The NCGENES Project:
Exploring the New World of Genome Sequencing

Ann Katherine M. Foreman, Kristy Lee, James P. Evans

Massively parallel sequencing (MPS) is now a clinical reality, promising improved diagnosis, targeted therapies, and population-based screening. To realize the potential of genomics, we must learn how to apply this technology optimally. The NCGENES project is designed to address several challenges that must be overcome in order to integrate MPS into clinical care.

On April 14, 2003, the Human Genome Project was completed, resulting in a high-quality human reference genome and the promise of a new era of genomic medicine. Massively parallel sequencing (MPS), also known as next-generation (NextGen) sequencing, has made the rapid sequencing of vast quantities of DNA practical. The technology has transitioned rapidly to the bedside, where whole-genome sequencing (WGS) and whole-exome sequencing (WES) are now being pursued widely in a research context and are clinically available from a few laboratories nationally. WES and WGS have fueled excitement about the potential for improved diagnosis of genetic disease, targeted cancer therapies, and even population-based screening. Although the promise of genomics is great, its application presents a multitude of challenges, which stem from the massive amounts of heterogeneous data generated when a patient's genome is analyzed.

Ideally, the clinical implementation of genomic medicine will be guided by evidence-based best practices. The National Human Genome Research Institute, an institute of the National Institutes of Health, has developed the Clinical Sequencing Exploratory Research (CSER) program to investigate the optimal integration of MPS into clinical practice. The North Carolina Clinical Genomic Evaluation by NextGen Exome Sequencing (NCGENES) project, a grantee of the CSER program, is an effort to systematically study the most critical challenges in the integration of genomic medicine into clinical care.

The human genome is simply the total complement of DNA in a person. It includes *exons*—those parts of our genes that encode proteins—as well as noncoding *introns* and the regions of DNA between genes. The *exome* is the totality of those parts of our genome that encode proteins; it constitutes approximately 1% of the human genome [1]. Analysis of the exome is less costly than WGS (Table 1) and provides most of the clinically relevant information, because the vast majority of harmful variants or mutations that cause genetic diseases reside in the exome. Thus WES is currently used more often than WGS in clinical settings.

TABLE 1.
A Comparison of Traditional Sanger Sequencing, Whole-Exome Sequencing, and Whole-Genome Sequencing

Variable	Traditional Sanger sequencing	Whole-exome sequencing (WES)	Whole-genome sequencing (WGS)
Scope of test	1 gene or panel of genes for a specified phenotype	All protein-encoding genes	All protein-encoding genes and intervening DNA
Total number of variants	Few; dependent on the number of genes sequenced	~80,000–100,000	~3–4 million
Scope of results	Includes only information pertinent to gene(s) requested	Includes results pertinent to clinical indication and possibly significant incidental findings	Includes results pertinent to clinical indication and possibly significant incidental findings; higher likelihood of uncertain results
Turnaround time	~1–12 weeks	~10–15 weeks for targeted analysis; ~24–28 weeks for complete analysis	~12 weeks
Cost	~$500–$17,000, depending on the number of genes sequenced	~$4,500–$6,000	~$9,500

Note. WES and WGS are already cost-competitive with Sanger sequencing panels; however, the 3 types of sequencing differ greatly in the scope of testing and the number of resulting variants.

Genome-scale sequencing has been made possible through advances in technology that allow millions of sequencing reactions to occur simultaneously on a single microchip or flow cell—hence the term *massively parallel sequencing* (Figure 1). DNA derived from a blood sample

Electronically published November 19, 2013.

Address correspondence to Dr. James P. Evans, Department of Genetics, CB 7264, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7264 (jpevans@med.unc.edu).

N C Med J. 2013;74(6):500-504. ©2013 by the North Carolina Institute of Medicine and The Duke Endowment. All rights reserved.
0029-2559/2013/74610

Crowdsourcing to Define the Clinical Actionability of Incidental Findings of Genetic Testing

Matthew S. Krantz, Jonathan S. Berg

Genome-scale sequencing may soon be cheaper than targeted assays as a clinical diagnostic tool. However, these larger queries will turn up many incidental findings—that is, unanticipated information discovered during the course of testing. Implementation of genome-scale sequencing in the clinical setting will require novel methods for managing these incidental findings.

In order to grapple with our ever-changing knowledge of genetic disease and to make recommendations regarding minimal standards for reporting of incidental findings, a "binning" system for the management of such findings has been proposed; this system is based on both clinical utility and clinical validity [1]. Clinically actionable incidental findings (eg, Marfan disease) go into bin 1; these incidental findings are likely to be rare and would be reported to an individual because of high penetrance and the existence of evidence-based management recommendations. Incidental findings that have clinical validity but no clinical actionability go into bin 2; these findings require informed decision making on the part of the individual and would only be reported to the patient on request. In these cases, there may be a strong association between genotype and phenotype, but no immediate intervention exists; an example is *APOE* risk alleles that are associated with the onset of Alzheimer disease. As effective evidence-based interventions emerge for bin 2 variants, their classification will change. Finally, most genes currently have no known clinical significance. Variants of these genes go into bin 3, and these incidental findings would not be reported.

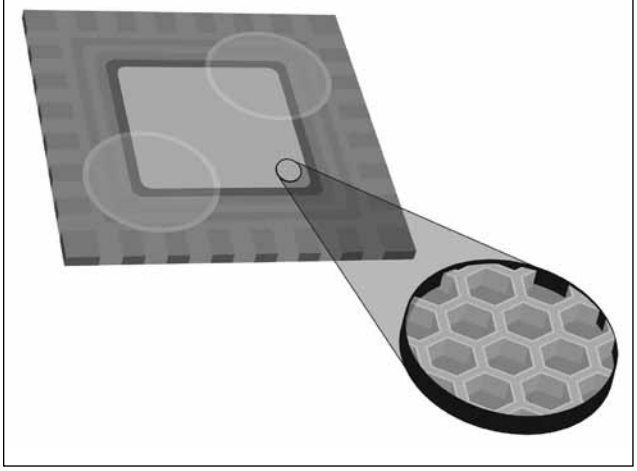
Although not strictly analogous, the bin 1 variants described above are similar to the incidental findings included on the minimal reporting list published by the American College of Medical Genetics and Genomics (ACMG) [2]. The ACMG working group that developed this list noted that their recommendations were hampered by lack of data on clinical utility and the need for a process by which recommendations could be updated regularly. The binning system described here tackles the lack of data on clinical utility and creates a process amenable to revision as new literature emerges.

So far, binning of genes has been provisional [3] and has been based on consensus among a small number of evaluators. In order to refine this approach into a more fully transparent, reproducible, evidence-based process, members of the North Carolina Clinical Genomic Evaluation by NextGen Exome Sequencing (NCGENES) team developed a semiquantitative metric that scores an incidental finding based on several key criteria: threat to health, chance of disease, efficacy of intervention, intervention acceptability, and knowledge base. This process yields a minimum total score of 0 (for genes with no known clinical relevance) and a maximum score of 15. A higher total score on this semiquantitative metric correlates with a greater degree of clinical actionability, which will hopefully allow us to

or saliva sample is sheared into fragments and processed to form a genomic "library." The processed fragments are loaded onto test chips for sequencing, and then each of the millions of sequenced fragments is mapped to the human reference genome, allowing assembly of an individual's genome. Once assembled, all variants from the reference sequence are interpreted with respect to whether they are innocent variations or whether they might cause disease. The magnitude of this task can be appreciated by noting that each person harbors approximately 100,000 variants that would be detected by exome sequencing. Most of these variants are innocuous, but a small minority might be responsible for disease. Identifying the few needles of clinically useful information in this haystack of data is a complex undertaking. Initial analysis is automated using computer software that classifies variants by type (eg, nonsense, missense, synonymous, or splice site variants). Interpretation is further aided by information such as how often a variant is found in the general population and prediction models that help assess the possible biological implications of variants. However, interpretation of a WES or WGS test still requires that experienced personnel spend substantial time combing through pertinent literature and databases to determine whether individual variants might be relevant to a patient's disease.

As is the case with any complex medical technology, the clinical potential of genomics cannot be fully realized until

FIGURE 1.
Massively Parallel Sequencing Involves the Use of Microchips That Can Perform Millions of Sequencing Reactions Simultaneously



set thresholds for determining placement of an incidental finding into bin 1, 2, or 3.

Reaching consensus regarding medical actionability will require multiple evaluations, and it is neither feasible nor desirable for a single working group to score all genome-scale incidental findings. Instead, having a robust number of evaluators will allow for greater diversity of opinion and expertise.

Crowdsourcing has been shown to be a powerful tool for answering scientific questions that require a wide array of input. Crowdsourcing employs distributed problem solving by engaging the public through open-source interfaces. Proteomics research has been accelerated by utilizing the collective intelligence of the crowd through the online game Foldit, in which players attempt to solve protein structures. Some compelling successes have been achieved using this approach; for instance, players modeled the crystal structure of the Mason-Pfizer monkey retroviral protease, which has provided insights that may be useful in the development of anti-HIV drugs [4].

Other areas of genetic research have also harnessed crowdsourcing. SNPedia is an online, open-access wiki project that allows users to input information about single nucleotide polymorphisms obtained from peer-reviewed journals into a computer-friendly format [5]. In another example of crowdsourcing, the Personal Genome Project aims to pair genomic and health data supplied by participants. The project is approved to study 100,000 participants and shares all information in the public domain, making it available for research [6]; to date, more than 1,800 people are enrolled. Although efforts are made to remove personal identifiers from the data, the Personal Genome Project operates under the premise of open consent, meaning that participants are not given promises of privacy, confidentiality, or anonymity.

We propose that the scoring of incidental findings using a semiquantitative metric could also be amenable to crowdsourcing. Defining medical actionability through crowdsourcing allows multiple annotators to provide scores for a gene-phenotype pair, and information can be updated as new evidence emerges. More genetic conditions will likely become medically actionable over time. It is essential, then, to choose an evaluation process that is highly adaptable to evolving medical research; crowdsourcing offers this flexibility. NCMJ

Matthew S. Krantz, BS medical student, Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.
Jonathan S. Berg, MD, PhD assistant professor, Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

Acknowledgment

Potential conflicts of interest. J.S.B. is a co-primary investigator for NCGENES. M.S.K has no relevant conflicts of interest.

References

1. Berg JS, Khoury MJ, Evans JP. Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time. Genet Med. 2011;13(6):499-504.
2. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15(7):565-574.
3. Berg JS, Adams M, Nassar N, et al. An informatics approach to analyzing the incidentalome. Genet Med. 2013;15(1):36-44.
4. Khatib F, DiMiao F; Foldit Contenders Group; Foldit Void Crushers Group, et al. Crystal structure of a monomeric retroviral protease solved by protein folding game players. Nat Struct Mol Biol. 2011;18(10):1175-1177.
5. Ball MP, Thakuria JV, Zaranek AW, et al. A public resource facilitating clinical use of genomes. Proc Natl Acad Sci U S A. 2012;109(30):11920-11927.
6. Cariaso M, Lennon G. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. Nucleic Acids Res. 2012;40(Database issue):D1308-D1312.

Electronically published November 19, 2013.

Address correspondence to Dr. Jonathan S. Berg, 5092 Genetics Medicine Building, CB 7264, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7264 (jonathan_berg@med.unc.edu).

N C Med J. 2013;74(6):501-502. ©2013 by the North Carolina Institute of Medicine and The Duke Endowment. All rights reserved.
0029-2559/2013/74624

we learn how to apply it optimally. This task is complicated by the limited scientific understanding of the impact of genetic variation on health. For the time being, WES will be most fruitful if it is performed in patients who are suspected of having a disorder that primarily arises from disruption of a single gene and who have defied diagnosis by standard means. Good candidates for WES include undiagnosed patients who exhibit a strong family history of the enigmatic disorder; those who appear to have a genetic disorder that can be caused by disruption of a variety of different genes; and those who exhibit complex patterns that are likely to result from a genetic disruption, such as developmental disorders or progressive neurological conditions. In the case of cancer, a strong genetic component to disease would be suggested by a young age at diagnosis or other unusual features (male breast cancer, for instance).

Even when several of these indicators are present, a more specific single-gene test or limited test panel will often be a more appropriate choice than WES or other broad, MPS-based genetic tests. For example, consider a 25-year-old woman who was recently diagnosed with bilateral renal cysts and whose family history includes a father who died of end-stage renal disease related to renal cysts; in this case, sequencing of the 2 genes responsible for autosomal dominant polycystic kidney disease, *PKD1* and *PKD2*, is likely to identify the underlying reason for the patient's disease more quickly and more cost-effectively than WES or WGS. MPS is likely to be more advantageous in cases of genetic heterogeneity, in which a patient's phenotype could be caused by a mutation in any one of a large number of genes. For example, more than 12 genes have been associated with long QT syndrome, and comprehensive testing for this condition using traditional sequencing can be a lengthy and costly endeavor.

NCGENES is seeking to define optimal applications of

WES by analyzing approximately 750 children and adults with suspected Mendelian genetic disorders who have eluded diagnosis by traditional means. Care has been taken to enroll participants with a variety of indications, including cancer, cardiogenetic diseases, neurodevelopmental disorders, and retinal diseases. Thus data from NCGENES will help to identify which types of patients are most likely to benefit from application of WES.

Costs of genomic analysis are expected to become comparable to the cost of single-gene tests in the near future. When that happens, there still may be compelling reasons to select a test with a specific focus rather than sequencing all of a person's genes. The breadth of information that WES yields will often result in false positives, identification of genetic variants of uncertain clinical significance (VUS), and incidental findings that are unrelated to the reason for ordering the test. These challenges must be understood before genome-scale sequencing can truly become a routine part of clinical care. In some circumstances, and for some patients, a traditional genetic test with a narrower focus will remain the best choice.

The chance of identifying a VUS increases with each gene analyzed, and identification of a VUS is virtually guaranteed when obtaining a genome-level test. The presence of a VUS essentially represents a potential false-positive result and can be confusing to patients and practitioners alike. Harm can ensue if a VUS is inappropriately treated like a positive finding; for example, follow-up of a VUS could prompt unnecessary screening or intervention, which exposes the patient to the risks of needless intervention and increases costs to the patient and the health care system. Even when providers appropriately recognize that medical decisions cannot be based on a VUS and offer genetic counseling, patients still may ascribe meaning to the variant, thus affecting their perception of their own health. In NCGENES, VUS are reported to participants only if they are considered a possible explanation for the participant's primary health concern; incidental VUS are not reported. The NCGENES team includes social scientists who will study the ethical and social implications of WES, including what impact, if any, the finding and reporting of a VUS has on participant perceptions of health compared with the impact of positive or negative diagnostic findings.

Incidental findings are not new in medicine; for example, unexpected tumors are sometimes identified on medical imaging. However, implementation of WES in clinical medicine invites incidental findings on a larger scale than has previously been seen in medicine. The majority of variants identified by WES or WGS will not be directly relevant to a patient's disease. Although the vast majority of these variants will have no clinical significance, it is important to recognize that a small portion of individuals undergoing WES will have incidental findings (ie, findings unrelated to the reason for sequencing) that have profound clinical importance. If variants influencing reproductive risk are considered (for instance, carrier status for recessive diseases such as cystic fibrosis or sickle-cell disease), then every person who has a WES or WGS test performed will have genetic variants identified that are clinically valid yet unrelated to the indication for testing and thus incidental to the purpose of the test. In an effort to begin to define the obligations of laboratories and clinicians, the American College of Medical Genetics and Genomics has published general guidelines regarding those genes that should be routinely examined for deleterious mutations when genome-scale sequencing is performed in a clinical context [2].

About 1% of individuals undergoing WES are found to have an unexpected mutation (ie, an incidental finding) that confers a high risk of severe disease that can be prevented or largely mitigated by medical intervention (eg, incidental discovery of a *BRCA1* mutation). Table 2 provides examples of such genes and the conditions they strongly predispose to when mutated. At the present time, this is a small group of genes, and it is recommended that the data of patients who undergo WES be examined for mutations in these genes in order to detect profound predisposition to preventable but

TABLE 2.
Examples of Medically Actionable Incidental Findings From Genetic Sequencing

Genetic syndrome	Genes that may have mutations	Likely adverse outcome if untreated	Typical intervention
Hereditary breast and ovarian cancer	*BRCA1, BRCA2*	Breast cancer	Prophylactic bilateral mastectomy or surveillance every 6 months alternating between mammogram and breast MRI
Lynch syndrome	*MLH1, MSH2, MSH6, PMS2*	Colon cancer	Colonoscopy with removal of polyps every 12–24 months
Hypertrophic cardiomyopathy	*MYBPC3, MYH7, TNNT2, TNN13, TPM1, MYL3, ACTC1, PRKAG2*	Sudden cardiac death	Placement of implantable cardioverter defibrillator
Long QT syndrome	*KCNQ1, KCNH2, SCN5A*	Sudden cardiac death	Beta blockers and periodic ECG
Familial hypercholesterolemia	*LDLR, APOB, PCSK9*	Myocardial infarction	Twice yearly cholesterol screening, with statins to be prescribed when hypercholesterolemia is documented

Note. ECG, electrocardiogram; MRI, magnetic resonance imaging.
These are examples of genetic conditions that are recommended to be reported as incidental findings if known disease-causing variants are identified in the genes listed. Such genes may also be candidates in the future for routine screening in the general population to prevent disease.

unexpected disease [2]. All NCGENES participants' exomes are analyzed for such variants regardless of the original indication for WES. If a mutation is found in one of these genes, results are returned to the research participant (or his or her parent in the case of minors). NCGENES participants with medically actionable incidental results will be interviewed about their experience of receiving such a result.

For most of the variants discovered through WES, the medical response is not clear. For example, carrier status for recessive diseases does not have direct relevance to an individual's health and may have little or no personal relevance if they have no plans for future children. On the other hand, mutations in the *PSEN1* gene cause a highly penetrant, early-onset form of Alzheimer disease, which is relevant to health; however, there are no known interventions that can modulate the disease outcome. How many of us would want to know that we were all but guaranteed to develop Alzheimer disease by age 65 years? Would we even want to be offered that type of information? NCGENES is seeking to answer these questions by finding out what other information in the genome (besides medically actionable and diagnostic information) is desired by patients who undergo WES.

To address these questions, adult participants in NCGENES are randomized into 1 of 2 groups. Individuals in the control group receive diagnostic results (ie, results related to the suspected genetic condition for which they were referred to the study) and medically actionable information. Those in the experimental group likewise receive diagnostic results and medically actionable information, plus they will be asked to decide whether they want various other categories of incidental information, such as carrier status for recessive diseases, and whether they want to know about variants that affect the risk of Alzheimer disease and other conditions [3]. By studying participants in the experimental group and the extent to which they seek such information, and by comparing the control and experimental groups with respect to their satisfaction with the decision to undergo WES and their personal perception of health, this study will help guide future decisions about how to handle non–medically actionable incidental findings that arise during clinical WES and WGS.

Genomic sequencing holds great promise and presents significant challenges in the clinical arena. These challenges are heightened with regard to genomic approaches to public health [4]. It is unlikely that individuals in the general population will benefit anytime soon from WES or WGS. However, one can readily imagine that it might be useful to screen members of the general population for mutations in carefully selected genes that confer a very high risk of severe but preventable or treatable disorders, such as colon cancer or breast cancer [4]. Additional studies to address the potential of such efforts are under way now at the University of North Carolina at Chapel Hill.

The world of genomics is moving quickly and will play a growing role in the care of patients. It is vital that we work together to define best practices for implementation of this new, promising, and highly complex technology. NCMJ

Ann Katherine M. Foreman, MS, CGC clinical assistant professor, Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.
Kristy Lee, MS, CGC clinical assistant professor, Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.
James P. Evans, MD, PhD clinical professor, Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

Acknowledgments

The authors wish to thank Noah H. Foreman for his assistance in the creation of Figure 1.

Potential conflicts of interest. A.K.M.F. and K.L. are molecular analysts and genetic counselors for NCGENES, and J.P.E. is a principal investigator for NCGENES.

References

1. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009;461(7261):272-276.
2. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15(7):565-574.
3. Berg JS, Khoury MJ, Evans JP. Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time. Genet Med. 2011;13(6):499-504.
4. Evans JP, Berg JS, Olshan AF, Magnuson T, Rimer BK. We screen newborns, don't we?: realizing the promise of public health genomics. Genet Med. 2013;15(5):332-334.