From Sequence Data to Returnable Results: Ethical Issues in Variant Calling and Interpretation

Ingrid A. Holm,^{1,2} Timothy W. Yu,^{1–3} and Steven Joffe^{4–6}

A discussion of return of genetic research results requires a common understanding of how final results are generated and what the scope of potential results may be. To this end, we provide a brief overview of the steps by which human genomic data, whether in the clinical or research setting, are generated and interpreted. We cover (1) DNA targeting methods, (2) sequencing, (3) mapping, (4) variant calling, (5) annotation, and (6) interpretation. As powerful as this technology is, we point out technical, scientific, and clinical limitations that inject uncertainty into interpretations based on genotypic data alone. Given these considerations, we then discuss ethical issues that arise as decisions are made regarding how human genomic data are generated and interpreted in the research setting, and we propose an ethical framework by which researchers can assert policies at the points of control that maximize rewards, while minimizing risks.

Keywords: ethics, gene variants, sequencing analysis

Introduction

NO RECENT CLINICAL LABORATORY advance has drawn greater enthusiasm and scrutiny than the advent of genome sequencing. This attention is deserved; compared to, say, checking a patient's thyroid hormone level, decoding an individual's genome can yield results with potential health ramifications for her or his lifetime. At the same time, discussions of return of research results from genome sequencing ought also to reflect that the science and art in this field are still evolving, both at the technical and interpretive levels. To this end, we provide a brief primer on genome sequencing, highlighting current practices as well as practical limitations along the way. We then briefly describe the ethical implications of the many choices that must be made when sequencing a genome and the uncertainties that pervade sequencing results.

Indications for Sequencing

The human genome is about 3 billion base pairs and includes about 25,000 genes. Any two individuals share about 99.9% of their DNA—the last 0.1% (i.e., 3 million base pairs) explains the diversity in the human population. The great majority (>97%) of those 3 million variants are common,

meaning that they are found in 1% or more of the human population. Rare variants, that is, those seen in 1% or less of the population, constitute ~3% of the variants observed in any given individual.

Although this article is focused on sequencing for research purposes, which is usually targeted toward gene discovery, most genome sequencing performed today is for (1) the purpose of diagnosing rare genetic conditions, which involves detecting rare, gene-disrupting variants, or (2) cancer diagnostics, which involves detecting recurrent acquired genetic alterations in tumors. Sequencing of somatic (tumor) DNA for cancer diagnostics may or may not include parallel analysis of germline DNA (Raymond *et al.*, 2016).

Sequencing Technologies

The sequencing of DNA has evolved rapidly since completion of the human genome project in 2001 (Lander *et al.*, 2001; Venter *et al.*, 2001). Milestones over the past 15 years include completion of the first single individual's genome (Venter *et al.*, 2001) in 2007 (Levy *et al.*, 2007), James Watson's genome (Wheeler *et al.*, 2008) as well as the first Chinese (Wang *et al.*, 2008) and Yoruban genomes (Bentley *et al.*, 2008) in 2008, the Korean genome (Kim *et al.*, 2009) and the first 12 human exomes (Ng *et al.*, 2010) in 2009,

¹Division of Genetics and Genomics, The Manton Center for Orphan Diseases Research, Boston Children's Hospital, Boston, Massachusetts.

²Department of Pediatrics, Harvard Medical School, Boston, Massachusetts.

³Broad Institute of MIT/Harvard, Cambridge, Massachusetts.

⁴Department of Medical Ethics and Health Policy, and ⁵Leonard Davis Institute for Health Economics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania.

⁶Division of Oncology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania.

followed by publication of 200 exomes (Ng *et al.*, 2010), and the 1000 genomes project (1000 Genomes Project Consortium *et al.*, 2010) in 2010. Sequencing costs have plummeted from \$2.7 billion for the first human genome (which took over 10 years to complete) to <\$1400 per genome in 2015 (www.genome.gov/sequencingcosts). It is important to note, however, that these figures include only the technical cost of generating the sequence, and exclude the costs associated with annotating, interpreting, and reporting medically relevant results to patients or research participants.

The rapid evolution of sequencing technologies over the past 10 years has made these advances possible [see Rizzo and Buck (2012) for a review]. The previous standard, semiautomated “Sanger sequencing” has been used in clinical testing for many years. Sanger sequencing produces 400–800 bp of high-quality genomic sequence at a time, and remains the gold standard for confirmation of findings found through next-generation sequencing (NGS). Although Sanger sequencing is the most accurate method for sequencing is currently available, the major limitation of Sanger sequencing is throughput, one machine can only carry out a limited number of reactions (typically 1–100) simultaneously (Rizzo and Buck, 2012).

NGS, also known as massively parallel sequencing, involves machines that have employed sophisticated miniaturization to be able to analyze many DNA sequences at one time. The result is vastly increased throughput. A wide variety of technologies are used, but most involve the generation of millions to billions of short (typically 20–250 bp) overlapping sequence reads, which can then be compared to one another as well as to a reference sequence and aligned (“mapped”) to reconstruct a contiguous consensus “read” of the DNA. A variant is detected when enough of the reads do not match the reference sequence at one or more base positions.

At any given base position in an NGS experiment, the “read depth” (i.e., the number of overlapping shorter sequence reads that include that base pair) can vary dramatically. Since individual NGS reads are more error prone than individual Sanger sequences (error rates $\sim 1/100$ – $1/1000$ vs. $\sim 1/10,000$), to be comfortable that the base call at any given position is correct, a minimum read depth is typically required (thresholds vary depending on the specific platform, but typically at least ~ 8 – 10 overlapping sequences covering the position). Inadequate read depth at a given site is one of the major limitations of any given NGS experiment, and can lead to both false negatives and false positives. This is why variants found using NGS typically need to be confirmed by Sanger sequencing before being considered definitive.

Another potential source of error in NGS is that the mapping to the reference genome may not be exact. This can be due to a variety of reasons, including “pseudogenes” (areas of the DNA that look like a particular gene, but are slightly different), repetitive sequences, and regions of duplications or deletions, all of which naturally occur in all individuals. At read lengths typically employed in NGS, 5–10% of the genome cannot be confidently mapped due to these factors [reviewed in Li and Freudenberg (2014)].

Finally, variant calling also depends on the type of variant that is being ascertained. Single base pair substitutions ($\sim 90\%$ of the total burden of genetic variation in an individual), the most frequently observed type of variant, are well detected, and sensitivities upwards of 95% with good speci-

ficity can be achieved (Zook *et al.*, 2014). Small (1–50 bp in size) insertions and deletions ($\sim 10\%$ of total) are harder to detect [sensitivities of 50–80% with more frequent false positives (Zook *et al.*, 2014)]. Larger insertions and deletions and more complicated alterations, like inversions and translocations, are more difficult still, and are considered out of scope by many clinical NGS providers.

In the future, some of these technical limitations may be addressed with newer, “third-generation” technologies (e.g., PacBio, Oxford nanopore) (Goodwin *et al.*, 2016); these employ single-molecule approaches to generate read lengths that can measure up to several 1000 bases in length. In principle, longer sequencing reads solve many problems, including allowing sequencing of otherwise inaccessible repetitive elements in the human genome, delineation of complicated structural variants, and establishment of phase. Practically, however, these technologies currently come with significant trade-offs, high intrinsic error rates and additional analytic complexity, making them unsuitable for standalone clinical laboratory use. [Hybrid sequencing strategies, making combined use of long- and short-read technologies, are in development (Rhoads and Au, 2015), but are out of the scope of this review.]

For these reasons, the expansive power and throughput of NGS technology notwithstanding, the output of an NGS experiment may be likened to a rough draft of an individual’s genome. Turning this draft into a clinical-grade result that can be reported back to a patient or research participant still requires manual “polishing,” usually by Sanger sequencing. Since Sanger sequencing remains low-throughput, this finishing work is typically only performed in a focused manner on genomic findings felt to be of potential clinical significance.

Sequencing Interpretation

Beyond technological considerations, the clinical application of NGS involves many additional steps. The decisions made in these steps determine both the output of the sequencing (what DNA is actually sequenced) and how that output is analyzed. The end product is a report that summarizes the output of this complex process.

Selecting a field of view

The first decision to be made after a DNA sample is collected is the selection of a technical field of view (FOV), that is, what portion of the sample will be targeted and sequenced. There are typically three options: (1) targeted multigene panels, (2) whole exome sequencing (WES), and (3) whole genome sequencing (WGS).

WGS has theoretical advantages, including near-completeness (subject to the limitations described above), uniformity of coverage, and the ability to detect structural variants. WGS also suffers from several disadvantages that limit its use today. First, WGS is much more expensive than more selective approaches, not only because more DNA is sequenced but also because the resulting data files are extremely large, making them difficult to manipulate computationally and store. Another is that, 98% of the sequence generated by WGS lies in the noncoding portion of the genome, which remains poorly characterized and as yet largely clinically uninterpretable. Finally, given technological and resource limitations, WGS sequencing incurs a trade-off between breadth and depth. WGS sacrifices depth of coverage (average 30 – $50\times$ read depth in

WGS) relative to exomes (which typically achieve $>80\text{--}100\times$ across the target) and multigene panels (which often achieve $>200\times$ or greater coverage), resulting in more dropouts and less accurate variant calls.

With WES, only the 2% of the genome that contains the protein-coding exons is sequenced. WES takes advantage of a variety of techniques that allow one to selectively enrich for DNA corresponding to exons (typically by deploying sticky complementary DNA baits). Since enrichment is not uniform across the genome, this results in more variability in read depth than WGS, but the trade-off is much greater efficiency of capture in the portion of the genome that can be clinically interpreted.

Finally, the most limited, yet most accurate and easiest to interpret, option is targeted multigene panels. In this case, the capture method is restricted to predefined subset of genes preselected to fit the clinical reason for sequencing. For example, a panel for diagnosing hypotonia would have only genes that are known to cause myopathies, muscular dystrophies, and other relevant conditions characterized by hypotonia. The advantage of panels is that given their smaller targets, capture efficiency and sequencing depth can be maximized. The major disadvantage of panels is that one is limited to the genes on the panel. If the individual's condition is caused by a variant in a gene that is not on the panel, the causative variant will not be detected. Targeted panels are very useful for clinical purposes, but are typically not used for research and gene discovery.

Choosing a region of interest

Distinct from selection of a FOV is the definition of the "region of interest" (ROI). This may include the entire FOV, or it may be substantially smaller; depending on the indication, not all of the sequence generated in an NGS experiment necessarily needs to be (or should be) formally analyzed. While the FOV defines the part of the genome for which data are technically *generated*, the ROI limits the DNA sequence data to be *analyzed*; thus if there is a variant in a gene that falls outside either the FOV or ROI, it would not be reported. This has important implications for assessing secondary findings, which must fall within the ROI if they are to be detected.

The selection of an ROI depends on a number of factors ranging from technical to medical to ethical. In the case of a conventional gene panel, the ROI is typically by definition the same as the FOV, since the FOV is limited to a small group of genes known to be related to the phenotype of interest. Alternately, the sequencing laboratory could choose a WES FOV, but limit the ROI to the same genes as those in a targeted panel, ignoring the sequence data generated for other genes in the exome. In the case of a patient with an undiagnosed rare disease, where panel sequencing may not be informative, the ROI may be chosen so as to include genes for all of the 3000+ known Mendelian disorders described, or even the entire exome, especially for an individual enrolled in gene discovery research. Finally, ROIs may be tailored because of ethical considerations, for instance by masking sequence data from the *ApoE* locus in an individual who declines to know her familial Alzheimer risk.

Annotating, filtering, and interpreting variants

After these steps, the final and most challenging step in NGS analysis is sifting through the variants obtained from the

ROI and deciding which may be of clinical relevance to the individual, so that they may be clinically confirmed and reported. Depending upon the size of the ROI, this may require prioritizing dozens (gene panels) to tens of thousands (whole exomes) to millions (whole genomes) of variants according to a number of criteria, including filtering technical artifacts, excluding common variants, considering inheritance patterns, canvassing the literature, and applying computational predictive tools. While the processes for evaluating variants vary slightly between laboratories, they follow guidelines developed by the ACMG (Richards *et al.*, 2015) to classify variants into one of five categories indicating the degree of certainty that the variant is associated with disease: pathogenic, likely pathogenic, uncertain significance, likely benign, and benign. Fundamentally, these guidelines address the following core questions:

First, is this a variant that has been previously reported in association with clinical disease? There is as of yet no single authoritative source with which to answer this, but laboratories consult a number of resources, including ClinVar (www.ncbi.nlm.nih.gov/clinvar/intro), HGMD (Human Gene Mutation Database, www.hgmd.cf.ac.uk/ac/index.php), OMIM (Online Mendelian Inheritance in Man, <http://omim.org>), LOVD (Leiden Open Variation Database, www.lovd.nl/3.0/home), and PharmGKB (www.pharmgkb.org), as well as reviewing the primary literature (e.g., PubMed, <http://ncbi.nlm.nih.gov/pubmed>). These databases house lists of variants, together with their clinical classifications (benign, uncertain, or pathogenic), and, to varying degrees, references to evidence supporting these assertions. It is important to stress that as community-generated resources, standards for curation vary and errors in the database are well-documented. Nonetheless, these are valuable starting points for a clinical genomic evaluation.

Next, how frequently is this variant observed in the human population? As described earlier, over 97% of genetic variation observed in individuals is in fact common in the population and therefore not likely to contribute to rare disease. Filtering out common variants is typically performed by cross-checking against dbSNP (www.ncbi.nlm.nih.gov/projects/SNP), the 1000 Genomes Project (www.1000genomes.org), the NHLBI Exome Sequencing Project (<http://evs.gs.washington.edu/EVS>), and the Exome Aggregation Consortium (<http://exac.broadinstitute.org>), each of which contains research data from thousands to tens of thousands of individuals who have undergone genome sequencing. Variants that are frequently encountered in these databases can generally be excluded from consideration as a cause of severe disease (note that this does not hold for some categories of clinically relevant variants such as pharmacogenomic variants, which can be common).

Last, what is the predicted impact of this variant on protein function? For the vast majority of variants encountered in a clinical genomics evaluation, experimental data do not exist. Laboratories must instead rely on knowledge of gene structure and the genetic code to try to predict computationally whether variants are expected to alter gene function. Most readily interpretable are nonsense or frameshift mutations, which result in gene disruption by protein truncation. Similarly, synonymous mutations, in which a base is altered without a change in the underlying amino acid sequence, are almost always benign. Much more difficult to predict are the effects of missense mutations, resulting in substitution of one

amino acid for another. Assessing the effects of missense mutations requires examining the protein's structure as well as looking for evolutionary patterns indicative of local sequence conservation or constraint. Predicting the impact of these changes is difficult, as evidenced by the number of computational algorithms that have been developed (e.g., SIFT, Polyphen2, LRT, MutationTaster, MutationAssessor, FATHMM, CADD). Murkier still are the category of splice variants, which may lie near exon-intron boundaries (including some nominally "synonymous" variants) or may be deep within introns.

These evaluations, while based on principles of functional and population genetics, are some of the most difficult aspects of clinical genome analysis, and are subject to many potential pitfalls and limitations. For instance, largely unaddressed in prediction of variant impact are potential mutations in or near potential regulatory sites (for instance, 5' or 3' UTRs [untranslated regions], or gene promoters or enhancers). While this category of mutation is expected to account for a minority of genetic disease, they are essentially uninterpretable at this point. Also unaddressed is the gain of function mutations, which have been demonstrated to be of special importance in certain diseases (overgrowth syndromes, cancer, etc.). It is particularly difficult to predict what novel mutations may lead to gain of function, either by removing a regulatory domain or altering a protein in a particular way.

An important concept in assessing the pathogenicity of a variant, especially a novel variant, relates to the prior probability that the individual sequenced has the relevant predisposition or disease. If a person has a phenotype, and a known or novel variant is found in a gene associated with that phenotype, even a novel variant is more likely to be pathogenic in the individual. Assessing family members with and without the phenotype for the variant can be helpful. On the other hand, if a novel variant in a known gene is found in an individual with no phenotypic manifestations, the likelihood that the variant is associated with or predictive of the phenotype in that individual is much lower.

Even if a variant has been previously reported, analysts must consider the possibility that the reported association is a false or overstated one. One of the greatest concerns is "ascertainment bias," which refers to the fact that variants in causative genes (e.g., mutations in *BRCA1* that cause a strong susceptibility to breast cancer) were identified in women with breast cancer. Until large-scale studies of unselected populations are available, the true risk that a woman with a *BRCA1* mutation will develop breast cancer may be overestimated since the true penetrance of the variant (i.e., the proportion of individuals carrying the variant who also develop the disease) is not known. Ascertainment bias presents a serious challenge when interpreting variants found in individuals who do not have an elevated prior probability of genetic risk due to their phenotype or family history.

Interpreting variants in new genes is particularly daunting. Such "gene discovery" requires functional, family, population, or other studies to determine if the finding is truly a new gene responsible for a disease.

Finally, after each variant is assigned a classification, judgment must be applied as to which variants should be reported based on the patient's clinical presentation. Typically, this judgment depends, in part, on the presenting signs and symptoms of the patient, since that establishes the prior

probability that an impactful variant exists. For instance, if a variant is found in a gene that is associated with the clinical symptomatology that prompted sequencing (the primary indication), then the level of certainty needed to report the variant is generally considered low. In this situation, most laboratories report variants that are pathogenic, likely pathogenic, or of uncertain significance. On the other hand, if a variant is found in a gene that is unrelated to the primary reason for sequencing (a secondary finding), then the level of certainty needed to report the variant is typically higher, as the patient is not known to have the condition, reducing the prior probability that the variant is actually impactful. In this situation, laboratories, typically, only report variants that are deemed pathogenic or likely pathogenic.

What Secondary or Incidental Findings Are Possible with Different Levels of Analysis?

Genetic findings unrelated to the initial reason for sequencing are considered "incidental findings" if they are discovered unexpectedly and "secondary findings" if they are deliberately sought (Presidential-Commission-for-the-Study-of-Bioethical-Issues, 2013). The FOV, which determines the level of NGS analysis, will determine what incidental or secondary findings are possible. If the FOV is a targeted gene panel, there will not be either incidental or secondary findings since the genes sequenced are by design relevant to the indication for sequencing. If WES is performed, but the ROI is limited to a certain number of genes (either a panel of genes or the 3000+ known Mendelian disorders described), then there will only be secondary findings in genes that fall within the ROI, such as those on consensus list of genes that should be analyzed even if they are unrelated to the indication for sequencing (Green *et al.*, 2013). The broader the group of genes within the ROI (e.g., the Mendelian disorder genes), the greater the likelihood that there will be an incidental finding. If WES is performed for gene discovery and the ROI is the whole exome, the genes on the secondary finding list will be available for interrogation, and the possibility of an incidental finding will exist.

Ethical Implications in the Research Domain

The foregoing review demonstrates the intricacy of the process involved in getting from DNA to a "result." This complexity, which is much greater compared with most laboratory tests, has at least two broad implications for discussions regarding return of individual results from research sequencing. First, sequencing laboratories must make numerous choices that seem on the surface to be technical in nature, but that on closer inspection have deep ethical dimensions. Second, decisions about return of results, whether at the level of policy or of an individual study, must be based on a deep understanding of the meaning and limitations of NGS, including both its technical and interpretive aspects.

Sequencing laboratories face decisions about depth of coverage, FOV, ROI, and level of certainty about a variant's pathogenicity required to return that variant to the investigator or research participant. Each of these decisions involves trade-offs among potential for benefit, potential for harm, and cost. Laboratories can also choose to allow research participants to influence these trade-offs by setting preferences.

Doing so incorporates respect for autonomy into the ethical milieu, but adds complexity and, in many cases, cost.

Whatever the choices that laboratories make when conducting NGS, producing valid and reliable results requires considerable technical and clinical experience and expertise. There is considerable room for error in both generation and interpretation of sequence data; even in the most capable hands, results differ across laboratories (Amendola *et al.*, 2015). Thus, decisions about return of results in the research context need to take into account whether or not the investigator and the laboratory have the ability to perform high-quality sequencing and interpretation, or the resources to engage an expert clinical laboratory to do so.

Decisions about return of results must also take the resources required into account. Despite the declining costs of generating sequence data, high-quality sequencing and interpretation remain costly. The need, in most cases, to confirm findings has further implications for cost. Currently, most WES and WGS testing require imperfect high-throughput sequencing, followed by low-throughput confirmation by Sanger sequencing. Given this two-step process, careful judgment is needed to decide which variants undergo confirmation. Imposing broad obligations of return on investigators has the potential to substantially burden resources available for research.

Some advocates of offering individual genomic results posit a “duty to hunt,” which implies analyzing and interpreting sequence data from a broad ROI within a given FOV (Gliwa and Berkman, 2013). Interrogating a wider FOV and ROI increases the potential to identify incidental findings, which are associated with the possibility of both benefit and harm. When the laboratory responsible for generating and interpreting the sequence lacks clinical-grade expertise, as is often the case in the research setting, the balance tips toward harm as the risk of identifying and returning false-positive, uncertain, or overstated results increases. Concerns about harm in such studies support a *prima facie* argument for a parsimonious approach (Presidential-Commission-for-the-Study-of-Bioethical-Issues, 2013).

Restriction of returnable findings to those determined to be pathogenic or likely pathogenic reduces, but does not eliminate, harm from false-positive findings. Conversely, returning variants of uncertain significance increases the potential for harm due to false-positives. Concerns about false-positives are greatest when the variant is unrelated to the indication (i.e., family history or phenotype) for which sequencing was performed, and therefore for which the prior probability of identifying a pathogenic variant in the relevant gene was low. In the research setting, this observation creates a strong presumption in favor of limiting return, especially of incidental findings, to those findings judged highly likely to be pathogenic.

Conclusion

The arguments support offering research participants individual incidental findings, which are based in respect for autonomy, beneficence, and reciprocity (Wolf *et al.*, 2008, 2012; Fabsitz *et al.*, 2010; Presidential-Commission-for-the-Study-of-Bioethical-Issues, 2013) and do not fully account for the practicality and the numerous technical and interpretive limitations that characterize the current state of the

sequencing art. These limitations make decisions about what constitutes a result, and which results to return, difficult even for expert laboratories. As a result, investigators and IRBs considering sequencing in the research context must recognize the potential harms from overstated or false-positive findings, as well as the resource implications, when making decisions about what results should be offered. Understanding the decisions made in the process of generating and interpreting sequencing data in the research setting, and their ethical implications, is crucial to maximizing the benefits and minimizing the harms of return of results to participants and populations.

Acknowledgments

This work was supported by NIH grants HG006615, HD077671, and HG006828 (Holm), and HG006492 (Joffe).

Author Disclosure Statement

T.W.Y. is cofounder of Claritas Genomics, a gene diagnostics and precision medicine company. I.A.H. and S.J. have no conflicts of interest to disclose.

References

- 1000 Genomes Project Consortium, Durbin RM, Abecasis GR, *et al.* (2010) A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073.
- Amendola LM, Dorschner MO, Robertson PD, *et al.* (2015) Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Res* 25:305–315.
- Bentley DR, Balasubramanian S, Swerdlow HP, *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.
- Fabsitz RR, McGuire A, Sharp RR, *et al.* (2010) Ethical and practical guidelines for reporting genetic research results to study participants: updated guidelines from a National Heart, Lung, and Blood Institute working group. *Circ Cardiovasc Genet* 3:574–580.
- Gliwa C, Berkman BE (2013) Do researchers have an obligation to actively look for genetic incidental findings? *Am J Bioeth* 13:32–42.
- Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17:333–351.
- Green RC, Berg JS, Grody WW, *et al.* (2013) ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 15:565–574.
- Kim JI, Ju YS, Park H, *et al.* (2009) A highly annotated whole-genome sequence of a Korean individual. *Nature* 460:1011–1015.
- Lander ES, Linton LM, Birren B, *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
- Levy S, Sutton G, Ng PC, *et al.* (2007) The diploid genome sequence of an individual human. *PLoS Biol* 5:e254.
- Li W, Freudenberg J (2014) Mappability and read length. *Front Genet* 5:381.
- Ng SB, Buckingham KJ, Lee C, *et al.* (2010) Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* 42:30–35.
- Presidential-Commission-for-the-Study-of-Bioethical-Issues (2013) ANTICIPATE and COMMUNICATE: Ethical Management of Incidental and Secondary Findings in the Clinical,

- Research, and Direct-to-Consumer Contexts. Services US-DoHH, Washington, D.C.
- Raymond VM, Gray SW, Roychowdhury S, *et al.* (2016) Germline findings in tumor-only sequencing: points to consider for clinicians and laboratories. *J Natl Cancer Inst* 108:djv351.
- Rhoads A, Au KF (2015) PacBio Sequencing and Its Application. *Genomics Proteomics Bioinformatics* 13:278–289.
- Richards S, Aziz N, Bale S, *et al.* (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 17:405–424.
- Rizzo JM, Buck MJ (2012) Key principles and clinical applications of “next-generation” DNA sequencing. *Cancer Prev Res (Phila)* 5:887–900.
- Venter JC, Adams MD, Myers EW, *et al.* (2001) The sequence of the human genome. *Science* 291:1304–1351.
- Wang J, Wang W, Li R, *et al.* (2008) The diploid genome sequence of an Asian individual. *Nature* 456:60–65.
- Wheeler DA, Srinivasan M, Egholm M, *et al.* (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452:872–876.
- Wolf SM, Crock BN, Van Ness B, *et al.* (2012) Managing incidental findings and research results in genomic research involving biobanks and archived data sets. *Genet Med* 14:361–384.
- Wolf SM, Lawrenz FP, Nelson CA, *et al.* (2008) Managing incidental findings in human subjects research: analysis and recommendations. *J Law Med Ethics* 36:219–248, 211.
- Zook JM, Chapman B, Wang J, *et al.* (2014) Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat Biotechnol* 32:246–251.

Address correspondence to:
Ingrid A. Holm, MD, MPH
Division of Genetics and Genomics
Boston Children's Hospital
3 Blackfan Circle
Mailstop BCH3150
Boston, MA 02115

E-mail: ingrid.holm@childrens.harvard.edu