Review

Whole Genome Sequencing as a Diagnostic Test: Challenges and Opportunities

Caitlin C. Chrystoja[1,2] and Eleftherios P. Diamandis[1,2,3*]

BACKGROUND: Extraordinary technological advances and decreases in the cost of DNA sequencing have made the possibility of whole genome sequencing (WGS) as a highly accessible clinical test for numerous indications feasible. There have been many recent, successful applications of WGS in establishing the etiology of complex diseases and guiding therapeutic decision-making in neoplastic and nonneoplastic diseases and in various aspects of reproductive health. However, there are major, but not insurmountable, obstacles to the increased clinical implementation of WGS, such as hidden costs, issues surrounding sequencing and analysis, quality assurance and standardization protocols, ethical dilemmas, and difficulties with interpretation of the results.

CONTENT: The widespread use of WGS in routine clinical practice remains a distant proposition. Prospective trials will be needed to establish if, and for whom, the benefits of WGS will outweigh the likely substantial costs associated with follow-up tests, the risks of over-diagnosis and overtreatment, and the associated emotional distress.

SUMMARY: WGS should be carefully implemented in the clinic to allow the realization of its potential to improve patient health in specific indications. To minimize harm the use of WGS for all other reasons must be carefully evaluated before clinical implementation.

© 2013 American Association for Clinical Chemistry

The Human Genome Sequencing Project was completed in 2001, at a cost of approximately $3 billion. The sequence came with the great promise to revolutionize the understanding of the mechanisms and treatment of just about every human disease. Twelve years later, the human genome sequence has undoubtedly played a major role in the accelerated understanding of the pathobiology of many diseases but has yet to live up to the hype in the press of transforming all aspects of clinical practice. Currently this technology is being used clinically for select diagnostic scenarios, but its use is limited by guidelines from the American College of Medical Genetics (ACMG)[4] detailing specific instances that will be authorized for reimbursement. Since 2001, some spectacular technological advances have allowed much faster and cheaper delineation of genomic sequences, known as second and third generation sequencing. These technologies, and their possible applications in medicine, have been reviewed elsewhere *(1–3)*. Another important development has been the evolution of commercial organizations, which started to provide "direct-to-consumer" (DTC) genomic information, especially high-throughput single-nucleotide polymorphism (SNP) analysis, and preliminary interpretations related to future predisposition to various diseases *(4, 5)*. The continuing advances in sequencing technologies have now reached the point at which a genome can be sequenced within a few days at very reasonable costs (<$10 000, with the goal of the $1000 genome) *(2)*. These extraordinary technological advances have triggered discussions that whole genome sequencing (WGS) may become a relatively straightforward and highly accessible genomic test that could revolutionize diagnosis of any disease with a genetic component, facilitate understanding of the mechanisms behind development of many diseases, and consequently catalyze the evolution of effective treatments under the new buzzword of "personalized medicine" *(6)*. There are, indeed, some recent examples for which WGS has already made a difference as a method for establishing the diagnosis of difficult diseases *(7, 8)*, as a platform for selecting appropriate therapy of cancer patients *(9, 10)*, and as a tool to discover novel disease-associated mutations *(11)*. It is thus natural that a debate is rapidly evolving as to when and how WGS will

[1] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada; [2] Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, ON, Canada; [3] Department of Clinical Biochemistry, University Health Network, Toronto, ON, Canada.

* Address correspondence to this author at: Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 6th Floor, Rm. 6-201, Box 32, 60 Murray S., Toronto, ON, Canada M57 3L9. Fax 416-619-5521; e-mail ediamandis@mtsinai.on.ca.

[4] Nonstandard abbreviations: ACMG, American College of Medical Genetics; DTC, direct-to-consumer; SNP, single-nucleotide polymorphism; WGS, whole genome sequencing; CAP, College of American Pathologists; NGS, next generation sequencing; WES, whole exome sequencing; AMP, Association of Molecular Pathology.

become a standard tool for preventative, diagnostic, prognostic, and predictive tests and even for monitoring diagnostic tests *(12)*.

There are important and as of yet unanswered questions about the implementation of WGS beyond research and in select clinical applications. What are the technical and genotype–phenotype linkage false-positive rates of WGS and what could it take to resolve them? How many patients will be subjected to unnecessary, expensive, and potentially harmful procedures because of these false positives? Will knowledge of a disease predisposition be useful for disease risk modification (if we have such strategies) or a constant (and likely harmful) psychological burden? Will we need lengthy (probably lifelong) prospective trials to show if WGS will be ultimately beneficial, in terms of some defined end point, or harmful, and for whom and for which indications? Should we adopt WGS because we can, or because it is rather cheap, or because it is useful? Research in parallel with diagnostic applications is underway in a number of centers to study these questions.

The controversy of WGS as a diagnostic test is just beginning (Table 1). Below we analyze some issues related to WGS as a clinical test and evaluate where we currently stand. We seek to answer the questions of whether WGS could be a useful test now, and if this test could have a positive impact on people's health. Fig. 1 shows some putative clinical applications of WGS.

Cost Issues

Sequencing a genome now costs less than $10 000, and with the continuing advances in sequencing technologies the possibility of sequencing a genome for $1000 or less seems increasingly likely in the near future. However, the cost of sequencing a genome is misleading because it fails to include the exponentially higher costs of analyzing and translating raw sequences into a meaningful output that benefits patients (Table 2).

Informed consent is needed before WGS is performed. The process may require several meetings, totaling 6 to 8 hours of a clinician's or counselor's time *(11)*. This is primarily to discuss the ramifications of incidental findings of clinically relevant genetic variants. Each genome is expected to contain approximately 150 000 novel SNPs *(13)*, of which 250–350 will be disruptive variants in genes, 50–100 will be variants in human disease genes, and 20 will be variants that inactivate genes *(14, 15)*. The bioinformatic processing cost for alignment through variant calling (as described below) with a cloud-computing approach costs as little as $120 to identify only SNPs, with the cost increasing to identify greater variation *(16)*. Explaining results to patients adds additional time, with one institution finding that it required an additional 5 hours *(17)*. As

Table 1. Some currently debated issues with WGS.[a]	
Issue	**Comments**
Cost	Usually underestimated (see Table 2)
Technological	Error rate; sequence completeness; interinstrumental variability; sequencing depth; base-calling algorithms; aligning read algorithms.
Quality assurance	CAP checklist; minimal standards for clinical use.
Interpretative	Incidental findings, how to report them to patients and families (see Table 3); how to prioritize variants to predict future effects and use them to counsel patients; impact of ethnicity; lack of sufficient numbers of genetic counselors; physician education.
Ethical	Informed consent including pretest and posttest counseling; avoid harm (physical and psychological); disclosure of WGS data linked to behavioral issues or psychiatric disorders; WGS and discrimination; privacy and data security.
Efficacy	Weak evidence for efficacy of WGS to predict future disease risk in asymptomatic individuals [Roberts et al. *(57)*].
Patents	Infringement on current patents protecting disease-associated mutations.

[a] See text for more details.

institutions gain experience with WGS and guidelines continue to evolve, pre- and posttest counseling will become more streamlined, which may reduce the amount of time needed. Identified variants that have clinical importance need to be confirmed using a gold
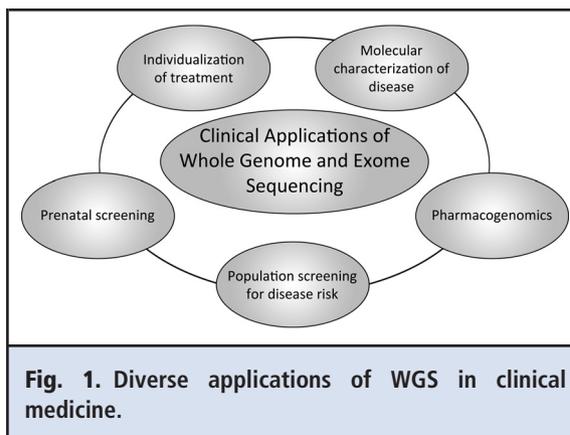

Fig. 1. Diverse applications of WGS in clinical medicine.

Table 2. The real costs of WGS for clinical use.		
Clinical procedure	**Service performed**	**Approximate cost, $ [reference]**
Sequence	Reagents and labor (for 30–40 fold coverage with Illumina's technology).	6500 [University of Texas at Austin (74)]
Informed consent	Clinician's time to discuss ramifications of incidental findings (6–8 h).	3000 [Bick and Dimmock (11)]
Interpretative consultation	Explaining results to patients (5 h).	2000 [Ormond et al. (17)]
Bioinformatic processing	Identify variants, disruptive variants, inactivating genes.	120 [Zhao et al. (16)]
Confirmatory testing[a]	Confirm mutations by Sanger sequencing. Not all laboratories follow recent ACMG guidelines and especially in targeted applications, only variants in genes considered clinically relevant to a patient's phenotype may be confirmed. Approximately 5 or more mutations × $200/mutation.	1000 [Prevention Genetics (18)]
Confirming disease presence[b]	Patients may harbor disease-associated variants but may not have the disease. Additional testing (endoscopic, imaging, laparoscopic, etc.).	2000–10 000
Genome data storage	Lifelong storage or resequencing in the future with more accurate methods.	5000

[a] Costs expected to drop in the future.
[b] Costs not determined for overdiagnosis and patient anxiety and distress.

standard method such as Sanger sequencing, which costs approximately $200 for each mutation, to verify that the variant is present in the genome and is not a sequencing-related error (18). As technological accuracy improves over the coming years, there will be a diminishing need for confirmatory testing and the resulting incurred cost. After confirmatory testing of a variants' presence in the genome, follow-up tests need to be performed. These after-sequencing costs dramatically increase the price of WGS (to approximately $24 000/test) and highlight the importance of being clear as to why a patient's genome is being sequenced in the first place, with a plan in place (in consultation with the patient) on what to do with the incidental findings.

After WGS has been performed, the patient's genome can be stored for future evaluation. As new medical questions arise, the genomic data can be reanalyzed. With improved software and the constantly increasing understanding about genome–phenotype linkages, revaluating a patient's genome can help identify new correlations. The cost of sequencing and analyzing a patient's genome can be amortized over the course of a patient's lifetime because their germline genome remains essentially constant. However, with the rapid development of new sequencing technologies with improved accuracy, it might be more cost-effective to resequence a patient's genome multiple times over the course of his or her lifetime rather than resorting to reanalysis of the previously sequenced genome on an older, less accurate sequencing platform, with the associated cost of storing the data. The cost of storing a patient's genome may be further amplified if the preliminary data (the raw reads and mapping results) are kept, in addition to the final output.

Technological and Quality Issues

Because WGS is still in its infancy, standardization practices have only recently come out. The College of American Pathologists (CAP) recently published the first accreditation checklist that addresses next generation sequencing (NGS) in its molecular pathology checklist. This is an important step for QC of genomic testing because the CAP checklists are used during the accreditation process to meet Centers for Medicare and Medicaid Services requirements, which regulate human clinical laboratory testing in the US through the Clinical Laboratory. Improvement Amendments. With different NGS platforms and bioinformatics software available to analyze genomic data (19), QC and standardization are critical. To address this issue, the CDC organized a conference, Next-Generation Sequencing—Standardization of Clinical Testing Meeting, the results of which have been recently published (20).

The Archon Genomics X Prize competition (21) calls for WGS of 100 genomes with a grand prize of $10 million to be awarded to the team(s) that can accomplish it within an accuracy of 1 error per million bases; completeness of 98%; identification of insertions, deletions, and rearrangements; and a complete haplotype for $1000 per genome. These criteria roughly characterize what will be the world's first medical-grade genome. With the possibility of errors throughout every step of analysis in WGS, medical-grade genomes will represent a clinical standard of the quality of a sequenced genome, with a consistent and minimized level of technological errors. In addition to sequencing an individual's genome to determine his or her geno-

type (using current methods, as described below), the ability to discern haplotypes is also important to provide the context of genomic variations. The recently described long fragment read technology enables the elucidation of an individual's haplotypes in a cost-effective and efficient manner (22). This allows the assignment of polymorphisms to a parental chromosome, the knowledge of which can provide clinically important information that a genotype cannot.

The recent high-profile retraction of a genome-wide association study that reported genetic variants associated with longevity in centenarians (23) highlights the importance of QC procedures. In that study (24), the errors arose from combining data from multiple genotyping platforms. Different sequencing platforms vary in their ability to identify variants, even when sequencing the same genome (25–27). In attempts to detect rare variants, sequencing a patient's genome on multiple platforms may help identify more possible candidates.

The first challenge in sequencing a patient's genome is accurate base-calling, which is the identification of bases from high-intensity data generated by the optical sensors of the platforms. Each platform has its own base-calling program supplied by the vendor, typically with alternative third party programs available (28). The error rate can be reduced by resequencing DNA samples to increase coverage and then combining the data into a more accurate consensus sequence; however, this increases costs (29). Third party base-calling algorithms provide increased accuracy and decreased coverage requirements, which simplifies downstream analysis at equal or even faster speeds than those provided by vendors, depending on computational configurations.

The reads must then be aligned to a reference genome. Current NGS technology read length varies from as short as 50 to closer to 1000 base pairs. Problems arise because billions of short reads need to be aligned to a reference genome consisting of billions of base pairs. An ever-increasing number of algorithms and software packages have been specifically developed for aligning reads, each with its own merits depending on user need (30). These relatively short reads may not be mapped to anywhere in the reference genome, which is currently a compilation of a relatively limited number of individuals and incomplete in highly repetitive and variable regions between individuals (centromeres and telomeres) due to sequencing errors, gaps in the reference genome, and balancing of appropriate analytical sensitivity with the run-time of the algorithm. As more genomes are added to the reference genome, the quality and usability of sequenced data will increase. Even when reads are correctly mapped, the accuracy of the mapped short reads varies, as seen

with differing alignment results when using different algorithms and by changing the parameters within a given algorithm (25–27). Future sequencing platforms that can increase read length could mitigate these issues. The sequencing company Oxford Nanopore is currently developing technology that uses nanopore sequencing to sequence read lengths of more than tens of thousands; however, this technology has yet to be perfected and commercialized (31).

The analytical sensitivity of variant calling represents another issue. Compared with the reference sequence generated by the Human Genome Project, any single individual's genome has about 4 million sequence variations (12). Of these, approximately 3.5 million are SNPs and thousands are structural variations consisting of insertions, deletions, rearrangements, and copy number variants (32). Algorithms vary in their ability to identify different types of variants, leading to researchers using multiple algorithms and then aggregating the results. More sophisticated algorithms need to be developed to address these issues. A particularly novel tool for structural variant discovery is forestSV (33), which can call structural variants effectively in a single genome.

Currently, sequencing and analyzing the human genome's 6 billion base pairs, in spite of an accuracy of only 1 false single-nucleotide variation per 500 kbp on the best platforms (34), still produces an astounding 12 000 errors per genome. These technical false positives may lead to erroneously identified clinically relevant disease-causing or risk-increasing variants that would mandate confirmatory testing, potentially subjecting individuals to unnecessary procedures and causing them undue concern.

During processing through the bioinformatic analysis pipeline, the huge data files generated need to be transferred between computing infrastructure components and software programs, requiring a significant capital investment for in-house information technology infrastructure. One promising option to offset some of the computing infrastructure requirements and costs for analysis is cloud computing. Cloud computing has the potential to offer a complete analysis pipeline in one easily accessible place. Importantly, the computationally intensive steps can be distributed over many other computers linked to the cloud, decreasing processing time. A complete data analysis pipeline using the cloud for WGS has been successfully described (16).

Interpretative and Ethical Issues

Incidental findings of variants not related to the reason WGS was performed may contain undesired information for patients. The information may be clinically im-

Table 3. Categories of genetic variants identified by whole genome sequencing (WGS) [Ayuso C et al. (35)].

Category	Description
1	Findings relevant to the reason WGS was performed.
2	Clinically relevant variants for which treatment is or is not available.
3	Variants causing high risk for future mendelian diseases.
4	Carrier status that can impact reproductive life decisions.
5	Variants of variable risk for future diseases.
6	Variants of unknown significance.

portant for the patients, their relatives, and/or their offspring. A recent report by Ayuso et al. (35) provides specific, clinically oriented recommendations about obtaining informed consent and disclosing incidental findings based on a comprehensive review of expert opinions relating to WGS. They propose a categorization system to classify genetic information related to the present or future effect of the variant and its actionability, carrier status, and penetrance (Table 3). The 6 categories are (a) findings relevant to the reason the test was performed, (b) clinically relevant variants for which treatment is or is not available, (c) variants causing high risk for future mendelian diseases, (d) carrier status that can impact reproductive life decisions, (e) variants of variable risk for future diseases, and (f) variants of unknown significance.

As with all genetic tests, pretest counseling and general information should be included in the informed consent form for WGS. Additionally, questions of disclosure of incidental findings and storage of genetic information need to be addressed. Genetic information that is of clinical relevance, even when not related to the reason for testing, and that will affect the patient, their relatives, and/or their offspring should always be disclosed for preventable or treatable diseases like Lynch syndrome, cancer-related genes such as breast cancer 1, early onset (*BRCA1*)[5] and breast cancer 2, early onset (*BRCA2*), and carrier status. In one study, patients were most interested in receiving WGS results for a preventable or treatable disease and determining carrier status (36). Information about risks for future

diseases that are nonpreventable and/or untreatable like Huntington disease and carrying the *APOE4* variant of the apolipoprotein E (*APOE*) gene (associated with Alzheimer disease) should be disclosed only with prior patient consent. When WGS is performed on minors, information about risks of nonpreventable and untreatable future diseases should be postponed until the individuals reach the legal age for consent and can make their own decisions based on the information. For genetic information patients do not wish to know about, the information can either be stored or destroyed. If stored, the information can potentially be disclosed in the future, but this raises the question of whether the patient or clinician is responsible to follow up on the information, especially in cases for which new information affecting actionability is discovered.

For disclosed incidental findings, posttest counseling must include discussion of indicated follow-up tests. Interpretation of incidental findings is complicated because even variants with evidence of impact on disease risk may have different risks in an asymptomatic population than they do in the population in which they were originally implicated. Even identification of a variant associated with a disease does not necessarily mean an individual has the disease (37). Follow-up studies for such false-positive incidental findings are potentially costly and invasive and may cause undue distress in patients. Moreover, they violate the medical imperative to do no harm.

Recently, the ACMG released recommendations for reporting of incidental findings in clinical exomes and WGS (38). This committee recommends that laboratories engaged in such clinical sequencing seek and report mutations in a specified group of genes that are known to be associated with disease (the list includes 57 genes associated with 24 diseases/syndromes). These recommendations apply to all normal and tumor specimens in all subjects irrespective of age, but excluding fetal samples. The committee defined incidental findings as "results of a deliberate search for pathogenic or likely pathogenic alterations in genes that are not apparently relevant to a diagnostic indication for which the sequencing test was ordered." It is important to mention here that the committee felt that it is the responsibility of the ordering clinician to provide comprehensive pre- and posttest counseling to the patient (38). The ACMG recommendations are controversial and have recently been contested because they may violate patient consent rights (39). The advantages and disadvantages of the guidelines have been discussed elsewhere (40). The guidelines will be annually reviewed and revised.

With the vast number of variants identified in any patient's sequenced genome, it is important to prioritize them for follow-up. Currently, algorithms to pri-

---

[5] Human genes: *BRCA1*, breast cancer 1, early onset; *BRCA2*, breast cancer 2, early onset; APOE, apolipoprotein E; *ERBB2* (formerly *HER2/Neu*), v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2; *CHD7*, chromodomain helicase DNA binding protein 7.

oritize variants, even when run on the same data, make different predictions about the effects of even common mutations (41). To prioritize variants with unknown disease associations on the basis of predicted effects will be even harder. New techniques use viruses and yeast growth over generations to elucidate the effects of hundreds of thousands of mutations in a single experiment (42, 43). The choice of experimental systems can lead to different results and thus multiple experimental systems (i.e., in vitro functional assays and model organisms such as zebrafish and mice) will be needed to investigate the functional significance of variants. In addition to experimental studies, statistical genetics and computational predictions can be used to help identify possible effects of poorly characterized variations and enable algorithms to more accurately prioritize them for follow-up (44). A recent tool to prioritize variants for follow-up, the GET-Evidence (Genome-Environment-Trait-Evidence) system, includes a peer production component for reaching consensus on variants with incorporation of phenotype information (37). It is based on the first 10 sequences of the Personal Genome Project, which aims to sequence the genomes of up to 100 000 individuals and integrate complete phenotype and other biological data (45). Other huge research projects such as the Genomes OnLine Database (46), which lists 12 000 NGS projects that are currently underway or will be shortly, are helping assign putative functions to variations.

One approach to deal with the amount of information generated is by using different approaches in managing variants depending on application. In identifying an etiology in symptomatic individuals, variant prioritizing algorithms should be tailored to identify only variants with demonstrated clinically relevant phenotypic implications that could then guide therapeutic interventions. If a definitive etiology cannot be obtained, variants with predicted effects related to patient symptoms can be prioritized for exploration in a research setting. In identifying asymptomatic individuals at high risk for preventable diseases, variants should be prioritized on the basis of strong evidence of association with preventable diseases, which would allow patient counseling to minimize risk. Triaging variants on the basis of targeted applications allows the most clinically significant variants to be discovered and used to guide patient counseling and treatment in a timely and cost-effective manner.

Interpreting the results from WGS is further complicated by the impact of ethnicity. Over the last several millennia, the exponential growth of the human population in a short time has resulted in an excess of rare variants that have not been subjected to natural selection (47). Because disease-causing alleles are distributed along ethnic lines (26), controls must be estab-

lished for ethnicity in interpreting the significance of variants. For example, it is possible that a variant could be population specific and not necessarily cause disease. More studies examining the impact of ethnicity on penetrance and phenotypic severity are needed in diverse ethnic populations.

There is a paucity of skilled geneticists and genetic counselors available to deal with the rapid growth of genetic testing. Clinicians have not been trained to provide genetic test counseling, in addition to not having the time and necessary resources. Medical schools are developing a formal curriculum to prepare future clinicians (48). The general population has also not been educated about the realities of WGS. An educational plan for clinicians and the population, in conjunction with reporting of test results with detailed interpretations and recommendations, will help bridge the knowledge gaps.

Ethical issues also arise on larger scales. Because genetic variations are increasingly associated with behavioral traits and psychiatric disorders (i.e., schizophrenia (49), depression (50), bipolar disorder (51)), uncertainty exists as to whether the information should be shared with the authorities if potentially dangerous behavioral traits or combinations of traits with psychiatric disorders are identified. Geneticists are examining the genome of Adam Lanza, the man who went on the Connecticut elementary school shooting spree, in search of a genetic cause that explains his actions (52). However, any identified genetic variant that may explain his actions ignores the critical impact of nongenetic factors (53).

It is important to ensure the confidentiality and privacy of WGS data. HIPAA (the Health Insurance Portability and Accountability Act), created by the US Congress, addresses privacy of patient data but it, and other laws and regulations, must be updated to reflect the complexity and sensitive nature of WGS. A recent report by the Presidential Commission for the Study of Bioethical Issues, *Privacy and Progress in Whole Genome Sequencing (54)*, has been released and provides recommendations about privacy protection and data security, while also maintaining data-sharing opportunities that will be important to help advance research.

Evidence for Efficacy

One of the public's areas of interest in the use of WGS is its ability to predict future disease risk in asymptomatic individuals. DTC genetic tests, which primarily use genotyping panels to assess a couple hundred SNPs to provide disease risk estimates, demonstrate the difficulties in providing meaningful results. Studies evaluating risk prediction among DTC vendors found substantial variability in estimates, except in the limited

number of SNPs with strong predictive value (i.e., the SNP for celiac disease) (55, 56). Different predictive models can change the accuracy of risk estimates, such as if/how literature-based estimates are combined with covariates. Further, the majority of tested SNPs provide only a modest change in an individual's risk estimates, and the influence of nongenetic factors (i.e., lifestyle) on the results is unknown for most of the tested SNPs. Roberts et al. (57) examined disease prevalence and the rates of concordance between monozygotic twins for 24 common diseases. In a best-case scenario that maximizes the potential clinical utility of WGS, the majority of individuals will have a clinically meaningful risk allele in at least 1 disease. However, for the remaining majority of diseases, the patients would receive negative results that would not meaningfully decrease their estimated risk for developing that disease. Prospective studies should be performed to evaluate the efficacy of disease risk prediction in altering health outcomes. Presently, the ability of genetic tests to reliably provide information for risk prediction is unreliable.

An immediate application of WGS in clinical practice is for the rapid examination of the etiology for complex diseases. WGS is more advantageous than single-gene sequencing or genotyping a series of known mutations because it can identify all variants in the entire genome, possibly increasing the speed of receiving care because only 1 test has to be performed (9). Whole exome sequencing (WES), which analyzes the exome (about 1% the size of the total genome) requires substantially less time for sequencing and analysis and can be performed at a substantially lower cost. WES has been used to identify the cause of a child's severe, intractable inflammatory bowel disease, in which other testing had not enabled the establishment of a diagnosis, allowing successful treatment (6). In an infant with acute liver failure, WES allowed improved medical decision-making by identifying mutations causing a recessive disorder, thus leading to counseling the parents that the infant would not be a suitable candidate for liver transplantation (58). WGS was used to diagnose congenital chloride diarrhea in a patient with a suspected diagnosis of Bartter syndrome (8). In another case, WGS has been used to find the gene responsible an unknown syndrome (59). Similar successes are now reported in an exponential fashion and more detailed coverage is beyond the scope of this review.

Pharmacogenomics uses knowledge of genetic variants to guide pharmacological treatment and selection of appropriate dosage, primarily in patients with oncological and hematological diseases, as well as in cardiac, pulmonary, rheumatologic, and infectious diseases (60). It was the earliest clinical implementation of genetic testing and remains an area in which WGS can provide clinically relevant information. A list of drugs with known pharmacogenomics information is described elsewhere (61). For example, evaluation of the status of the v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 (ERBB2, formerly HER2/Neu) status can stratify subsets of breast cancer patients who are likely to respond to trastuzumab. Targeted therapies are also available that are based on somatic variations for chronic myeloid leukemia, colorectal cancer, and lung cancer (61). WGS has been used to identify a cryptic fusion oncogene and direct the treatment plan in a patient with acute promyelocytic leukemia (9). In a patient with adenocarcinoma of the tongue, WGS was clinically used to identify the somatic driver mutation and direct the treatment to an inhibitor of the mutation (10). For nonneoplastic treatment decisions, WGS can be used to identify germline polymorphisms that confer sensitivity to warfarin and guide dosage (62). In another case WGS was used to guide therapy in twins diagnosed with dopa-responsive dystonia (63). Again, similar examples are now reported at an accelerated pace.

In reproductive health, WGS can be used for preconception carrier screening, preimplantation genetic diagnosis to select healthy embryos, prenatal diagnostic testing, and newborn screening. WGS can identify all genetic variations that previously either had no tests available or required multiple tests, such as the >3500 mendelian disorders with a known genetic basis (64) that contribute to disease and infant mortality (65). In contrast to current invasive prenatal screening options (i.e., amniocentesis and chorionic villus sampling), which pose a risk to the fetus, WGS can be used to noninvasively sequence the fetal genome by counting parental haplotypes in maternal plasma (66). Identification of any diseases will inform more appropriate prenatal counseling and treatment and postnatal care. A comparison of WGS strategies and current routine maternal serum screening protocols for Down syndrome confirmed the superior diagnostic sensitivity and specificity of WGS (67). Whole genome "jumping libraries" of a prenatal sample identified a translocation breakpoint that disrupted chromodomain helicase DNA binding protein 7 (CHD7), a causal locus in the CHARGE syndrome (coloboma of the eye, heart defects, atresia of the choanae, retardation of growth and/or development, genital and/or urinary abnormalities, and ear abnormalities and deafness), otherwise unidentifiable with current screening technology, the knowledge of which assists with prenatal counseling (68). In utero detection of conditions such as phenylketonuria, galactosemia, maple syrup urine disease, and severe combined immunodeficiency allows treatment to be immediately administered after birth, before the newborn manifests symptoms (66). In neonatal intensive care units, WGS can be used to rapidly identify

Fig. 2. Obstacles to be addressed to bring WGS into routine clinical use.
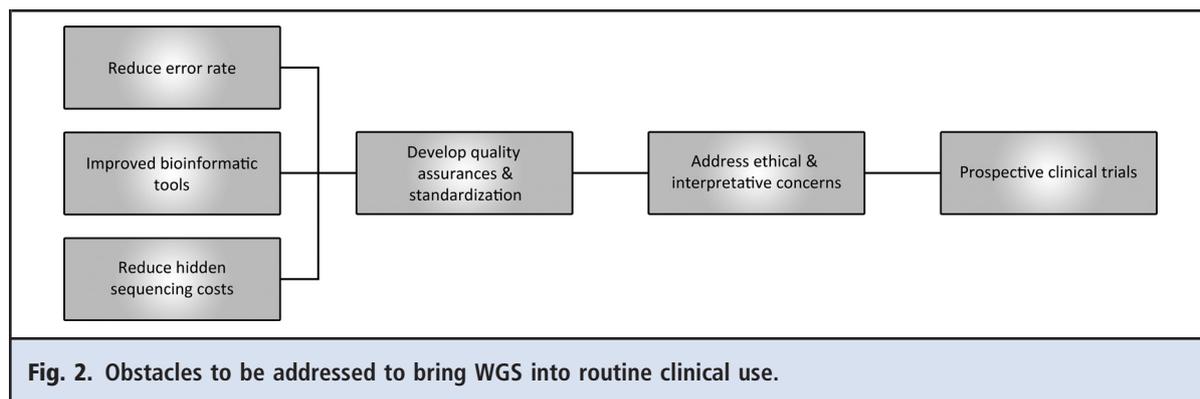
genetic disorders and guide treatment in neonates *(69)*.

Patenting Issues

An important concern for laboratories performing WGS is that they may infringe on hundreds of patents protecting the clinical significance of numerous sequence variations in the genome and thus face patent infringement liability. The US Supreme Court's recent ruling in the long-running case of Association of Molecular Pathology (AMP) v. Myriad Genetics on the eligibility of patenting isolated DNA serves to end concern about WGS being obstructed by genomic patents *(70)*. Myriad Genetics is a genetic testing company holding patents on the BRCA breast and ovarian cancer susceptibility genes and a diagnostic test to identify women at increased risk for those cancers. In the earlier stages of the case, the US Federal Circuit Appeals Court upheld the prior US District Court's ruling that Myriad's diagnostic method patent claim of comparing sequence variations in the BRCA gene to identify patient cancer risk was invalid *(71)*. The US Supreme Court found that isolated DNA is not patent eligible because it is a product of nature and not an act of invention. Myriad Genetics has used its patent rights to prevent other genetic testing companies and academic institutions from offering BRCA testing and engaging in research involving the genes. Now, other institutions can engage in testing and research without fear of litigation, increasing competition in the marketplace and driving test costs down. On the day of the Supreme Court's decision, several companies and university laboratories announced that they were offering BRCA mutation testing at lower costs than Myriad; however, Myriad has sued, alleging infringement on their remaining patents *(72)*.

The US Supreme Court ruled that isolated cDNA is patent eligible because introns are removed, making it novel and distinct from the DNA from which it was derived. Because the removal of introns is a scientifi-

cally routine activity, this contrasts with the Court's prior decision in the case of Mayo v. Prometheus, in which the Court ruled that combining a natural law (that individuals metabolize drugs differently) with a scientifically routine activity (measuring a metabolite concentration to adjust the dose of a thiopurine drug) was not patent eligible *(73)*. It remains to be seen how the Supreme Court's most recent decision will impact their prior one. The impact of the AMP v. Myriad case on protecting intellectual property and incentivizing genomic discovery is still uncertain.

Conclusions

WGS is a tremendously powerful test with important and diverse clinical utility. Nonetheless, there are multiple challenges preventing expanded clinical implementation of WGS, such as high costs beyond sequencing, the accuracy of NGS platforms and analysis algorithms, quality assurances, the ability to meaningfully interpret results, the need to address ethical concerns about incidental findings, and the lack of trained clinicians (Fig. 2). Targeted applications involved in processes such as identification of the genetic basis for complex diseases, pharmacogenomics, treatment personalization, and reproductive health represent areas in which WGS is currently being used successfully in clinical practice to guide treatment and improve patient outcomes, and where its expanded usage will have clinically significant benefits. Broader applications to implement WGS as a standard clinical practice, such as identifying disease risk in asymptomatic populations, will take a lot longer to reach the clinic. For such applications, prospective trials should be performed to show if, and for whom, the benefit of WGS outweighs the likely substantial costs associated with follow-up tests, risks of overdiagnosis and overtreatment, and the associated emotional distress. Until these obstacles can be surmounted, it seems prudent to offer WGS only for select diagnostic applications in which its continued

Review

and expanded clinical implementation maximizes the benefit to patients while minimizing harm.

References

1. Pasche B, Absher D. Whole-genome sequencing: a step closer to personalized medicine. JAMA 2011;305:1596–7.
2. Diamandis EP. Next-generation sequencing: a new revolution in molecular diagnostics? Clin Chem 2009;55:2088–92.
3. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. Clin Chem 2009;55:641–58.
4. Bloss CS, Darst BF, Topol EJ, Schork NJ. Direct-to-consumer personalized genomic testing. Hum Mol Genet 2011;20:R132–41.
5. Bellcross CA, Page PZ, Meaney-Delman D. Direct-to-consumer personal genome testing and cancer risk prediction. Cancer J 2012;18:293–302.
6. Diamandis M, White NM, Yousef GM. Personalized medicine: marking a new epoch in cancer patient management. Mol Cancer Res 2010;8:1175–87.
7. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, et al. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med 2011;13:255–62.
8. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci USA 2009;106:19096–101.
9. Welch JS, Westervelt P, Ding L, Larson DE, Klco JM, Kulkarni S, et al. Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. JAMA 2011;305:1577–84.
10. Jones SJ, Laskin J, Li YY, Griffith OL, An J, Bilenky M, et al. Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. Genome Biol 2010;11:R82.
11. Bick D, Dimmock D. Whole exome and whole genome sequencing. Curr Opin Pediatr 2011;23:594–600.
12. Drmanac R. Medicine. The ultimate genetic test. Science 2012;336:1110–2.
13. Pelak K, Shianna KV, Ge D, Maia JM, Zhu M, Smith JP, et al. The characterization of twenty sequenced human genomes. PLoS Genet 2010;6:e1001111.
14. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73.
15. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science 2012;335:823–8.
16. Zhao S, Prenger K, Smith L, Messina T, Fan H, Jaeger E, Stephens S. Rainbow: a tool for large-scale whole-genome sequencing data analysis using cloud computing. BMC Genomics 2013;14:425.
17. Ormond KE, Wheeler MT, Hudgins L, Klein TE, Butte AJ, Altman RB, et al. Challenges in the clinical application of whole-genome sequencing. Lancet 2010;375:1749–51.
18. Prevention Genetics. Clinical DNA testing price list. http://preventiongenetics.com/files/4613/8333/9106/PricelistByDisease_10-15-13.pdf (Accessed March 2013).
19. SEQanswers. Software list. http://seqanswers.com/wiki/Software/list (Accessed March 2013).
20. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol 2012;30:1033–6.
21. Kedes L, Campany G. The new date, new format, new goals and new sponsor of the Archon Genomics X PRIZE competition. Nat Genet 2011;43:1055–8.
22. Peters BA, Kermani BG, Sparks AB, Alferov O, Hong P, Alexeev A, et al. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. Nature 2012;487:190–5.
23. Sebastiani P, Solovieff N, Puca A, Hartley SW, Melista E, Andersen S, et al. Retraction. Science 2011;333:404.
24. Sebastiani P, Solovieff N, Puca A, Hartley SW, Melista E, Andersen S, et al. Genetic signatures of exceptional longevity in humans. Science 2010; 2010.
25. Nothnagel M, Herrmann A, Wolf A, Schreiber S, Platzer M, Siebert R, et al. Technology-specific error signatures in the 1000 Genomes Project data. Hum Genet 2011;130:505–16.
26. Moore B, Hu H, Singleton M, De La Vega FM, Reese MG, Yandell M. Global analysis of disease-related DNA sequence variation in 10 healthy individuals: implications for whole genome-based clinical diagnostics. Genet Med 2011;13:210–7.
27. Rieber N, Zapatka M, Lasitschka B, Jones D, Northcott P, Hutter B, et al. Coverage bias and sensitivity of variant calling for four whole-genome sequencing technologies. PLoS One 2013;8:e66621.
28. Ledergerber C, Dessimoz C. Base-calling for next-generation sequencing platforms. Brief Bioinform 2011;12:489–97.
29. Huang X, Madan A. CAP3: a DNA sequence assembly program. Genome Res 1999;9:868–77.
30. Hatem A, Bozdağ D, Toland AE, Çatalyürek ÜV. Benchmarking short sequence mapping tools. BMC Bioinformatics 2013;14:184.
31. Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat Rev Microbiol 2012;10:599–606.
32. Pang AW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF, et al. Towards a comprehensive structural variation map of an individual human genome. Genome Biol 2010;11:R52.
33. Michaelson JJ, Sebat J. forestSV: structural variant discovery through statistical learning. Nat Methods 2012;9:819–21.
34. Roach JC, Glusman G, Hubley R, Montsaroff SZ, Holloway AK, Mauldin DE, et al. Chromosomal haplotypes by genetic phasing of human families. Am J Hum Genet 2011;89:382–97.
35. Ayuso C, Millán JM, Mancheño M, Dal-Ré R. Informed consent for whole-genome sequencing studies in the clinical setting. Proposed recommendations on essential content and process. Eur J Hum Genet 2013;21:1054–9.
36. Facio FM, Eidem H, Fisher T, Brooks S, Linn A, Kaphingst KA, et al. Intentions to receive individual results from whole-genome sequencing among participants in the ClinSeq study. Eur J Hum Genet 2013;21:261–5.
37. Ball MP, Thakuria JV, Zaranek AW, Clegg T, Rosenbaum AM, Wu X, et al. A public resource facilitating clinical use of genomes. Proc Natl Acad USA 2012;109:11920–7.
38. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 2013;15:565–74.
39. Wolf SM, Annas GJ, Elias S. Point-counterpoint. Patient autonomy and incidental findings in clinical genomics. Science 2013;340:1049–50.
40. McGuire AL, Joffe S, Koenig BA, Biesecker BB, McCullough LB, Blumenthal-Barby JS, et al. Point-counterpoint. Ethics and genomic incidental findings. Science 2013;340:1047–8.
41. Hicks S, Wheeler DA, Plon SE, Kimmel M. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. Hum Mutat 2011;32:661–8.
42. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, Fields S. High-resolution mapping of protein sequence-function relationships. Nat Methods 2010;7:741–6.

43. Hietpas RT, Jensen JD, Bolon DN. Experimental illumination of a fitness landscape. Proc Natl Acad Sci USA 2011;108:7896–901.

44. Sunyaev SR. Inferring causality and functional significance of human coding DNA variants. Hum Mol Genet 2012;21:R10–7.

45. Jones B. Genomics: personal genome project. Nat Rev Genet 2012;13:599.

46. Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, et al. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 2012;40:D571–9.

47. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 2012;337: 64–9.

48. Walt DR, Kuhlik A, Epstein SK, Demmer LA, Knight M, Chelmow D, et al. Lessons learned from the introduction of personalized genotyping into a medical school curriculum. Genet Med 2011;13:63–6.

49. Rietschel M, Mattheisen M, Degenhardt F; Genetic Risk and Outcome in Psychosis (GROUP Investigators), Mühleisen TW, Kirsch P, et al. Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. Mol Psychiatry 2012;17: 906–17.

50. Schosser A, Butler AW, Ising M, Perroud N, Uher R, Ng MY, et al. Genomewide association scan of suicidal thoughts and behaviour in major depression. PLoS One 2011;6:e20690.

51. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–78.

52. abc News. School shooter's DNA to be studied. http://news.yahoo.com/school-shooters-dna-studied-214930321--abc-news-topstories.html (Accessed March 2013).

53. No easy answer. Nature 2013;493:133.

54. Presidential Commission for the Study of Bioethical Issues. Privacy and progress in whole genome sequencing. http://bioethics.gov/cms/node/764 (Accessed March 2013).

55. Zehnbauer B. Direct-to-consumer genetics testing—fair comparisons? Clin Chem 2011;57:369–71.

56. Imai K, Kricka LJ, Fortina P. Concordance study of 3 direct-to-consumer genetic-testing services. Clin Chem 2011;57:518–21.

57. Roberts NJ, Vogelstein JT, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE. The predictive capacity of personal genome sequencing. Sci Transl Med 2012;4:133ra58.

58. Goh V, Helbling D, Biank V, Jarzembowski J, Dimmock D. Next-generation sequencing facilitates the diagnosis in a child with twinkle mutations causing cholestatic liver failure. J Pediatr Gastroenterol Nutr 2012;54:291–4.

59. Mendel's Pod. Father/scientist finds gene responsible for daughter's unknown syndrome: Hugh Rienhoff talks personal genomics. http://www.mendelspod.com/podcast/father-scientist-finds-gene-responsible-for-daughters-unknown-syndrome-hugh-rienhoff-talks-personal-genomics (Accessed March 2013).

60. U.S. Food and Drug Administration. Table of pharmacogenomic biomarkers in drug labeling. http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm (Accessed March 2014).

61. Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ. Next generation sequencing in clinical medicine: challenges and lessons for pathology and biomedical informatics. J Pathol Inform 2012;3:40.

62. Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, et al. Clinical assessment incorporating a personal genome. Lancet 2010;375: 1525–35.

63. Bainbridge MN, Wiszniewski W, Murdock DR, Friedman J, Gonzaga-Jauregui C, Newsham I, et al. Whole-genome sequencing for optimized patient management. Sci Transl Med 2011;3:87re3.

64. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM). Nucleic Acids Res 2009;37:D793–6.

65. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci Transl Med 2011;3: 65ra4.

66. Fan HC, Gu W, Wang J, Blumenfeld YJ, El-Sayed YY, Quake SR. Non-invasive prenatal measurement of the fetal genome. Nature 2012;487: 320–4.

67. Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, Ehrich M, et al. DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. Genet Med 2011;13:913–20.

68. Talkowski ME, Ordulu Z, Pillalamarri V, Benson CB, Blumenthal I, Connolly S, et al. Clinical diagnosis by whole-genome sequencing of a prenatal sample. N Engl J Med 2012;367:2226–32.

69. Saunders CJ, Miller NA, Soden SE, Dinwiddie DL, Noll A, Alnadi NA, et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. Sci Transl Med 2012; 4:154ra135.

70. Kesselheim AS, Cook-Deegan RM, Winickoff DE, Mello MM. Gene patenting—the Supreme Court finally speaks. N Engl J Med 2013;369:869–75.

71. Association for Molecular Pathology v. United States Patent and Trademark office and Myriad Genetics, 689 F. 3d 1303 Fed. Cir. (2012).

72. New York Times. 2 Competitors sued by genetics company for patent infringement. http://www.nytimes.com/2013/07/11/business/2-competitors-sued-by-genetics-company-for-patent-infringement.html?_r=0 (Accessed July 2013).

73. Mayo Collaborative Services v. Prometheus Laboratories, Inc., 566 U.S., 132 S. Ct. 1289 (2012).

74. The University of Texas at Austin. Illumina's historical list prices. https://wikis.utexas.edu/display/GSAF/Illumina's+historical+list+prices (Accessed August 2013).