clinical purposes, with interpretation initially focused on specific multigene panels and possibly blinding the balance of genomic variants.

Although software tools are essential to assemble and provide initial analysis of raw NGS data, interpretation for clinical use will require professional training and judgment. This effort will be the most time-consuming and, therefore, costly component of the testing and may be a significant cost for the laboratory, if this is where the interpretation occurs. Hence, NGS billing codes should reside on the physician fee schedule, with a component for professional services. Although this will not be an issue for physicians, laboratory genetics providers with doctoral training who obtain independent board certification are not yet eligible to bill professional component services on the physician fee schedule. A multistakeholder coalition is attempting to resolve that issue. New payment models based on episodes of care would, over time, presumably include NGS tests based on evidence that these contribute meaningfully to patient management and outcomes.

Gene Patents and Genome Sequencing

Utility patents are a specific subtype of intellectual property that give their owners the right to exclude others from making, using, offering to sell, or selling an invention in the United States or importing an invention for the life of the patent,[73] which is typically 20 years.[74] Patents can be awarded for novel, nonobvious, and useful processes, machines, manufactures, and compositions of matter.[75] Despite judicial doctrines that prohibit the patenting of products of nature and natural phenomena,[76, 77] the US Patent Office has granted thousands of patents on human gene sequences, variants, and associations between gene variants and clinical phenotypes[78] because US patent law has generally regarded isolated DNA as a complex chemical, instead of a product of nature. In addition, patents on genotype-phenotype associations have been awarded when claims include correlation steps that generically frame the biological relationships as processes.[79]

Patents on human gene sequences, if ultimately valid, could present an enormous barrier to the incorporation of genome, exome, and transcriptome sequencing into medical practice. Making, using, or selling a patented invention without the consent of the patent holder

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

constitutes patent infringement. Genome sequencing by its nature potentially appears to involve infringement of numerous gene sequence patents, irrespective of the clinical use of the information obtained. Moreover, clinical application of the sequence information would also likely infringe on thousands of genotype-phenotype association patents. Although theoretically the data from patented genes during genome sequencing and interpretation could be blinded, this would greatly diminish the usefulness of the test results. Perhaps through the formation of gene patent pools, an affordable system of common access to genes and genetic information ultimately could be developed. However, valuation of thousands of individual genes and genetic relationships would prove problematic. When coupled with evolving knowledge and the continual need for the addition of new genetic relationships to this pool, the challenges associated with implementation appear daunting.

In *Association for Molecular Pathology v United States Patent and Trademark Office*, a lawsuit sponsored by the American Civil Liberties Union, several medical societies, health care providers, breast cancer patients, and women's groups sued Myriad Genetics and the Patent Office, seeking to invalidate key claims of patents covering the wild-type and mutated sequences of the *BRCA1* and *BRCA2* genes, as well as associations between those sequences and the predisposition to breast and/or ovarian cancer. In a landmark decision, the court held that both composition of matter claims on the human *BRCA1* and *BRCA2* gene sequences and process claims covering the correlations between mutations in these genes and a heritable predisposition to breast cancer are invalid. However, this ruling was partially overturned on appeal and it appears likely that this case and the patentability of human gene sequences and genotype-phenotype associations will eventually be decided by the US Supreme Court. *Association for Molecular Pathology v United States Patent and Trademark Office*, therefore, will have profound implications for the clinical application of genome sequencing.

Genomics Education

The human genome can be likened to a new human organ system for which current medical training provides rudimentary education at best. As with any field of medicine, it is imperative that ongoing and continuing medical education be used to learn about new modalities and emerging technologies that reflect current practice. No single program or fellowship can be expected to comprehensively train its residents, given the dynamically evolving milieu. However, it is essential to provide trainees with a good grasp of the current concepts while exposing them to a broad range of educational opportunities. Professional organizations, such as the AMP, the College of American Pathologists, and the American College of Medical Genetics, are working on various aspects of this, as are individual training programs. Within pathology and medical genetics training programs, molecular genetic pathology and clinical molecular genetics fellowships provide the most intense training in molecular testing, which can lead to board certification through the American Board of Pathology or the American Board of Medical Genetics. Anatomical and clinical pathology residency training programs are required to provide molecular pathology training for all residents, as reflected by the inclusion of molecular pathology questions on the pathology board examinations. However, the extent of molecular pathology education within residency programs is variable. Practicing pathologists and clinical laboratory geneticists

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

have a wide range of molecular knowledge, depending on their past training, length of time in practice, practice setting, and the types of continuing medical education and lifelong learning in which they have participated. Genomics and genome testing are just beginning to be integrated into existing training programs; however, without doubt, these practices will increase in prominence in the near future.

How do we transition to the incorporation of genomics into pathology training and practice? A group of pathology and genetics societies has formed a working group called Training Residents in Genomics, which is developing a pathology residency genomics curriculum that can be adopted by any pathology residency training program. The curriculum will consist of a set of Power- Point lectures, practical exercises, and literature references that will be available on the Internet. A similar initiative of an online genomics course, which will be made publicly available, has been developed at Stanford University (Stanford, CA). Clearly, these programs will provide basic genomics training to pathology residents and pathologists in practice, but will not be sufficient for the actual practice of genomics and use of NGS for clinical testing, or even for the proper interpretation of specific genomic results within the context of pathology practice. Broader inclusion of genomic test results into the remainder of pathology education will be required to more completely integrate NGS education. Another option for pathology residency training is to develop tracks with a genomics training focus, while maintaining the broader educational requirements of anatomical and clinical pathology training. For training in the performance and interpretation of large-scale genomic tests, molecular genetic pathology fellowship programs will probably continue to offer the best option for intensive training. After all, this type of testing is the next logical progression for the practice of molecular pathology and will likely replace many current molecular tests as the technology becomes less expensive. This transition could be driven by changes in the Accreditation Council for Graduate Medical Education requirements for molecular genetic pathology fellowship programs. For practicing pathologists, continuing medical education programs in the use of genomics, as applied to specific types of pathology practice settings, are available from several sources but will need to be expanded as the use of genomics and NGS increases.

Genomics education also is needed outside the practice of pathology. For other medical specialties, pathologists can act as the resource for the proper use and interpretation of genomic test results for patient care, as is already the case for all pathology testing, but only if pathologists truly understand this testing. In addition, each medical specialty will incorporate genomics aspects significant to its specialty into its educational pathways. Medical education also needs to incorporate genomics education into the medical school curriculum. With the revisiting of the Flexner Report in 2010, many medical schools are assessing and revising their curricula, with the opportunity to incorporate and emphasize the use of human genome data into medical practice and patient care. Finally, genomics education and literacy for patients, their families, and the public will assist with the acceptance of genomics into medical care. Patient and family education can be addressed within specific medical practice settings and through disease-based support organizations, whereas public education may be addressed prospectively through K-12 and undergraduate education curriculum enhancements and through public education campaigns.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Future Applications and Implications of NGS

To predict the future, we can only try to seek lessons from the past. As prescient and understated as Watson and Crick were when they wrote "*It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material*," [80] could they have imagined how the golden age of molecular biology would lead to molecular diagnostics? Since then, molecular biology has progressed to sequencing short DNA sequences in experiments that took days[1–3, 81] and evolved to sequencing individuals' genomes in a similar, indeed shorter, time frame. Given the dramatic technical progress in transportation, communication, information technology, and medicine over approximately 60 years, going from the elucidation of the double helix to sequencing entire genomes may have been a logical evolution, but only in hindsight.

Now that much of the genome is already available for technical interrogation by NGS, multiple new diagnostic opportunities exist for molecular pathologists and their patients. Examples are already beginning to abound (eg, in molecular cardiology,[11] molecular oncology,[82] and noninvasive trisomy detection[83–85]). The list of potential applications is limited only by our imaginations. Molecular pathologists and geneticists have been professional diagnosticians of genetic and now genomic conditions and have interpreted the data generated in clinical molecular laboratories since the field's inception. They serve as expert analysts and consultants to other physicians and will continue to do so as dependable stewards of these new technologies. Molecular pathologists and laboratory geneticists, in partnership with other genomicstrained clinicians and health care professionals, are essential to ensuring that these test results are obtained, used, and reported responsibly. The AMP and its members play a pivotal role in this area of molecular diagnostics and in its future. AMP members comprise the largest proportion of those setting the framework for the use of NGS and genome testing by developing and offering these services. In thinking strategically, AMP and its members are well positioned to productively collaborate with other pathology and genetics societies to author practice guidelines and to recommend quality assessment monitors and metrics. The evolving technology innovations are rapidly reducing the cost of materials and instrumentation, to the point at which many laboratories are considering leapfrogging from single-variation testing to NGS methods that evaluate much of the genome. Notably, although the costs of performing the assays have decreased, the fundamental process of applying clinical expertise to the interpretation of the identified variants remains largely unaltered and remains central to our contributions to patient care.

AMP, through its diverse expert membership, is uniquely positioned to proactively address these issues and to inform the development of proficiency testing and alternative assessment schemes necessary for ensuring the quality of clinical NGS testing. AMP is advancing the issues that are presented, and is shaping the landscape in a collaborative process with other professional organizations to accomplish the goals of excellent medical practice in the genome era. Now no longer an abstract concept for the future, the exciting reality of powerful genome testing has decisively arrived.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Disclaimer

The Association for Molecular Pathology (AMP) Clinical Practice Guidelines and Reports are developed to be of assistance to laboratory and other health care professionals by providing guidance and recommendations for particular areas of practice. The Guidelines or Report should not be considered inclusive of all proper approaches or methods, or exclusive of others. The Guidelines or Report cannot guarantee any specific outcome, nor do they establish a standard of care. The Guidelines or Report are not intended to dictate the treatment of a particular patient. Treatment decisions must be made based on the independent judgment of health care providers and each patient's individual circumstances.

AMP makes no warranty, express or implied, regarding the Guidelines or Report and specifically excludes any warranties of merchantability and fitness for a particular use or purpose. AMP shall not be liable for direct, indirect, special, incidental, or consequential damages related to the use of the information contained herein.

Acknowledgments

Supported by the Association for Molecular Pathology.

References

1. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437:376–380. [PubMed: 16056220]
2. Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A. 1977; 74:560–564. [PubMed: 265521]
3. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977; 74:5463–5467. [PubMed: 271968]
4. Smith L, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. Fluorescence detection in automated DNA sequence analysis. Nature. 1986; 321:674–679. [PubMed: 3713851]
5. Rosenblum BB, Lee LG, Spurgeon SL, Khan SH, Menchen SM, Heiner CR, Chen SM. New dye-labeled terminators for improved DNA sequence patterns. Nucleic Acids Res. 1997; 25:4500–4504. [PubMed: 9358158]
6. Huang XC, Quesada MA, Mathies RA. DNA sequencing using capillary array electrophoresis. Anal Chem. 1992; 64:2149–2154. [PubMed: 1416049]
7. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [PubMed: 11237011]
8. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004; 431:931–945. [PubMed: 15496913]
9. Shendure J, Ji H. Next-generation DNA sequencing. Nature Biotechnol. 2008; 26:1135–1145. [PubMed: 18846087]
10. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. Clin Chem. 2009; 55:641–658. [PubMed: 19246620]
11. Anderson MW, Schrijver I. Next generation DNA sequencing and the future of genomic medicine. Genes. 2010; 1:38–69. [PubMed: 24710010]
12. Metzker ML. Sequencing technologies: the next generation. Nat Rev Genet. 2010; 11:31–46. [PubMed: 19997069]
13. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011; 175:348–352.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

14. Cooper DN, Schmidtke J. Diagnosis of genetic disease using recombinant DNA. Hum Genet. 1986; 73:1–11. [PubMed: 3011642]
15. Cooper DN, Schmidtke J. Diagnosis of genetic disease using recombinant DNA: second edition. Hum Genet. 1989; 83:307–334. [PubMed: 2572533]
16. Cooper DN, Schmidtke J. Diagnosis of genetic disease using recombinant DNA: third edition. Hum Genet. 1991; 87:519–560. [PubMed: 1916757]
17. Cooper DN, Schmidtke J. Diagnosis of human genetic disease using recombinant DNA: fourth edition. Hum Genet. 1993; 92:211–236. [PubMed: 8406430]
18. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet. 2010; 42:790–793. [PubMed: 20711175]
19. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, McGuire AL, Zhang F, Stankiewicz P, Halperin JJ, Yang C, Gehman C, Guo D, Irikat RK, Tom W, Fantin NJ, Muzny DM, Gibbs RA. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N Engl J Med. 2010; 362:1181–1191. [PubMed: 20220177]
20. Hardy J, Singleton A. Genomewide association studies and human disease. N Engl J Med. 2009; 360:1759–1768. [PubMed: 19369657]
21. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet. 1999; 22:139–144. [PubMed: 10369254]
22. Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med. 2010; 363:166–176. [PubMed: 20647212]
23. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Metaanalysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet. 2003; 33:177–182. [PubMed: 12524541]
24. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet. 2002; 3:391–397. [PubMed: 11988764]
25. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005; 6:95–108. [PubMed: 15716906]
26. Margaglione M, Grandone E. Population genetics of venous thromboembolism: a narrative review. Thromb Haemost. 2011; 105:221–231. [PubMed: 20941456]
27. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PW, D'Agostino RB Sr, Cupples LA. Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med. 2008; 359:2208–2219. [PubMed: 19020323]
28. Paynter NP, Chasman DI, Paré G, Buring JE, Cook NR, Miletich JP, Ridker PM. Association between a literature-based genetic risk score and cardiovascular events in women. JAMA. 2010; 303:631–637. [PubMed: 20159871]
29. Eichler EE, Flint J, Gibson G, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010; 11:446–450. [PubMed: 20479774]
30. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature. 2010; 463:191–196. [PubMed: 20016485]
31. McDermott U, Downing JR, Stratton MR. Genomics and the continuum of cancer care. N Engl J Med. 2011; 364:340–350. [PubMed: 21268726]
32. Schubert C. Cancer drugs find a companion with new diagnostic tests. Nat Med. 2011; 17:1157. [PubMed: 21988963]
33. Burt RW, Barthel JS, Dunn KB, David DS, Drelichman E, Ford JM, Giardiello FM, Gruber SB, Halverson AL, Hamilton SR, Ismail MK, Jasperson K, Lazenby AJ, Lynch PM, Martin EW Jr, Mayer RJ, Ness RM, Provenzale D, Rao MS, Shike M, Steinbach G, Terdiman JP, Weinberg D. NCCN. NCCN clinical practice guidelines in oncology: colorectal cancer screening. J Natl Compr Canc Netw. 2010; 8:8–61. [PubMed: 20064289]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

34. Daly MB, Axilbund JE, Buys S, Crawford B, Farrell CD, Friedman S, Garber JE, Goorha S, Gruber SB, Hampel H, Kaklamani V, Kohlmann W, Kurian A, Litton J, Marcom PK, Nussbaum R, Offit K, Pal T, Pasche B, Pilarski R, Reiser G, Shannon KM, Smith JR, Swisher E, Weitzel JN. National Comprehensive Cancer Network. Genetic/familial high-risk assessment: breast and ovarian. J Natl Compr Canc Netw. 2010; 8:562–594. [PubMed: 20495085]
35. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR. Genomewide in situ exon capture for selective resequencing. Nat Genet. 2007; 39:1522–1527. [PubMed: 17982454]
36. Taylor BS, Ladanyi M. Clinical cancer genomics: how soon is now? J Pathol. 2011; 223:318–326. [PubMed: 21125684]
37. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008; 456:66–72. [PubMed: 18987736]
38. McLendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogianakis GM, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455:1061–1068. [PubMed: 18772890]
39. Bell D, Berchuck A, Birrer M, Chien J, Cramer DW, Dao F, et al. Integrated genomic analyses of ovarian carcinoma. Nature. 2011; 474:609–615. [PubMed: 21720365]
40. Collins FS, Barker AD. Mapping the cancer genome: pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. Sci Am. 2007; 296:50–57.
41. Foekens JA, Atkins D, Zhang Y, Sweep FC, Harbeck N, Paradiso A, Cufer T, Sieuwerts AM, Talantov D, Span PN, Tjan-Heijnen VC, Zito AF, Specht K, Hoefler H, Golouh R, Schittulli F, Schmitt M, Beex LV, Klijn JG, Wang Y. Multicenter validation of a gene expression–based prognostic signature in lymph node–negative primary breast cancer. J Clin Oncol. 2006; 24:1665–1671. [PubMed: 16505412]
42. Mamounas EP, Tang G, Fisher B, Paik S, Shak S, Costantino JP, Watson D, Geyer CE Jr, Wickerham DL, Wolmark N. Association between the 21-gene recurrence score assay and risk of locoregional recurrence in node-negative, estrogen receptor-positive breast cancer: results from NSABP B-14 and NSABP B-20. J Clin Oncol. 2010; 28:1677–1683. [PubMed: 20065188]
43. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002; 415:530–536. [PubMed: 11823860]
44. Lam HY, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M, Dewey FE, Habegger L, Ashley EA, Gerstein MB, Butte AJ, Ji HP, Snyder M. Performance comparison of whole-genome sequencing platforms. Nature Biotechnol. 2011; 30:78–82. [PubMed: 22178993]
45. Beck JS, Kwitek AE, Cogen PH, Metzger AK, Duyk GM, Sheffield VC. A denaturing gradient gel electrophoresis assay for sensitive detection of p53 mutations. Hum Genet. 1993; 91:25–30. [PubMed: 8454284]
46. Greiner TC. Enhanced detection of TP53 mutations using a GC-clamp in denaturing high performance liquid chromatography. Diagn Mol Pathol. 2007; 16:32–37. [PubMed: 17471156]
47. Ogino S, Kawasaki T, Brahmandam M, Yan L, Cantor M, Namgyal C, Mino-Kenudson M, Lauwers GY, Loda M, Fuchs CS. Sensitive sequencing method for KRAS mutation detection by Pyrosequencing. J Mol Diagn. 2005; 7:413–421. [PubMed: 16049314]
48. Li J, Wang L, Mamon H, Kulke MH, Berbeco R, Makrigiorgos GM. Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. Nat Med. 2008; 14:579–584. [PubMed: 18408729]
49. Green ED, Guyer MS. National Human Genome Research Institute. Charting a course for genomic medicine from base pairs to bedside. Nature. 2011; 470:204–213. [PubMed: 21307933]
50. Altman RB. Direct-to-consumer genetic testing: failure is not an option. Clin Pharmacol Ther. 2009; 86:15–17. [PubMed: 19536117]
51. Khoury MJ, McBride CM, Schully SD, Ioannidis JP, Feero WG, Janssens AC, et al. Centers for Disease Control and Prevention. The scientific foundation for personal genomics:

recommendations from a National Institutes of Health-Centers for Disease Control and Prevention multidisciplinary workshop. Genet Med. 2009; 11:559–567. [PubMed: 19617843]

52. Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. Am J Hum Genet. 2009; 85:142–154. [PubMed: 19679224]
53. Drmanac R. The advent of personal genome sequencing. Genet Med. 2011; 13:188–190. [PubMed: 21311341]
54. Sharp RR. Downsizing genomic medicine: approaching the ethical complexity of whole-genome sequencing by starting small. Genet Med. 2011; 13:191–194. [PubMed: 21311340]
55. Grody W. Ethical issues raised by genetic testing with oligonucleotide microarrays. Mol Biotechnol. 2003; 23:127–138. [PubMed: 12632697]
56. Ten Bosch J, Grody W. Keeping up with the next generation massively parallel sequencing in clinical diagnostics. J Mol Diagn. 2008; 10:484–492. [PubMed: 18832462]
57. International Organization for Standardization. Medical Laboratories: Particular Requirements for Quality and Competence. Geneva, Switzerland: International Organization for Standardization; 2003.
58. Clinical Laboratory Standards Institute (CLSI). Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine: Approved Guideline, CLSI Document MM9-A. Wayne, PA: Clinical and Laboratory Standards Institute: 2007.
59. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2011; 13:36–46. [PubMed: 22124482]
60. Su Z, Ning B, Fang H, Hong H, Perkins R, Tong W, Shi L. Next-generation sequencing and its applications in molecular diagnostics. Expert Rev Mol Diagn. 2011; 11:333–343. [PubMed: 21463242]
61. Masys DR, Jarvik GP, Abernethy NF, Anderson NR, Papanicolaou GJ, Paltoo DN, Hoffman MA, Kohane IS, Levy HP. Technical desiderata for the integration of genomic data into Electronic Health Records. J Biomed Inform. 2012; 45:419–422. [PubMed: 22223081]
62. Flicek P, Birney E. Sense from sequence reads: methods for alignment and assembly. Nat Methods. 2009; 6:S6–S12. [PubMed: 19844229]
63. 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; 7:e1002280. [PubMed: 21935354]
64. Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE. Molecular Subcommittee of the ACMG; Laboratory Quality Assurance Committee. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. Genet Med. 2008; 10:294–300. [PubMed: 18414213]
65. Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA, Pushkarev D, Neff NF, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreiya H, Whaley R, Knowles JW, Chou MF, Thakuria JV, Rosenbaum AM, Zaranek AW, Church GM, Greely HT, Quake SR, Altman RB. Clinical assessment incorporating a personal genome. Lancet. 2010; 375:1525–1535. [PubMed: 20435227]
66. Ali-Khan SE, Daar AS, Shuman C, Ray PN, Scherer SW. Whole genome scanning: resolving clinical diagnosis and management amidst complex data. Pediatr Res. 2009; 66:357–363. [PubMed: 19531980]
67. Centers for Disease Control and Prevention (CDC) (2) Centers for Medicare & Medicaid Services (CMS), HHS. Medicare, Medicaid, and CLIA programs: laboratory requirements relating to quality systems and certain personnel qualifications: final rule. Fed Regist. 2003; 68:3639–3714. [PubMed: 12545998]
68. Peterson JC, Hill RH, Black RS, Winkelman J, Tholen D. Review of Proficiency Testing Services for Clinical Laboratories in the United States: Final Report of a Technical Working Group, CDC.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

69. Clinical Laboratory Standards Institute (CLSI). Using Proficiency Testing to Improve the Clinical Laboratory: Approved Guideline. 2. Wayne, PA: Clinical and Laboratory Standards Institute; 2007.
70. Clinical Laboratory Standards Institute (CLSI). Assessment of Laboratory Tests When Proficiency Testing Is Not Available: Approved Guideline. 2. Wayne, PA: Clinical and Laboratory Standards Institute; 2007.
71. Ramsden SC, Daly S, Geilenkeuser WJ, Duncan G, Hermitte F, Marubini E, Neumaier M, Orlando C, Palicka V, Paradiso A, Pazzagli M, Pizzamiglio S, Verderio P. EQUALquant: an international external quality assessment scheme for real-time PCR. Clin Chem. 2006; 52:1584–1589. [PubMed: 16740649]
72. Patton SJ, Wallace AJ, Elles R. Benchmark for evaluating the quality of DNA sequencing: proposal from an international external quality assessment scheme. Clin Chem. 2006; 52:728–736. [PubMed: 16455867]
73. Infringement of patent. Code of Federal Regulations. Title 35, Pt. 271, 2010 ed.
74. Contents and term of patent; provisional rights. Code of Federal Regulations. Title 35, Pt. 154, 2010 ed.
75. Inventional patentable. Code of Federal Regulations. Title 35, Pt. 101–103, 2010 ed.
76. Diamond v Chakrabarty, 447 US 303. 1980.
77. Diamond v Diehr, 450 US 175. 1981.
78. Jensen K, Murray F. Intellectual property: enhanced: intellectual property landscape of the human genome. Science. 2005; 310:239–240. [PubMed: 16224006]
79. Klein RD. Gene patents and genetic testing in the United States. Nature Biotechnol. 2007; 25:989–990. [PubMed: 17846625]
80. Watson JD, Crick FHC. A structure for deoxyribose nucleic acid. Nature. 1953; 171:737–738. [PubMed: 13054692]
81. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975; 94:441–448. [PubMed: 1100841]
82. Ross JS, Cronin M. Whole cancer genome sequencing by nextgeneration methods. Am J Clin Pathol. 2011; 136:527–539. [PubMed: 21917674]
83. Chiu RW, Akolekar R, Zheng YW, Leung TY, Sun H, Chan KC, Lun FM, Go AT, Lau ET, To WW, Leung WC, Tang RY, Au-Yeung SK, Lam H, Kung YY, Zhang X, van Vugt JM, Minekawa R, Tang MH, Wang J, Oudejans CB, Lau TK, Nicolaides KH, Lo YM. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ. 2011; 342:c7401. [PubMed: 21224326]
84. Ehrich M, Deciu C, Zwiefelhofer T, Tynan JA, Cagasan L, Tim R, Lu V, McCullough R, McCarthy E, Nygren AO, Dean J, Tang L, Hutchinson D, Lu T, Wang H, Angkachatchai V, Oeth P, Cantor CR, Bombard A, van den Boom D. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. Am J Obstet Gynecol. 2011; 204:205.e1–11. [PubMed: 21310373]
85. Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, Ehrich M, van den Boom D, Bombard AT, Deciu C, Grody WW, Nelson SF, Canick JA. DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. Genet Med. 2011; 13:913–920. [PubMed: 22005709]