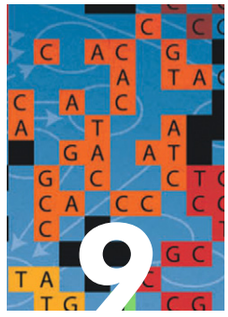
To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

CHAPTER



c0009

Clinical Genome Sequencing

Robert C. Green, Heidi L. Rehm, and Isaac S. Kohane

s0010

INTRODUCTION

p0065 The completion of the first draft of the Human Genome Project in 2001, and subsequent refinements since then, have stimulated an explosion of research about the human genome, the inherent variability of DNA sequences among humans, and the relationship of sequence variation to human health. The scientific impact of this research in understanding the pathophysiology of disease and in spurring new lines of pharmaceutical development has been profound in numerous disease areas, but the impact on the day-to-day practice of medicine has been modest so far, because genomic technologies are still expensive, because the management and interpretation of genome-scale data is challenging, and because the value of genomic data in the practice of medicine has not yet been demonstrated to practicing physicians.

p0070 This situation is rapidly changing. The cost of sequencing individual exomes and genomes continues to drop and will soon be comparable to other common medical tests and procedures. The accuracy of next-generation sequencing (NGS) is improving and innovative methods for efficiently analyzing and interpreting vast amounts of genomic data are being developed. Examples of how whole-exome or whole-genome information may be used in clinical diagnosis and decision-making are no longer rare, and both commercial and academic molecular laboratories all over the world are beginning to offer whole-exome (WES) or whole-genome sequencing (WGS) as a clinical service. Due to lower costs, laboratories have initially focused on WES, but there is general agreement that WGS will eventually predominate since, in addition to knowledge of variation within protein-coding genes, it provides additional information about genome structure and regulation. In anticipation of these developments, this chapter will focus primarily upon WGS. The arrival of genomic medicine, so long anticipated (Feero et al., 2010; Guttmacher and Collins, 2003; Guttmacher et al., 2010), is truly underway, and patients will be looking to the medical establishment for guidance in the use of this technology to improve their health. Yet serious near-term, longer-term and ongoing challenges remain as genome sequencing begins to be integrated into the daily practice of medicine.

Challenges to the Implementation of Genomic Medicine

s0015

p0075 In the near term, sequencing is still expensive, computational costs can be high, and the accuracy of current NGS techniques is not well established, particularly for certain types of variation and for certain regions of the genome. This is particularly true for regions of the genome with repetitive or biased sequence content (rich in GC or in AT base pairs) and for complex sources of variation such as deletions, duplications, and rearrangements. Standards for acceptable accuracy of sequencing in the clinical enterprise are needed, keeping in mind that the applications of sequencing technologies are rapidly changing and that analytical accuracy for one type of variant in one part of

Genomic and Personalized Medicine, 2nd edition
by Ginsburg & Willard. DOI: 10.1016/B978-0-12-382227-7.00009-4

Copyright © 2012, Elsevier Inc.
All rights reserved.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

the genome will not always be generalizable to other types of variants in other parts of the genome. Processing and managing terabytes of raw genomic data is unwieldy, and while the finished textual sequence may only need storage on the order of a gigabyte, the informatics challenge remains large because raw data may be needed for archival or medicolegal purposes or to enable iterative improvements in data analysis methods.

p0080 In the longer term, automation is necessary to prioritize clinically meaningful information by performing the complex and time-consuming task of filtering the large amount of variation from individual genomes. The clinical annotation of the human genome today is an artisanal enterprise, mixing well-established associations with unverified anecdotes, and clinical-grade computational algorithms are required to replace the current elaborate and frequently manual process of evaluating novel variants. Additionally, clinically vetted collections of variants are currently not available in a reliable or wholesale form. Databases such as the Human Gene Mutation Database (HGMD) and Online Mendelian Inheritance in Man (OMIM) are not constructed to facilitate automated searches and are replete with non-standardized and incorrect nomenclature and with errors in clinical interpretation (Tong et al., 2011). Proprietary laboratories maintain their own databases of clinically curated variants for their specific test targets, but do not often share their warehouses of interpreted variants. For systematic and routine analysis of WGS data, access to accurate databases of trustworthy, clinically validated disease-associated variants will be necessary.

p0085 The ongoing challenges to genomic medicine will be to establish and refine the value of using such information in the clinical environment in a cost-effective manner. There will be tension between genomic testing based upon clinical symptoms within the patient or family, and genomic testing that occurs in persons without prior symptoms or family history and, in effect, constitutes population screening. This is because the exact same variants will often have different implications based upon the presence or absence of symptoms or family history. Clinical guidelines will be needed to decide when genomic information can support clinical decision-making, and reimbursement guidelines will be needed to determine what clinical indications can and should be covered. Yet attempting to develop guidelines that can keep pace with the constantly changing science will tax expert organizations and other medical institutions in ways never before experienced. The application of genomic sequencing to large numbers of individuals has the potential to create a torrent of unanticipated findings implicating some degree of risk that will be, for some time, difficult to quantify. The resulting confusion, coupled with the instinct of medical clinicians to order medical tests "just to be safe," has the potential to needlessly inflate costs and increase iatrogenic harm (Kohane et al., 2006; McGuire and Burke, 2008). There is consequently insufficient evidence and enormous uncertainty in how to direct the clinical use of genome sequencing in clinical medicine (Evans et al., 2011; Khoury et al., 2008).

The Case for Genomics in Clinical Medicine s0020

p0090 Still, it seems that the application of genomic information in individual healthcare is inevitable. The technological appeal of sequencing and the potential for automated interpretation, coupled with the paucity of health professionals skilled in genetics, have created an enormous opportunity for new business ventures. If clinical medicine does not adapt to genomics and learn to manage such information, it could be disseminated entirely outside of conventional medical care. This possibility is presaged by direct-to-consumer (DTC) genetic testing companies that currently provide microarray results based upon genome-wide association studies (Frueh et al., 2011), and have begun offering common variants associated with Mendelian disorders as well as tests that reveal carrier status of recessive syndromes. Many of these companies are poised to begin offering sequence information and interpretation as soon as it is economically feasible. While the innovation demonstrated by the best of the personal genomics companies is laudable, many believe that the most appropriate setting for the contextualization of genome sequence information and its integration within the healthcare plan of an individual is within the physician–patient relationship.

p0095 Patients concur that their physicians should be involved, as 78% of those responding to surveys through social networks reported that they would ask their physician for help interpreting genetic test results, and 61% felt that physicians had a professional obligation to help them with the interpretation of genetic findings (McGuire et al., 2009). In addition, a survey of patients enrolled in the Coriell Personalized Medicine Collaborative found that over 90% of respondents were likely to share their results with their physicians (Gollust et al., 2012). To adequately support these patients, physicians must address the dilemma of interpreting and acting upon genomic data before there is sufficient evidence to fully guide its use. Genomic information in medicine has been singled out as especially "difficult to interpret" (Varmus, 2010), and as proving that "we understand … even less well than we currently suppose" (Feero et al., 2010). Some have coined the term "evidence dilemma" (Khoury et al., 2008), while others have urged us to downsize our expectations with more limited testing, or "deflate the genomic bubble" (Evans et al., 2011; Sharp, 2011). Yet managing patients without a complete evidence base is a familiar situation to clinicians, and the slow accretion of reliable data and the absence of sufficient evidence in clinical medicine is the subject of a considerable literature (Downing, 2009; Lenfant, 2003; Petitti et al., 2009). In fact, the traditional practice of medicine often necessitates use of tests, tools, and procedures with insufficient evidence (Brauer and Bozic, 2009; Moussa, 2011; Travis, 2006). As the US Preventative Services Task Force has written on this subject, "even though evidence is insufficient, the clinician must still provide advice, patients must make choices, and policymakers must establish policies" (Petitti et al., 2009).

p0100 As an introduction to the challenges in this rapidly evolving field, we present a summary of current technical and

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

interpretational challenges pertaining to clinical laboratory genomics. This is followed by an overview of the ways in which genome sequencing may be used in the clinical practice of medicine and the challenges associated with these different applications.

CHALLENGES OF GENOME SEQUENCING IN THE CLINICAL LABORATORY

Since a genome sequence may be revisited and reinterpreted throughout an individual's lifetime, the initial generation of the data should be subject to high analytical standards, and subsequent analyses should iteratively reanalyze the data to ensure genetic findings are accurate and comprehensive, especially in the context of the particular indication. Clinicians making use of genome information should be aware of the limitations of genome sequence data, potential sources of error inherent in its generation, and have a working familiarity with important concepts relevant for interpreting the data (see Box 9.1). In this section, we will explore the steps taken to transform a blood or tissue sample into electronic signals representing the individual base pairs, and to use these sequences to generate a report that will allow clinicians to diagnose, manage, or predict disease. Despite their intrinsic complexity, current laboratory protocols for whole-genome sequencing can be summarized in relatively few steps, each of which have distinct challenges that can be addressed through robust laboratory workflows, stringent quality control metrics, and rigorous data handling. These steps are outlined in Figures 9.1 and 9.2 and are described in more detail in the sections below.

Sample Continuity and Tracking

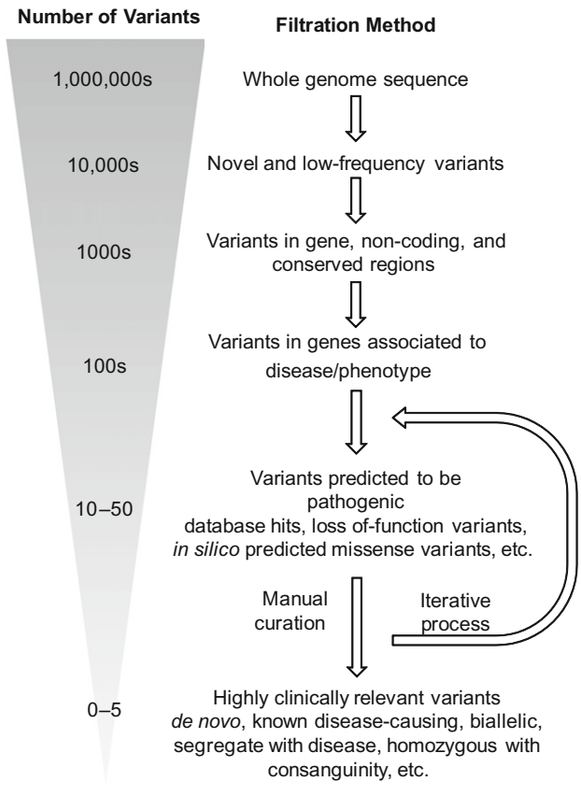
The crucial hand-off between caregiver and laboratory is the first point at which potential errors can occur, and established quality standards for sample provenance and tracking are part of Clinical Laboratory Improvement Amendments (CLIA) certification. All patient samples sent to a laboratory should be accompanied by (1) sufficient clinical information to ensure appropriate and accurate testing and interpretation of results, and (2) at least two unique identifiers that are used to identify a sample during its time in the lab, such as the patient's name, date of birth, or medical record number (Hull et al., 2008). In the context of WGS, providing accurate phenotype (and, if relevant, family history) information to the laboratory is essential as it provides a framework for variant interpretation and identification of disease-causing variation among the sea of data produced. Unique laboratory numbers are also assigned through bar-coded labels and physically affixed to the sample tube and to subsequent reaction tubes or plates. As nucleic acids are often extracted from several patient samples in parallel, robust, computerized sample tracking systems are often

BOX 9.1 Common terms and tools used to describe genome sequencing data

- **Whole-genome sequencing –** analysis in its entirety of genetic material from a sample provided by an individual
- **Whole-exome sequencing –** analysis of the ~1% of the genome thought to be functionally relevant, specifically exons from genes encoding proteins and functional RNAs
- **Targeted sequencing –** analysis of one or a small panel of genes. Depending on the number and size of the genes, this can be achieved using traditional sequencing methods, although currently panels of more than 30 genes are generally more amenable to targeted next-generation sequencing methods for cost and technical reasons
- **Genome position** – a sequential number explicitly referring to a single location or a range of positions in a reference sequence. Often reported as chr1:234567-234568 (UCSC format) or g.234567_234568 (HGVS format)
- **UCSC (University of California, Santa Cruz) Genome Browser** – web-based tool commonly used to examine annotations and data available describing a genome position. Accessible at http://genome.ucsc.edu/
- **Integrative Genomics Viewer** – downloadable tool commonly used to examine alignments of sequencing and other types of genomic data. Accessible at http://www.broadinstitute.org/igv
- **Coverage or depth** – the number of overlapping sequencing reads that map to, and overlap, a genome position
- **Allele balance or allelic fraction** – the ratio of reads supporting a non-reference over a reference base at a genome position; this ranges from 0 to 1. A value of 0.5 would describe a perfect heterozygote although variation around this value is common
- **Strand bias** – a metric that attempts to capture whether a variant is supported by an expected distribution of reads corresponding to positive (forward) and negative (reverse) strand sequences. An unbiased position should have near-equal numbers of forward and reverse reads, although this ratio can be skewed by random sampling of a sequencing library, particularly at low coverage
- **Sequencing library** – specially prepared, often random, fragments of genomic DNA suitable for sequencing

Figure 9.1 Laboratory workflow for whole-genome sequencing.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.



f0015 **Figure 9.2** Variant analysis and filtration in whole-genome sequencing.

used to ensure no samples are confounded or cross-contaminated. While laboratory processes are in place to ensure sample mix-up is rare, potential use of sequencing data over the lifetime of an individual underscores the importance of perfectly pairing sequence data and the individual. Therefore, a number of additional processes should be used to conclusively demonstrate that the data generated from a sequencing laboratory correspond to the original sample received for testing, including:

o0010 1. Genotyping of a small set of informative DNA markers, ideally in an independent sample, and comparing these results with the final sequencing result. This can also be done using an orthogonal technology

o0015 2. Confirmation of clinically relevant variants detected using an independent method, such as Sanger dideoxynucleotide sequencing

o0020 3. Confirming reported gender, ethnicity, family relationship, or other provided demographic features with what is predicted from variants sequenced as a part of the sequencing assay.

p0130 The selection of molecular methods to verify sample identity is laboratory-specific and, unlike the ubiquitous paper barcode label, is not uniformly implemented or standardized. Guided by additional validation data, a combination of these methods will likely be used to ensure accurate tracking of a DNA sample through the genome sequencing process. Challenges to implementing these methods include the need to protect a patient's right to privacy with regard to linking family and sequence information (Cassa et al., 2008), increased burdens to clinicians and hospital groups of collecting extra samples for testing, and extra costs associated with confirmatory testing. Resolution of these issues is largely institutional and requires testing laboratories to work closely with hospitals, clinicians, and patients to ensure that only desired information is returned, that validation data are available and trustworthy, and that additional sample collection and testing are conducted with minimal impact on budgets and current workflows.

Library Construction s0035

p0135 A whole-genome sequencing "library" consists of specially prepared fragments of genomic DNA suitable for sequencing. Ideally, fragmentation of the genome across thousands of cells is performed, so that the resulting data are sampled randomly across the genome with overlapping fragments providing redundant, confirmatory evidence of the genomic sequence. The success of library construction is heavily dependent on the purity, integrity, and quantity of the original DNA extracted from a patient sample. High-quality genomic DNA is readily extracted from most blood and fresh-frozen tissue samples, and protocols continue to improve for preparing libraries from suboptimal specimens such as formalin-fixed paraffin-embedded specimens (Wood et al., 2010). Several methods are used to randomly fragment DNA, including enzymatic digestion (non-random cutting at specific recognition sequences), nebulization (aerosolizing a DNA solution under high pressure), sonication (generation of microcavitation bubbles within solution using sound waves), and hydrodynamic shearing (forcing DNA through a small hole under high pressure) (Joneja and Huang, 2009). With current sequencing technologies, once DNA fragments are generated, common adaptor sequences are added to each end and the products are amplified by polymerase chain reaction (PCR). In addition to facilitating the subsequent sequencing reaction, the adaptor sequences might contain short, synthetic sequences used to uniquely identify a sample's fragments in case of cross-contamination or, more commonly, to enable pooling of samples at subsequent steps to reduce cost and streamline workflows. These identifying sequences are commonly referred to as "molecular bar codes" or "indexes" and are often employed in exome and targeted gene panel sequencing, but are not common in whole-genome sequencing. With a finished sequencing library in hand, the sample can proceed directly to sequencing of the whole genome, or can be taken through an optional target selection step.

p0140 If the sequencing analysis is limited to a portion of the genome, such as the exome or targeted genes, then it is necessary to isolate specific library fragments encoding regions of interest. To capture these fragments, libraries can be hybridized to hundreds of thousands of synthetic probes designed

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

to encode targeted regions, leaving the unbound fragments to be separated away (Albert et al., 2007; Gnirke et al., 2009). Other options exist for targeting subsets of the genome that obviate the need for library construction, including PCR-based techniques that incorporate addition of sequencing adapters (Porreca et al., 2007; Tewhey et al., 2009); however, these methods are limited in the amount of total sequence that can be captured and are not currently used for whole-exome capture. Regardless of the method used, isolated fragments only represent a small fraction of the initial whole-genome sequencing library, ~1% in the case of whole-exome sequencing. This allows for reduced sequencing costs and may enable the division of a sequencing machine's capacity across multiple samples through pooling of samples, either prior to target selection or to sequencing. As sequencing costs drop, target selection is expected to become increasingly unnecessary, as the cost of performing the capture becomes greater than sequencing an entire genome itself.

s0040 ## Sequencing

p0145 Historically, DNA fragments were sequenced individually using Sanger dideoxynucleotide sequencing. In contrast, next-generation sequencing (NGS) technologies produce billions of fragments simultaneously in a single machine run. There are several companies that have commercialized methods to sequence DNA in a massively parallel fashion (e.g., Illumina, Life Technologies, 454 Life Sciences, Complete Genomics), and emerging technologies are under development by new companies (e.g., Pacific Biosciences, Oxford Nanopore, GnuBio). The sequencing reads generated from NGS are substantially shorter than traditional Sanger methods [35–250 base pairs (bp) vs 650–800 bp]. However, the billions of short reads overlap substantially, which effectively blankets a region in multiple redundant reads rather than the two bidirectional reads generated by Sanger sequencing assays. The ability to generate billions of reads in a single assay allows for high coverage across a substantial portion of the genome, making possible the move away from single genes toward sequencing gene panels and entire genomes.

p0150 While a discussion of the strengths and weaknesses of the individual technologies is beyond the scope of this article and has been addressed elsewhere (Mardis, 2008), the end result is the same: a massive text file containing billions of As, Cs, Ts, and Gs (representing the bases adenine, cytosine, thiamine, and guanine) along with associated quality scores representing a statistical prediction of the accuracy of each base that was called. However, the length and quality of these sequences is dependent on the sequencing technology used. The accuracy of most current technologies is reduced in sequence motifs such as homopolymers (long stretches of the same base), simple repeats, or GC- or AT-rich regions. In addition, sequencing artifacts can arise from sample contamination, variable machine performance, and error rates and biases inherent to polymerases used during library construction and sequencing. These technical challenges are being continually addressed by vendors, resulting in greater quality sequences with each improvement of the existing technology. The challenge of subsequent analyses is to ensure that the data generated continue to be of high quality, to detect and address sample or laboratory problems, and to differentiate true sequence variants from inevitable sequencing errors.

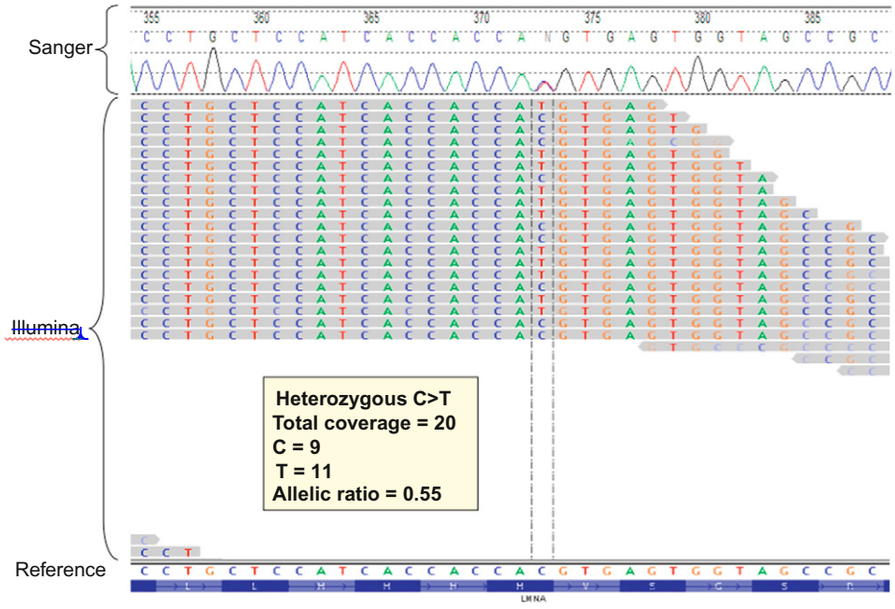
p0155 Due to their continual iterations, the exact error rates of the most widely used sequencing technologies from Illumina (sequencing-by-synthesis), Life Technologies (SOLiD or supported oligo ligation detection), and 454 (pyrosequencing) have been difficult to pinpoint (Mardis, 2008). Sequences are highly accurate (>99.99%) when measured against known bacterial genomes or the few publicly available, well-characterized human genomes (Bentley et al., 2008; Margulies et al., 2005; McKernan et al., 2009). However, a recent article comparing two platforms sequencing the same individual (Lam et al., 2012) revealed a lack of concordance for 19.9% of single nucleotide variations and for 73.5% of indels (small insertion/deletion polymorphisms). While some of this discordance may reflect incomplete optimization of each platform by the investigators, these data suggest that there is still room for improvement until these technologies are truly "clinical grade."

p0160 While there is currently no large set of gold-standard human genome sequences for rigorous assessment of assay sensitivity and specificity, efforts are underway to generate reference materials that have been well characterized by a variety of sequencing approaches (Lam et al., 2012). As sequencing costs decline over the next few years, thousands of genomes sequenced using multiple platforms and with very deep coverage will be available for cross-comparison and validation purposes. In the meantime, validation of clinical tests relies on comparing variant detection to data generated from historical technologies. Given the extraordinarily small portion of the genome that current clinical tests assay, extrapolating general sequencing accuracy across the entire genome likely provides an inaccurate estimation of specificity and sensitivity. This is particularly important as certain classes of variation, such as large repeat expansions, are not currently resolvable by short-read technologies. Therefore, initial genome sequencing test launches will likely focus on the analysis of "well-trodden" portions of the genome to ensure that results are consistent with those from established technologies. As an understanding of error rates and technical limitations accrues, these methods will eventually ramp up to provide fully validated analyses of entire genomes at base-pair resolution, likely guided by increasingly available datasets for validation.

s0045 ## Alignment

p0165 Alignment is the process of assigning or "mapping" each NGS read to a corresponding position in a reference sequence (see Figure 9.3). The most widely used human genome reference sequence is maintained by the Genome Reference Consortium (Church et al., 2011), which provides a critical service continually updating versions of the genomic sequences initially published in 2001 by the Human Genome Project (Lander et al.,

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.



f0020 **Figure 9.3** An alignment of NGS reads supporting a heterozygous single-base substitution confirmed by Sanger sequencing.

2001) and Celera Genomics (Venter et al., 2001). To facilitate accurate mapping and therefore accurate variant detection, aligners must be tolerant of differences between reads and the reference without being too permissive. Requiring perfect matches to the reference will result in false-negative calls, as any variation by its definition is a difference from the reference. However, being too tolerant of mismatches will align reads to incorrect regions of the genome or align poor-quality reads, resulting in false-positive variant calls, i.e., more mismatches that actually exist. Several software packages exist to perform sequence alignment, and like the sequencing hardware, each is under continual development and each has pros and cons beyond the scope of this article (Li and Homer, 2010). The choice of aligner can greatly affect the sensitivity and specificity of subsequent variant detection algorithms, and different aligners may work better for detecting different types of variation. Due to the large size of the genome and the sheer number of reads generated, the alignment process can be computationally intensive, making accurate and efficient aligners particularly valuable. The accuracy of the alignment process will continue to improve as algorithms become more sophisticated and as the "read" length of sequencing technologies increases.

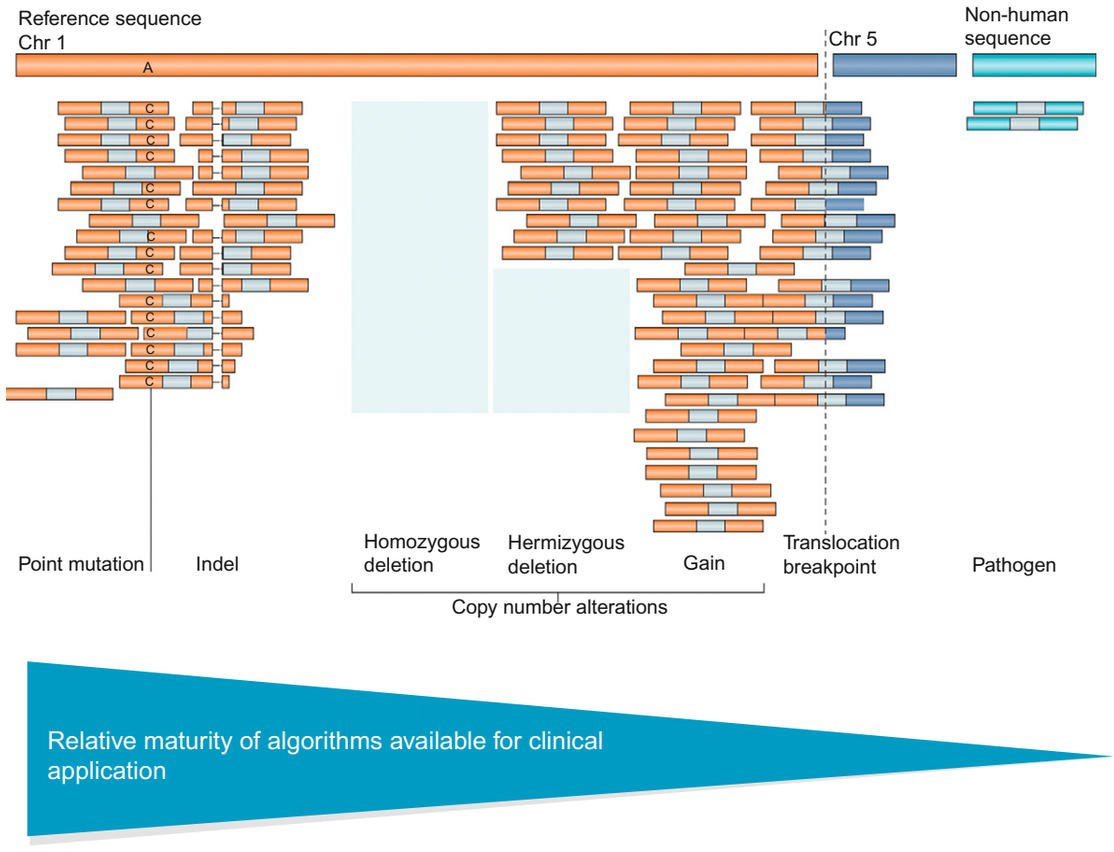
s0050 ## Variant Detection

p0170 The alignment is the first step in detecting the diverse types of genomic variation. The number of reads aligned, or "mapped," to a base within the reference sequence is referred to as "coverage." The greater the coverage available at a given position, the greater the confidence that can be placed in a variant call at that position, as multiple, unique reads derived from different fragments combine to support or refute the presence of a non-reference allele. Currently, detection of single nucleotide substitutions is fairly robust. Other types of variations are more challenging to detect, including small (1–100 bp) insertions and deletions or "indels," as they require the previous alignment step to be tolerant of missing or additional bases. Several tools have been developed to address this problem, but large deviations from the reference still require further algorithm development and/or longer reads for accurate detection. For smaller indels, local realignment within the region is usually employed to refine the mapping.

p0175 In addition to base-pair changes, larger structural changes can be detected from genome sequence data, although the accuracy using current technologies is variable (Meyerson et al., 2010). Gains or losses of large segments of genomic material (amplification or deletions of hundreds to millions of base pairs) in an individual can be inferred from differences in coverage across a region compared to a reference set of sequences. Furthermore, since sequences are often generated from both ends of a single fragment, discrepancies in the expected location of these paired-end reads can be used to infer structural variants including translocations, inversions, and complex rearrangements. Lastly, reads that do not map to the human genome reference may map to other genomes from potentially pathogenic species such as bacteria and viruses. The sensitivity of detection of all of these genomic events is greatly improved by increased depth of coverage across bases within a sequencing library and effective use of available variant detection algorithms (see Figure 9.4).

p0180 Academic and private researchers are largely driving the development of variant detection algorithms, creating a

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.



f0025 **Figure 9.4** Types of genome alteration that can be detected by second-generation sequencing. *Reproduced from Meyerson et al., 2010.*

bottleneck in the transfer of validated, high-quality methods to clinical laboratories. Hence, it is currently the responsibility of the laboratory to iteratively evaluate combinations of aligners and variant callers to ensure accurate detection of the widest range of genetic variation possible. However, as next-generation genome sequencing becomes commonplace, algorithms will undoubtedly mature to form a full collection of tools upon which a lab can draw to drive a comprehensive genome analysis and report. Furthermore, just as the right algorithms have been embedded in magnetic resonance imaging (MRI) scanners' workflow (often without detailed knowledge on the part of the radiologists using these machines), once the standard of care for genome assembly is met, alignment and variant detection will likely disappear into the software of the genomic interpretation pipeline, only to be examined in cases of clinical ambiguity or quality control.

s0055 ## *Variant Interpretation*

p0185 Over three million variants can be found in a single human genome, and simple detection of these is insufficient to generate a clinically useful result. Initial clinical use of whole-genome and -exome sequencing will continue to primarily focus on identifying the cause of a single primary indication and therefore the main strategy for interpreting a genome will be to filter the set of identified variants down to a reasonable number to evaluate manually as candidates for the indication (see Figure 9.2). The first step in, for example, searching for the cause of a rare disease, is to remove all common and known benign variation. This is often performed by excluding all variants found at a high frequency in general population databases such as dbSNP and the more recent dataset from the ESP cohort (National Heart, Lung and Blood Institute, 2012). For example, removal of variants recorded in dbSNP can reduce the length of a variant list by about 90% (Bainbridge et al., 2011; Worthey et al., 2011). Many other filters can then be employed to help narrow down the number of variants to a manageable set, some dependent on the clinical symptoms and suspected mode of inheritance of disease. These include selection of novel variants, *de novo* variants, protein-truncating variants (nonsense, frameshift, splice site), non-synonymous (missense) variants, variants present in existing mutation databases, homozygous variants in consanguineous families, biallelic variants for recessive inheritance, variants in genes with expression patterns matching the organ sites involved, variants in a

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

conserved region of the genome, or variants predicted to be deleterious to protein structure by computational algorithms.

p0190 However, caution must be exercised in using certain approaches. Comparing discovered variants to single nucleotide polymorphism (SNP) databases can be problematic as these sources of variation may contain pathogenic variants, either because the data were generated from sequencing disease cohorts, or because healthy controls have not yet manifested disease, or even because many healthy patients carry recessive disease variants. Additionally, while computational approaches such as PolyPhen (Adzhubei et al., 2010) and SIFT (Kumar et al., 2009; Ng and Henikoff, 2003) are now used in clinical laboratories to augment the semi-manual interpretation of variants, they have not been fully validated and caution should be exercised in trying to use them independently to define the pathogenicity of variants (Jordan et al., 2011).

p0195 Disease- or loci-specific databases containing lists of known pathogenic and benign variants are useful for interpretation of genetic tests (Fokkema et al., 2011; Samuels and Rouleau, 2011; Stenson et al., 2009). Often manually curated from published literature, these databases represent a Herculean effort to capture a relatively small, but important, portion of the overall variation in the human genome. Historically, proprietary knowledge of sequence variation from thousands of individuals has been a competitive advantage among clinical testing laboratories and access to these knowledge bases is often unavailable for outside groups. However, this is likely to change as the cost of sequencing drops, as variation from thousands of genomes becomes readily available, and as laboratories start to report on the thousands of diseases that can now be tested. As laboratories and clinicians will soon be interpreting genome-scale datasets, there is a clear need for a unifying database that comprehensively records frequencies and clinical associations of all variants seen in all genomes with known clinical phenotypes (Ratner, 2012). There have been significant efforts to systematically annotate variants from across the genome (Cotton, 2009; Stenson et al., 2009) and it will be crucial to engage keepers of gene-specific databases to ensure valuable gene-specific knowledge is captured by new, high-content efforts.

p0200 Once a short list of candidate novel variants has been identified, a manual interpretation process is invoked, beginning with careful assessment of relevant variant features, including annotations used for the original filters, and published reports. The available data are synthesized and an interpretation is made based on laboratory- and disease-specific guidelines. Published classification guidelines for molecular genetics laboratories (American College of Medical Genetics, 2008) are available to help guide laboratories in assigning variants to categories like "benign," "unknown significance," and "deleterious" (pathogenic). Some laboratories will also choose to include intermediate categories, such as "likely benign" or "likely pathogenic," in their classification scheme.

p0205 If likely causal variants are not identified during the manual curation of the smaller list, there will likely be a reiterative process that applies separate, often less stringent, sets of criteria to the larger set of variants. There is a danger in expansion of the candidate list, as more variants will be manually reviewed that have a lower probability of being associated to disease. Current anecdotal evidence suggests that whole-genome sequencing may find the causal variant in 10–25% of cases with rare, likely genetic, disorders. While historical genetic testing has a clear definition of a negative result, this is less defined when there is the potential to identify millions of variants, most with no available information. Parameters and guidelines need to be developed for what constitutes a negative case, and if/when such cases should be reanalyzed with regard to evolving knowledge.

Variant Confirmation s0060

p0210 To ensure the accuracy of genetic results, most laboratories confirm clinically relevant variants according to recommended guidelines (American College of Medical Genetics, 2008). Confirmation may use the same technology if the primary rationale is ruling out sample mix-up. In other cases, if the analytical accuracy or resolution of the technique is not sufficient, an orthogonal method may be used. This requirement presents a particular challenge to clinical whole-genome sequencing as it demands a secondary assay capable of assessing virtually any variant that may be detected, or routinely would necessitate complete resequencing of every genome. Current validation methods for substitutions and small indels utilize today's gold standard, PCR followed by Sanger sequencing. Variants that detect larger portions of the genome may require alternative procedures to be in place [multiplex ligation-dependent probe amplification (MLPA), real-time PCR (qPCR), microarray], some of which may also be new to the laboratory. While this approach to confirmation may be feasible in the short term for validating results from targeted sequencing tests, it does not scale well to confirm genome-scale variant sets.

p0215 Aside from confirming true positives, laboratories, clinicians, and patients must realize that whole-genome sequencing in its current form is not 100% complete. Traditionally, sequencing assays on individual genes or panels of genes have interrogated every base pair covered by a particular test. Due to technical and cost limitations, whole-genome sequencing does not have adequate coverage for every base in the genome, and these non- or low-coverage regions can be either recurrent or spurious. There thus exists an inherent risk of false-negatives that will diminish as sequencing and computational methods improve.

Reporting s0065

p0220 The final step in the genome sequencing process is delivery of a clinical report to the ordering physician. This will be a particular challenge in the genomics era as text descriptions traditionally used for reporting the results of single gene tests may not adequately convey the scope and complexity of the interpreted data, and since the sheer number of variants may obscure an established pathogenic variant. Reporting methods

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

for genome sequencing continue to evolve and no standard method has yet emerged. Laboratories will need to decide which of the approximately three million variants that could be found in each patient to include in a clinical report, and how to describe each variant with respect to its relevance to health and disease.

p0225 Open communication between physicians and molecular laboratories is paramount to ensure relevant results are returned for the indication for testing, particularly in these early days of medical sequencing. Since scientific discovery and methodologies in genomics are progressing so rapidly, infrastructure supporting regular updates of analyses and interpretation has tremendous potential to improve patient care, tempered by the need to not overwhelm patients and physicians with extraneous, inconsequential alerts. Innovative methods of communicating and updating variant interpretations and alerting physicians of clinically relevant changes are being implemented and evaluated in electronic medical record systems, a step toward fully revisable laboratory reports (Aronson et al., 2011, 2012). Currently, this model assumes that that raw genomic information will be housed in the laboratory, separate from hospital or physician notes. Alternative routes for variant reporting have been investigated by DTC genetic testing companies, that rely heavily on interactive websites to guide consumers through the interpretation of their genotyped variants, typically one at a time. However, even these sites suggest follow-up with physicians or genetic counselors (of which there is a national shortage) for detailed interpretation.

p0230 In the short term, delivery of genome sequence information will likely follow the current reporting model where reports are crafted around single, typically rare, disease indications and returned to the ordering doctor. However, as genomic data are used more widely as part of the daily practice of medicine, practice parameters may call for repeatedly interrogating the sequence in the context of new information, symptoms, or other indications as they arise. Furthermore, as labs "incidentally" uncover medically important information unrelated to the primary indication for testing, the need to establish a process for returning incidental or secondary findings has emerged. This process has been fueled both by the laboratories as well as patients who are increasingly requesting that all information of personal relevance be returned. However, secondary results may not be relevant at the time of initial sequencing. Thus, it may become necessary to integrate sequence information into a patient's medical record and build infrastructure for physicians and clinical decision support systems to interact with the data over the lifetime of a patient. Much as electronic health record systems now provide order entry decision support to avoid adverse drug interactions or dangerous chemotherapeutic doses, so will electronic health records provide clinical interpretations of each variant singly and in combination (where the epistatic effects are known) and use these to guide clinical decision-making.

CHALLENGES OF USING GENOME SEQUENCING IN MEDICAL PRACTICE

s0070

p0235 How will clinicians utilize genome sequencing in the practice of medicine and where will this eventually lead? As the cost and labor associated with issuing an accurate and interpreted whole-exome or whole-genome sequence continue to fall, one can imagine using this test for almost any situation in which sequencing of individual genes or gene panels is used today, as well as for numerous population screening purposes. In this section, we describe a number of categories of clinical interaction, and briefly discuss the current and future clinical scenarios related to each as sequencing becomes more commonly used in medicine. Table 9.1 lists some of the potential benefits and limitations of whole-genome sequencing applications.

t0010 **TABLE 9.1** Potential benefits and possible limitations associated with clinical applications of whole-genome sequencing

WGS application	Potential benefits	Limitations
Molecular characterization of rare diseases	Definitive diagnosis, family planning, and familial risk assessment	Many variants of unknown significance will be identified
Individualized cancer treatment	Targeted therapeutics specific for a cancer genotype with improved outcomes	There is currently limited prognostic information available
Pharmacogenomics	Inform choice or dosage of medications	Clinical utility has not been demonstrated in most cases
Preconception and prenatal screening	Inform parents about the risk of disease in offspring	Ethical problems arising from acting on such information
Population screening for highly penetrant Mendelian disorders	Identification of diseases where surveillance or intervention could alter outcome	Limited knowledge of penetrance of disease alleles within unaffected families makes interpretation difficult
Population screening for common disease susceptibility	Increased health knowledge and potentially positive health changes	Clinical utility has not been demonstrated

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

s0075

Molecular Characterization of Rare Diseases

p0240 Genome sequencing is becoming increasingly used to understand the molecular pathology of unusual disease presentations and in some cases to lead to new treatment options. The first study to demonstrate that exome sequencing could identify a putative disease gene compared exome data from four unrelated individuals affected with Freeman–Sheldon syndrome (FSS) with similar data from eight healthy HapMap individuals (Ng et al., 2009). This report demonstrated that through identifying genes with one or more non-synonymous coding SNPs, splice site disruptions, or coding indels in one or several affected exomes, and filtering out common variants with dbSNP or HapMap data, the candidate list could be narrowed to a single gene, *MYH3*. This gene had previously been identified by a candidate gene approach as associated with FSS, consistent with the clinical presentation. This proof-of-concept study was followed by a number of reports in which exome or genome sequencing discovered a novel gene or influenced treatment choices in patients with rare conditions. For example, an accurate diagnosis of congenital chloride-losing diarrhea was made based on the detection of a novel homozygous variant in *SLC26A3* from a patient with a preliminary diagnosis of Bartter syndrome born to healthy but consanguineous parents (Choi et al., 2009). Similar bioinformatics filtration techniques identified genes underlying recessively inherited Miller syndrome (Ng et al., 2010a), *de novo* Schinzel–Giedion syndrome (Hoischen et al., 2010) and some cases of *de novo* Kabuki syndrome (Ng et al., 2010b). The search for a common candidate gene in the individuals with Kabuki syndrome was successful only after ranking the affected individuals by canonical phenotypes, reinforcing the limitations imposed by genetic heterogeneity and the importance of phenotype characterization. Also in that study, exome sequencing was supplemented in some cases with Sanger sequencing to detect small indels, reinforcing the current limitations of next-generation detection of some types of variation.

p0245 Whole-exome sequencing of a child with inflammatory bowel disease garnered national attention when discovery of the causative gene supported a therapeutic course that was probably life-saving (Worthey et al., 2011). After assuming a recessive mode and searching for homozygous, hemizygous, or compound heterozygous variants, the investigators found a hemizygous change in a highly conserved residue of the X-linked inhibitor of apoptosis gene (*XIAP*). This mutation was present in the patient's asymptomatic mother, who had skewed X-linked inactivation of the pathogenic allele. Based on this result, previously discarded theories for immunological pathophysiology in this child were reconsidered, and the child underwent a hematopoietic stem cell progenitor transplantation, resulting in apparently sustained resolution of symptoms.

p0250 As of this writing, disease genes have been identified in several dozen additional rare conditions by sequencing a single affected individual, a number of affected individuals, or family groupings such as a trio of affected child and two unaffected parents (Gilissen et al., 2011; Gonzaga-Jauregui et al., 2012). In some cases, the recognition of a causative mutation offers opportunities for reproductive planning. In other cases, identification of the responsible gene has helped with the choice of treatment. For example, whole-genome sequencing in a pair of fraternal twins with childhood-onset dystonia implicated three genes with two or more variants, consistent with an expected recessive inheritance pattern. One of these genes, *SPR* (sepiapterin reductase), was previously associated with dopa-reponsive dystonia. These children had already been treated with L-dopa, and reportedly improved further with 5HTP therapy after recognition that *SPR* mutations can also dysregulate the serotonin pathway (Bainbridge et al., 2011). In still other cases, discovery of a genetic variant has extended scientific knowledge about related conditions. For example, uncovering the molecular etiology in a case of Charcot–Marie–Tooth neuropathy using whole-genome sequencing led to insights into a genetically more complex disease, carpal tunnel syndrome (Lupski et al., 2010).

p0255 There are several thousand syndromes that are thought to have Mendelian inheritance but in which the responsible gene or genes are not yet known, and many of these will gradually be elucidated in coming years. However, even among conditions that appear to have clear Mendelian inheritance, molecular characterization will be more difficult for syndromes that have variable penetrance and substantial locus heterogeneity. Certain forms of apparently Mendelian autism are an example of this, since sequencing family trios of patients with autism spectrum disorder has not yet pinpointed specific genes. However, higher frequencies of protein-altering mutations have been observed in highly conserved residues, identifying a host of potential candidates for responsible genes (O'Roak et al., 2011). Moving beyond analyses of the sequence alone, integration of high-throughput functional analyses of genes that are suspected to be responsible for disease should accelerate molecular characterization of rare diseases (Patwardhan et al., 2009).

Individualized Cancer Treatment

s0080

p0260 Genome sequencing has great potential for disease monitoring and therapeutic decisions. While this will be true of many conditions, in the near term, sequencing will likely be utilized most quickly and most extensively for this purpose in cancer treatment. A full review of the ways in which genome sequencing may be used in the prognosis and management of cancer is beyond the scope of this chapter and may be found elsewhere (see Chapters 57–67) (MacConaill and Garraway, 2010; McDermott et al., 2011). However a brief summary of evolving themes in prognosis and treatment monitoring in cancer is presented.

p0265 Cancer is fundamentally a genomic disease, in that most tumors arise and persist due to genomic changes that often contribute to dysregulated cell growth and survival. Germline variants can also confer increased disease risk or be associated with cancer treatment options by altering drug metabolism. Genome sequencing of both germline and somatic

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

tissue has prompted extensive analyses of changes in cancer genomes, including copy-number changes, rearrangements, small insertions and deletions, and point mutations, culminating in comprehensive catalogues of somatic mutations and insights into the genes that contribute to cellular transformation (McDermott et al., 2011; Pleasance et al., 2010; Stratton et al., 2009). The classification of cancers is still predominantly done through histological analysis of tissue sections or cells, but in some tumor types, such as breast cancers and leukemias, molecular markers have also been utilized for years. Microarray-based expression profiling that measures the expression level of mRNA transcripts is rapidly improving the ability to classify cancers, particularly in early-stage breast cancers, colon cancers, and hematologic cancers (Dave et al., 2006; Lo et al., 2010a; O'Connel et al., 2010; Paik et al., 2004; Rosenwald et al., 2002; Van't Veer et al., 2002; Wang et al., 2005).

p0270 Genes involved in specific cellular pathways are frequently mutated as a consequence of somatic alterations that directly contribute to the abnormal growth of the cancer cell. Sequencing to determine the presence or absence of mutations within these genes can help predict a patient's response to a specific targeted therapy. For example, small-molecule inhibitors of epidermal growth factor receptor (EGFR) kinase activity were originally developed for treatment of cancer because of the role of EGFR in regulating cellular proliferation and because the gene is overexpressed in many cancers. The discovery of activating somatic mutations of EGFR in non-small-cell lung cancer highlighted a subgroup of patients who showed a good response to EGFR inhibitors (Lynch et al., 2004; Paez et al., 2004). A large prospective study has now shown that the response rate to targeted EGFR inhibitors in patients with non-small-cell lung cancer whose tumors harbor an activating *EGFR* mutation is 71%, as compared with 1% for those without a mutation (Mok et al., 2009). In addition, evidence suggests that individuals with a mutation in *KRAS*, a gene downstream of EGFR in the signaling pathway, will not respond to EGFR inhibitors, as could be expected (Pao et al., 2005). In light of trials like this, the analysis of tumor biopsy samples for a subgroup of key mutations in cancer genes that confer sensitivity to targeted agents has been introduced as a routine diagnostic test in some centers. In addition to gefitinib or erlotinib in tumors harboring *EGFR* mutations, a small but growing number of targeted therapeutics have been deployed successfully based on key tumor genetic events, including all-*trans* retinoic acid against acute promyelocytic leukemias with t(15;17) (*PML-RARα*) translocations, traztuzumab against ERBB2-amplified breast cancers, and imatinib in tumors with *BCR-ABL* gene fusions or *KIT* amplification (Buchdunger et al., 2000; Castaigne et al., 1990; Demetri et al., 2002; Huang et al., 1988; Prenen et al., 2006; Vogel et al., 2002). Newer kinase inhibitors targeting mutant BRAF in melanoma and *EML4-ALK* fusions in lung cancer have shown similarly promising results in clinical trials (Flaherty et al., 2010; Kwak et al., 2010). Rapidly expanding knowledge of such alterations in the clinical and translational arenas, including mutations, chromosomal copy-number alterations, and polymorphisms affecting drug metabolism, can be expected to facilitate individualized approaches to cancer treatment.

Pharmacogenomics

s0085

p0275 The underlying premise of pharmacogenomics is that an individual's genetic makeup will significantly determine whether they will respond to, or have adverse reaction to, a given medication (Pirmohamed, 2011; Weiss et al., 2008). An important potential application of genome sequence information is the refinement of drug selection and dosage for a variety of diseases. It is widely expected that knowledge of genetic variation will provide more accurate starting points for drug dosing and for predicting adverse events, as well as dosing management once therapies are underway. And it is frequently mentioned that avoidance of ineffective or potentially harmful drugs for a particular genetic background has potential for savings in healthcare costs and more timely, effective treatment for patients. However, to date there have been a tremendous number of review articles and perspectives pieces about these topics and many fewer examples of pharmacogenomics testing that have been adopted clinically (Holmes et al., 2009). Part of the reason that pharmacogenomics has not been more quickly adopted in clinical practice is that many of the studies demonstrating associations with efficacy or adverse events have suffered from small sample sizes, poor phenotyping, poor study design, or a lack of control for clinical and environmental covariates. However, the proliferation of next-generation sequencing devices and decreased sequencing costs will help overcome many of these limitations, and, combined with systematic patient phenotyping, better-quality studies in pharmacogenomics should soon be emerging.

p0280 There are notable examples linking genetic variants to drug outcomes. Currently, pharmacogenetic tests are used for assigning the dose of 6-mercaptopurine based on *TPMT* (Relling et al., 1999; Schaeffeler et al., 2004; Yates et al., 1997); to decide whether or not to use codeine or tamoxifen based on *CYP2D6* (Caraco et al., 1996); and to design optimal treatment regimens for colon cancer using irinotecan based upon the *UGT1A1* genotype. In addition, pharmacogenetic tests have been developed to define warfarin dosage more accurately on the basis of *VKORC1* and *CYP2C9* genotypes (Jonas and McLeod, 2009) and *HLA-B* variants have been associated with hypersensitivity and liver damage from carbamazepine (McCormack et al., 2011), flucloxacillin (Daly et al., 2009), and abacavir (Pirmohamed, 2010). Yet, the actual utilization of pharmacogenomic variants in clinical practice remains controversial in all but a few situations. For example, with clopidogrel and *CYP2C19* polymorphisms, although there is consistent evidence to implicate the *CYP2C19**2 allele in predisposition to stent thrombosis, the evidence for adverse cardiovascular outcomes following stenting, or in those patients with acute coronary syndrome who have not been stented, is less clear cut (Mega et al., 2010b). For this example, there are several issues at play, which represent the common types of challenge in this field: (1) there is a lack

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

of agreement on whether pharmacogenomics or platelet function tests should be used by the clinician (Azam and Jozic, 2009; Bonello et al., 2010; Fitzgerald and Pirmohamed, 2011); (2) there is insufficient evidence at present as to whether polymorphisms in other genes besides *CYP2C19* (e.g., *ABCB1* and paraoxonase) are also important in defining therapy and/or dose (Mega et al., 2010a); (3) it is unclear what dosing strategy should be used in those patients with either one or two variants in the *CYP2C19* gene for both loading and maintenance to further improve the efficacy of clopidogrel (Bonello et al., 2010; Gladding et al., 2009; Price et al., 2011); (4) the role of genotype-based drug choice and/or drug dose with respect to clopidogrel, and its use, in comparison with the newer anti-platelet agents, such as prasugrel and ticagrelor, is unclear (Yin and Miyata, 2011).

p0285 In an attempt to define which pharmacogenomics associations have evidence of sufficient quality to be used clinically, the Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network has created a "PharmGKB" knowledge base, listing associations that meet particular evidence thresholds (Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network, 2011; Relling et al., 2010; Relling and Klein, 2011). Despite the relative paucity of evidence supporting clinical efficacy, a number of third-party vendors have emerged to provide decision support systems for physicians prescribing specific medications where genotyping could potentially assist at the point of care. As whole-exome and whole-genome sequencing emerge in the clinic, the potential exists to define a large number of pharmacogenomics variants at a single point in time and either offer this compendium to a patient, or have it stored as a personal "library" of variants to be queried whenever a new medication is to be started. In this sense, pharmacogenomics variants constitute the least controversial of the incidental or secondary findings (see below) that could be revealed in any given patient through sequencing.

s0090 ## Preconception and Prenatal Screening

p0290 There are over 1100 known recessive Mendelian disorders that have an established molecar basis. While the diseases themselves are individually uncommon, these syndromes account for considerable pediatric morbidity and mortality (Costa et al., 1985; Kumar et al., 2001). Historically, carrier testing has been offered to populations known to be at risk for specific recessive diseases, or in families where an affected first child has already been born. Preconception screening, along with genetic counseling of carriers, has resulted in steep declines in the incidence of some severe recessive diseases such as cystic fibrosis and Tay–Sachs disease. In some ethnic groups, such as Ashkenazi Jews, panels of carrier tests have been recommended by the American Congress of Obstetricians and Gynecologists (ACOG) (Committee on Genetics, 2009) and commercialized. However, preconception screening has not yet been applied broadly to the general population, in part because the cost of screening for numerous rare carrier states has been prohibitive. The advent of clinical WES and WGS with appropriate informatics analysis will permit simultaneous screening for large numbers of recessive carriers, both known and novel, potentially ushering in new paradigms for reproductive planning.

p0295 In a recent proof-of-concept study (Bell et al., 2011), carrier status for 448 severe recessive pediatric conditions was determined through targeted sequencing of 26 controls and 104 cases, including 76 people known to be carriers or affected by 1 of 37 diseases. Mutation detection had 95% sensitivity and 100% specificity for substitution, insertion/deletion, splicing, gross deletions and single nucleotide polymorphisms (Bell et al., 2011). The average individual carrier burden for severe pediatric recessive disease mutations was estimated at 2.8 mutations. Of note, among the 104 subjects, 27% of the disease-annotated mutations were omitted because of misclassification or lack of evidence for pathogenicity and 26 new nonsense mutations were identified. This report also highlighted the inaccuracy of current mutation databases, as 12% of the literature-annotated mutations previously cited in the 74 affected individuals or carriers were incorrect.

p0300 This study demonstrated the impending feasibility of using targeted sequencing, and by extension genome sequencing, for carrier screening. As noted by the authors and others (Kobelka, 2011), there are a number of technical hurdles and ethical considerations to overcome before such testing could be implemented on a larger scale, many of which have been mentioned above. Testing strategies need to account for large copy-number variants, and informatics analyses require automation and less manual review to streamline the process. Current databases of pathogenic variants are too inaccurate to be reliably used without further curation. The clinical use of preconception screening to avoid affected children presupposes the utilization of preimplantation genetic diagnosis (PGD), which is expensive, or prenatal testing and termination of pregnancy, which is controversial. Current preconception testing has largely been limited to detecting carrier states for pediatric conditions that are fatal or severely debilitating. However, inexpensive carrier testing for hundreds of recessive conditions could well include less severe childhood diseases such as deafness or hemoglobinopathies – or adult-onset diseases, such as neuropathies – and either PGD or pregnancy termination for these conditions would be considered by many to be ethically questionable.

p0305 Advancements in detection and sequencing of circulating cell-free fetal DNA have opened the door for non-invasive *in utero* genetic testing and genome sequencing of fetuses (National Institute of Child Health and National Human Genome Research Institute, 2010). In an early report (Lo et al., 2010b), full maternal and fetal genomes were detected and sequenced using only a maternal blood sample. Next-generation sequencing has been used in validation studies to identify cases of trisomy 13, 18, and 21 with sensitivities of over 98% and false-positive rates less than 0.5% (Chiu et al., 2011; Palomaki et al., 2011, 2012). In fact, such sequencing has the potential to supplant some aspects of ultrasound or newborn biochemical screening and diagnose disorders *in utero* (National Institute of

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

Child Health and National Human Genome Research Institute, 2010). This technology will transform prenatal diagnostics and will present new clinical and ethical dilemmas for clinicians regarding which test results should be delivered to expectant parents. While screening for fetal gender has been technically feasible for some time (Devaney et al., 2011), fetal sex selection has been opposed as unethical by the American Congress of Obstetricians and Gynecologists (ACOG Committee on Ethics, 2007). However, application of similar technologies to uncover lethal or debilitating diseases prenatally might be warranted, particularly as the timeliness, specificity, and sensitivity are likely superior to current methods. However, once all fetal variation is knowable within weeks of conception, there is the potential for selection against virtually any genetic feature, including physical traits. The challenge to caregivers and regulatory agencies will be defining where the line is drawn between deleterious disease and undesirable traits.

s0095

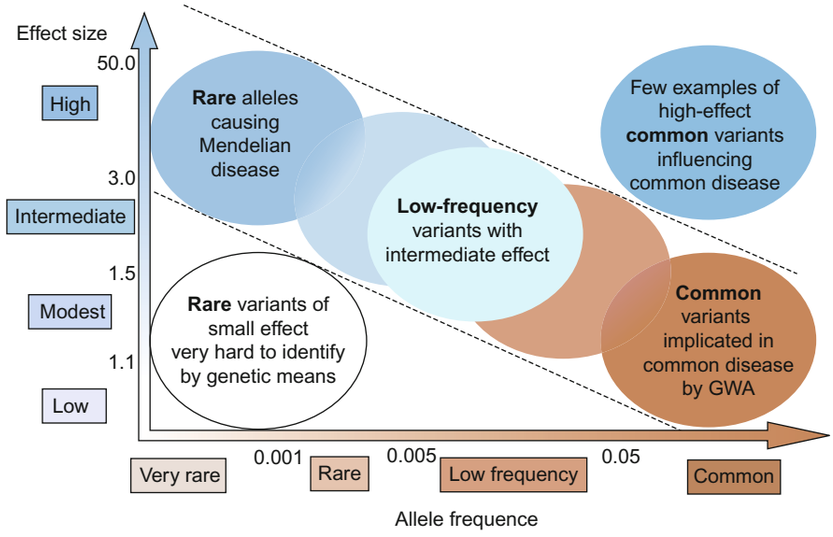
Population Screening for Disease Risk

p0310 The use of genetic tests within clinical medicine has typically been instigated by the recognition of symptoms in a particular patient, or been directed toward an asymptomatic patient with a specific family history of disease. The only exception to this has been state-mandated newborn screening. But the rapid expansion of genetic knowledge and the shrinking cost of genomic technologies, combined with social trends toward personal empowerment and prediction and prevention of disease, are making various forms of population screening for genetic disease risk inevitable. It is helpful to divide the discussion into screening for rare and highly penetrant Mendelian disorders and screening for common disease susceptibility, even though not all conditions are easily divisible into these categories (see Figure 9.5) and the genomic analysis of the future may report on both. It is also useful to consider screening of newborn infants with sequencing, and to consider the overall problems associated with incidental or secondary findings that result from sequencing. These two areas are also described separately below.

s0100

Population Screening for Highly Penetrant Mendelian Disorders

p0315 In the current practice of medical genetics, genetic testing is typically ordered from a menu of thousands of existing tests where there is judged to be a prior probability of that disease after evaluation of patient symptoms or family history. Many of these disorders cannot currently be treated or prevented, and the testing is done to make a definitive molecular diagnosis, ending the "diagnostic odyssey," to permit the family to consider reproductive options, or to inform potential risk in family members. Population screening for these conditions would be difficult to justify at our current state of knowledge. However, population screening might be more clearly warranted for diseases where early intervention could alter outcomes, such as highly penetrant hereditary cancer syndromes. For example, approximately half of the women who are *BRCA1/2*-positive and develop breast cancer have no family history of breast cancer (Rubenstein et al., 2009), raising the question of whether universal screening for these mutations and the resulting surveillance and/or surgery triggered by their discovery could save lives. Likewise, pathogenic mutations for cardiac syndromes associated with sudden cardiac death are relatively common: estimated to be on the order of 1 in 500 for hypertrophic cardiomyopathy and 1 in 3000 for longQT syndrome (GeneTests, 2011). Given the increased risk of persons with these



f0030 Figure 9.5 The spectrum of risk associated with rare and common genetic variants. *Figure reproduced from Manolio et al., 2009, cited in that paper as modified from a similar figure in McCarthy et al., 2008.*

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

conditions to suffer cardiac arrest, a case could be made for population screening, followed by surveillance and even some forms of intervention for those who were mutation-positive.

p0320 Complicating this seemingly straightforward population-based approach is the fact that even Mendelian disorders with high penetrance in affected families may have variable penetrance in the general population, and thus a significant proportion of people carrying a known pathogenic variant will not develop symptoms at any point in their lifetimes. In other words, the association between certain pathogenic variants and expression of disease may be relatively well understood when that variant occurs in the context of symptoms or family history, but the association is far less well understood when found in an apparently healthy individual without a family history. A striking example of this can be seen in cases where a homozygous 845G→A (Cys282Tyr) mutation in *HFE* is highly predictive of hemochromatosis in the context of family history, but poorly predictive in the context of population screening (Beutler et al., 2002). Given this uncertainty, widespread population screening may find pathogenic mutations that label an individual as "at risk" without any clear guidance as to the degree of risk and, thus, the need for intervention. In the example of *BRCA1/2* mutations, clinicians confronted with such mutations might feel obligated to increase breast imaging surveillance and even prophylactic oophorectomies and mastectomies. In the example of cardiomyopathy or longQT syndrome variants, clinicians could feel obligated to follow serial echocardiograms or electrocardiograms, or to consider placement of implantable cardioverter-defibrillatory devices. Given the tendency for clinicians to order tests and procedures "to be safe" when faced with unknown risks, it is easy to imagine that surveillance and proactive interventions could be administered in a highly variable and perhaps even overly aggressive manner. And without empirical trials or even broad clinical experience, it would be exceedingly difficult to determine the benefits, risks, and liabilities associated with such interventions. Medical costs would almost certainly increase as a result of such screening.

p0325 Population screening for highly penetrant Mendelian variants using exome or genome sequencing may be further complicated by whether the individual being sequenced wishes to know all, or only a portion, of the information available from their sequenced genome. Managing this issue will be challenging because patient autonomy and the "right not to know" diagnostic or risk assessment information is firmly established in medicine, medical ethics, and especially in genetics (Erez et al., 2010; Takala, 1999). However, in recent years it has become increasingly controversial as to whether the right not to know should be respected when it puts the patient, or others in the patient's family, at risk of harm (Erez et al., 2010; Malpas, 2005; Wilson, 2005). There has been a countervailing dialogue about "duty to warn," particularly in cases where the relatives of cancer patients with specific mutations may have a 50% risk of a highly penetrant cancer variant (Offit et al., 2004). Analyzing the ethics of the "right not to know" in a single non-actionable variant such as Huntington disease is exceedingly complicated (Erez et al., 2010), and the complexity of allowing patients to decide what they wish to learn, or not learn, from among over a hundred risk variants in each individual genome would seem untenable. One solution is to issue a "general genome report" for each genome using a prespecified list of agreed-upon genes or variants that are screened in every genome (Ashley et al., 2010). However, reaching consensus on the contents of such a list will be difficult (Green et al., 2012), because of the enormous number of "incidental" or secondary findings of unclear pathogenicity as described further below.

Population Screening for Common Disease Susceptibility s0105

p0330 Genome-wide association studies (GWAS) investigating genetic contributions to common complex diseases have revealed thousands of common variants that provide modest risk information, typically with relative risks of less than 2.0. Many of these variants are not in protein-coding regions of the genome, and the data from GWAS have been scientifically valuable in identifying previously unsuspected genetic associations with common complex diseases that may lead to new treatments (Khor et al., 2011). But the use of these variants to offer individual predictions for diseases has been controversial, particularly in light of the highly publicized launch of genetic testing companies that have provided such information directly to consumers (Evans and Green, 2009; Janssens and Van Duijn, 2009; Kraft and Hunter, 2009). It is possible that genetic risk information could promote increased health knowledge and empowerment and motivate positive health changes, and, regardless of its accuracy, could provide "teachable moments" for healthy interventions. Nonetheless, there is concern that the presentation of genetic risk information without the context of environmental or family history may provide an inaccurate measure of overall risk and may lead to unnecessary anxiety or false reassurance (Frueh et al., 2011). This is true in part because for most complex diseases for which risk alleles have been examined, currently identified genetic factors only account for a small proportion of the overall phenotypic variation (Visscher et al., 2012).

p0335 The impact of learning genetic susceptibility information has been studied from a number of angles. Many have been concerned that learning the risk of diseases for which there is no proven prevention could be psychologically damaging. Indeed, when James Watson, codiscoverer of the double-helix structure of DNA, received a report of his sequenced genome, the only information he requested to remain unaware of was the Apolipoprotein E (*APOE*) variant that confers robust information about the risk of developing Alzheimer's disease (AD) (Farrer et al., 1997). Using *APOE* genotyping and risk of AD as a paradigm, the REVEAL Study has conducted a series of randomized, controlled trials that have illuminated many of the issues associated with genetic risk disclosure. At least among persons who volunteered for such studies and were screened for pre-existing psychological instability, it appears that probabilistic risk information can be communicated without excessive anxiety or distress (Green et al., 2009). In fact, those who

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

receive this information value it (Kopits et al., 2011), find personal utility in terms of insurance purchasing (Taylor et al., 2010; Zick et al., 2005), and are more likely to take actions, including unproven therapies, that they hope will reduce their risk of AD (Chao et al., 2005; Vernarelli et al., 2010). The REVEAL Study has also described many of the reasons people seek genetic risk information (Roberts et al., 2003, 2004), how such testing influences self-perception of risk (Christensen et al., 2011; Hiraki et al., 2009; LaRusse et al., 2005; Linnenbringer et al., 2010), and the degree to which participants recall their test results or discuss them with others (Ashida et al., 2009, 2010; Eckert et al., 2006). Other randomized controlled trials of receiving susceptibility information following the REVEAL Study model are underway for other common complex diseases such as diabetes (Cho et al., 2012; Grant et al., 2011, 2012) and obesity (Catherine Wang, personal communication).

p0340 Customers of DTC genetic testing companies have also been queried (Bloss et al., 2011; Kaufman et al., 2012), including one study in which customers were randomized to receive DTC testing or not (James et al., 2011). Each of these studies has documented that customers do not seem to be distressed by the information they receive, but that a small proportion of customers do share this information with their physicians and intend to have follow-up tests as a result of these conversations. The full impact of this information upon medical economics remains unclear (Goldsmith et al., 2012), but there is concern that clinician time and diagnostic studies performed in response to misunderstood, possibly overestimated, risk information could increase medical costs and divert medical resources from areas of greater need (McGuire and Burke, 2008).

s0110 *Population Screening of Newborns or Children*

p0345 In discussions of genomic sequencing and newborns, it is helpful to distinguish between newborn screening and screening of newborns (National Institute of Child Health and National Human Genome Research Institute, 2010). Newborn screening is the established US public health system of blood spot collection and analysis, mandated through state laws, that is used to identify a range of actionable childhood diseases. Screening of newborns is a broader and more controversial concept that involves the use of sequencing or other technologies in newborns or very young children to identify risks for diseases throughout the lifespan.

p0350 The addition of sequencing to state-mandated newborn screening is likely to move slowly for several reasons. Current newborn screening through state laboratories is cost sensitive, currently taking place for less than $20 per child, and while some molecular testing is performed, particularly for secondary characterization of abnormal biochemical results, the cost necessary to convert laboratories and implement universal sequencing would be enormous. Newborn screening is also firmly anchored in a culture that prefers to focus upon the detection of treatable conditions (Wilson and Jungner, 1968). While that culture has faced recent challenges, and there is much controversy around efforts to expand the mandated battery to include untreatable disorders such as Duchenne's muscular dystrophy, fragile X syndrome or lysosomal storage diseases, the sequencing of hundreds or thousands of genes as part of this service, with all of the attendant uncertainties that would result, is not likely to be easily adopted in state-mandated screening, even if cost were not a factor.

p0355 The notion of voluntarily providing genomic screening of newborns or young adults with genome sequencing in order to better understand future health risks has attracted considerable speculation, but as yet no widespread implementation. Population screening of newborns or young children with genomic technologies presents the same opportunities and challenges for detection of highly penetrant Mendelian and susceptibility testing as described above, but with an added layer of complexity because the information would be delivered to parents. The idea is attractive to some because highly penetrant Mendelian variants currently linked to modifiable conditions like cancer or cardiac disease could be identified early and appropriate surveillance initiated. In addition, variants providing susceptibility information for common complex diseases could be identified at an early age and lifelong diet and exercise habits could theoretically be influenced for public health benefit. However, concerns over the difficulty of interpreting risk for even a limited set of variants identified by population screening, coupled with concerns over misunderstanding of risk information and potential damage to the parent–child bond (Waisbren et al., 2003) or to a child's self-image, have inhibited implementation (Goldenberg and Sharp, 2012). Moreover, as children age from dependence to emancipated minor to adult, the obligations for disclosure and data sharing will change (Taylor, 2008). There is little infrastructure in place in most academic health centers to support this change of child status in a systematic fashion.

p0360 Whether the focus is on newborn screening or screening in newborns, the implementation of large-scale genome testing in children will require considerable prenatal education of parents; an extensive infrastructure to process, analyze, and bank samples and data; and clear guidelines for the selection and delivery of results. Due to the public nature of newborn screening programs, legislative barriers may also impede development of the next generation of tests. Furthermore, even for traditional single-gene tests, there continues to be a tension between informed consent and the desire to provide the best possible care for a newborn (President's Council on Bioethics, 2008).

Incidental Findings with Genomic Sequencing s0115

p0365 As briefly described above, one of the most difficult challenges to the use of genome sequencing in clinical medicine will be managing the hundreds to thousands of incidental or secondary findings that can be generated by the analysis of each person's genome (Kohane et al., 2012). Incidental findings may be defined as variants of potential health relevance in the genome that are unrelated to the medical reason for which the testing was ordered. This notion presupposes that genome sequencing will typically be ordered for a specific genetics indication

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

(such as a symptomatic disease process or family history of known or suspected genetic disease). Following this definition, when genome sequencing is obtained in a healthy individual or newborn without a specific family history, then all findings could be incidental. But as clinical sequencing becomes available, the reality may be more complex. Individual patients will come to the experience of genome sequencing with various pre-existing perceptions of risk, and with subtle symptoms or family histories that do not imply Mendelian forms of a particular disease, but that nevertheless trigger concerns. In these cases, patients may be particularly interested or concerned about a specific condition or organ system and when genome sequencing is readily available, may specifically request interrogation of genes for that condition or that general category of condition (e.g., cancer or heart problems). Findings in response to these interrogations might be considered "quasi-incidental" as they may not have the accepted prior probabilities that are utilized in conventional genetic testing, but might nonetheless be investigated in response to some specified interest or (possibly imprecise) family history from the patient.

p0370 Decisions about the return of incidental and quasi-incidental findings in clinical medicine have been foreshadowed by an extensive debate about the return of incidental research results and findings to persons participating in genomic research. Since genotyping and sequencing in research has preceded the use of these technologies in clinical medicine, this debate has largely taken place to date in the absence of any clinical standards of care. Research subjects consistently report that they would prefer to have most or all of their results returned to them (Murphy et al., 2008), and some authors have held that returning at least some incidental findings is an ethical obligation (Fabsitz et al., 2010; Shalowitz and Miller, 2005; Wolf et al., 2008, 2012). Others have suggested that large genomics research studies should have mechanisms for subjects to express their preferences, and to request certain categories of research results in an anonymous and automated fashion, supervised by a board of experts in medicine, genetics, and ethics (Kohane et al., 2007). Caution has been urged against this enthusiasm for sharing research results with subjects, lest researchers disclose preliminary findings that have not yet been proven to be true, or somehow take on the responsibilities of clinicians by returning findings that may be valid, but are outside of research expectations or established standards of care (Clayton and McGuire, 2012).

p0375 Genome sequencing will continue to occur in research, and the return of results to research participants will continue to be controversial. But as genome sequencing enters medical care, whether performed for a conventional medical genetics analysis, for true population screening, or for some intermediate indication, there would seem to be a greater fiduciary responsibility and a more clear-cut expectation that the genome would be routinely evaluated for at least some incidental or secondary findings. Currently, there are no guidelines for return of secondary findings from clinical sequencing, although there have been proposals for lower- and higher-risk categories or "bins" based upon clinical validity and actionability (Berg et al., 2011). An informal effort to explore concordance and discordance among genetics specialists as to which incidental variants should be returned when clinical sequencing is ordered demonstrated considerable diversity of opinion (Green et al., 2012), and as of this writing, consensus statements by professional organizations are being crafted to assist laboratories and clinicians.

p0380 Simply estimating the total number of variants that might qualify for disclosure is difficult. One analysis estimated that between 4000 and 17,000 known variants could meet existing recommendations for return of incidental findings in research, and that this number would likely grow by 37% over the next four years (Cassa et al., 2012). The analysis of thousands of variants per individual is so complex that many feel it can only be reasonably done with a combination of highly curated databases and computational algorithms. The number of variants that could be meaningfully returned in any given individual is also hard to estimate because highly penetrant dominant Mendelian mutations will be quite rare, but it is likely that most people will carry between two and six recessive variants, as well as a full complement of common pharmacogenetic and common disease risk variants. An exhaustive analysis of sequencing pioneer Stephen Quake's genome revealed nine previously described rare disease variants that were either recessive or of unknown importance, four novel variants in disease genes (two recessive, two likely dominant with unclear penetrance), and a number of known and suspected pharmacogenomic variants.

p0385 In analysis of the Quake genome, variants for common diseases were superimposed upon pre-test probabilities of disease prevalence. Many have argued that the eventual interpretation of human genomes will be so complex that algorithms that make computational estimates along these lines will be required. But by any measure, the "incidentalome" looms as a daunting challenge to such efforts (Kohane et al., 2006, 2012). One analysis has suggested four categories of false-positive findings: (1) highly penetrant Mendelian variants that are incorrectly annotated in existing databases, (2) measurement errors in technical sequences, (3) unknown degrees of association for variants that are reported incidentally without the prior probabilities of symptoms or family history, and (4) false-positives resulting from the sheer volume of disease-associated variants in patients without prior high probabilities of disease. The first three of these can be addressed by the expected advances in science over the coming years, but the fourth will remain challenging (Kohane et al., 2012).

CONCLUSIONS

s0120

p0390 The rapidly falling cost of genomic technologies, coupled with the hunger for understanding oneself and the attractive social narrative that personalized genomic medicine will improve health, are accelerating the integration of genomics

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

into medical care. Yet challenges to successful integration are substantial. In the laboratory, sequencing technologies are evolving so quickly that it is difficult to understand and standardize sequencing accuracy, particularly since accuracy will vary across different regions of the genome and among different types of genomic variation. The absence of a unified publicly available database that accurately and comprehensively holds established pathogenic variants is a major impediment to the interpretation of variant information. The sheer volume and complexity of the genetic information that is potentially available in every person's genome and the fact that such information will cross specialty lines will make interpretation by a single clinician difficult. There is considerable controversy about whether the focus of genomic medicine should only be upon targeted testing, or whether it is critically important that incidental genomic findings be looked for and reported. This debate is complicated by the fact that the implications of potentially pathogenic variants will be very different depending upon whether or not there is a prior probability of a given genetic disease in the patient history, family history, or physical examination. As yet, clinicians outside of a few specialty areas are not comfortable with genomic information, and there are very little data available to help clinicians with decisions about cost/benefits or risk/benefits of follow-up tests for surveillance and other interventions. But there is every expectation that these challenges will be addressed, ushering in a new and vibrant age of genomic medicine.

ACKNOWLEDGMENTS

s0125
p0395
We thank Trevor Pugh, PhD, and Matt Lebo, PhD, for extensive assistance in the writing of this chapter, and Chris Cassa, PhD, for helpful comments. The authors are supported by NIH grants HG02213, HG006500, AG027841, HG005092, HG003178, HG00603, HG003170, HG006615, LM008748, and LM010526.

REFERENCES

ACOG Committee on Ethics, 2007. ACOG committee opinion No. 360: sex selection. Obstet Gynecol 109, 475–478.

ACOG Committee on Genetics, 2009. ACOG committee opinion No. 442: preconception and prenatal carrier screening for genetic diseases in individuals of Eastern European Jewish descent. Obstet Gynecol 114, 950–953.

Adzhubei, I., Schmidt, S., Peshkin, L., et al., 2010. A method and server for predicting damaging missense mutations. Nat Methods 7, 248–249.

Albert, T.J., Molla, M.N., Muzny, D.M., et al., 2007. Direct selection of human genomic loci by microarray hybridization. Nat Methods 4, 903–905.

American College of Medical Genetics, 2008. ACMG Laboratory Standards and Guidelines [online]. American College of Medical Genetics, Bethesda, <http://www.acmg.net/AM/Template.cfm?Section=Publications1>.

Aronson, S.J., Clark, E.H., Babb, L.J., et al., 2011. The GeneInsight Suite: a platform to support laboratory and provider use of DNA-based genetic testing. Hum Mutat 32, 532–536.

Aronson, S.J., Clark, E.H., Varugheese, M., Baxter, S., Babb, L.J., Rehm, H.L., 2012. Communicating new knowledge on previously reported genetic variants. Genet Med., 5. doi: 10.1038/gim.2012.19 [Epub ahead of print].

Ashida, S., Koehly, L.M., Roberts, J.S., Hiraki, S., Green, R.C., 2009. Disclosing the disclosure: factors associated with communicating the results of genetic susceptibility testing for Alzheimer's disease. J Health Commun 14, 768–784.

Ashida, S., Koehly, L.M., Roberts, J.S., Chen, C.A., Hiraki, S., Green, R.C., 2010. The role of disease perceptions and results sharing in psychological adaptation after genetic susceptibility testing: the REVEAL Study. Eur J Hum Genet 18, 1296–1301.

Ashley, E., Butte, A., Wheeler, M., et al., 2010. Clinical assessment incorporating a personal genome. Lancet 375, 1525–1535.

Azam, S.M., Jozic, J., 2009. Variable platelet responsiveness to aspirin and clopidogrel: role of platelet function and genetic polymorphism testing. Transl Res 154, 309–313.

Bainbridge, M.N., Wiszniewski, W., Murdock, D.R., et al., 2011. Whole-genome sequencing for optimized patient management. Sci Transl Med 3, 87re3.

Bell, C., Dinwiddie, D., Miller, N., et al., 2011. Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci Transl Med 3, 65ra4.

Bentley, D., Balasubramanian, S., Swerdlow, H., et al., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53–59.

Berg, J.S., Khoury, M.J., Evans, J.P., 2011. Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time. Genet Med 13, 499–504.

Beutler, E., Felitti, V., Koziol, J., Ho, N., Gelbart, T., 2002. Penetrance of 845G→A (C282Y) HFE hereditary haemochromatosis mutation in the USA. Lancet 359, 211–218.

Bloss, C., Schork, N., Topol, E., 2011. Effect of direct-to-consumer genomewide profiling to assess disease risk. N Engl J Med 364, 524–534.

Bonello, L., Palot-Bonello, N., Armero, S., Camoin-Jau, L., Paganelli, F., 2010. Impact of loading dose adjustment on platelet reactivity in homozygotes of the 2C19 2* loss of function polymorphism. Int J Cardiol 145, 165–166.

Brauer, C., Bozic, K., 2009. Using observational data for decision analysis and economic analysis. J Bone Joint Surg Am 91A, 73–79.

Buchdunger, E., Cioffi, C.L., Law, N., et al., 2000. Abl protein-tyrosine kinase inhibitor STI571 inhibits in vitro signal transduction mediated by c-kit and platelet-derived growth factor receptors. J Pharmacol Exp Ther 295, 139–145.

Caraco, Y., Sheller, J., Wood, A.J., 1996. Pharmacogenetic determination of the effects of codeine and prediction of drug interactions. J Pharmacol Exp Ther 278, 1165–1174.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

Cassa, C., Schmidt, B., Kohane, I., Mandl, K., 2008. My sister's keeper? Genomic research and the identifiability of siblings. BMC Med Genomics 1, 32.

Cassa, C.A., Savage, S.K., Taylor, P.L., Green, R.C., McGuire, A.L., Mandl, K.D., 2012. Disclosing pathogenic variants to research participants: quantifying an emerging ethical responsibility. Genome Res 22 (3), 421–428.

Castaigne, S., Chomienne, C., Daniel, M.T., et al., 1990. All-trans retinoic acid as a differentiation therapy for acute promyelocytic leukemia. I. Clinical results. Blood 76, 1704–1709.

Chao, S., Roberts, J.S., Marteau, T.M., Silliman, R., Cupples, L.A., Green, R.C., 2008. Health behavior changes after genetic risk assessment for Alzheimer disease: The REVEAL study. Alzheimer Dis. Assoc. Disord. 22 (1), 94–97.

Chiu, R.W., Akolekar, R., Zheng, Y.W., et al., 2011. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ 342, c7401.

Cho, A., Killeya-Jones, L., O'Daniel, J., et al., 2012. Effect of genetic testing for risk of type 2 diabetes mellitus on health behaviors and outcomes: Study rationale, development, and design. BMC Health Serv. Res. ~~18 (12), 16~~.

Choi, M., Scholl, U., Ji, W., et al., 2009. Genetic diagnosis by whole exome capture and massive parallel DNA sequencing. Proc Natl Acad Sci USA 106, 19,096–19,101.

Christensen, K., Roberts, J., Uhlmann, W., Green, R.C., 2011. Changes to perceptions of the pros and cons of genetic susceptibility testing after APOE genotyping for Alzheimer disease risk. Genet Med 13, 409–414.

Church, D.M., Schneider, V.A., Graves, T., et al., 2011. Modernizing reference genome assemblies. PLoS Biol 9, e1001091.

Clayton, E.W., McGuire, A.L., 2012. The legal risks of returning results of genomics research. Genet Med, 14. doi: 10.1038/gim.2012.10 [Epub.].

Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. 2011. Pharmacogenomics knowledge base [online]. <http://www.pharmgkb.org/index.jsp>.

Costa, T., Scriver, C.R., Childs, B., 1985. The effect of Mendelian disease on human health: a measurement. Am J Med Genet 21, 231–242.

Cotton, R.G., 2009. Collection of variation causing disease – the human variome project. Hum Genomics 3, 301–303.

Daly, A.K., Donaldson, P.T., Bhatnagar, P., et al., 2009. HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. Nat Genet 41, 816–819.

Dave, S.S., Fu, K., Wright, G.W., et al., 2006. Molecular diagnosis of Burkitt's lymphoma. N Engl J Med 354, 2431–2442.

Demetri, G.D., Von Mehren, M., Blanke, C.D., et al., 2002. Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. N Engl J Med 347, 472–480.

Devaney, S.A., Palomaki, G.E., Scott, J.A., Bianchi, D.W., 2011. Noninvasive fetal sex determination using cell-free fetal DNA: a systematic review and meta-analysis. JAMA 306, 627–636.

Downing, G.J., 2009. Key aspects of health system change on the path to personalized medicine. Transl Res 154, 272–276.

Eckert, S.L., Katzen, H., Roberts, J.S., et al., 2006. Recall of disclosed Apolipoprotein E genotype and lifetime risk estimate for Alzheimer's disease: the REVEAL study. Genet Med 8, 746–751.

Erez, A., Plunkett, K., Sutton, V.R., McGuire, A.L., 2010. The right to ignore genetic status of late onset genetic disease in the genomic era: prenatal testing for Huntington disease as a paradigm. Am J Med Genet 152A, 1774–1780.

Evans, J., Meslin, E., Marteau, T., Caulfield, T., 2011. Genomics. Deflating the genomic bubble. Science 331, 861–862.

Evans, J.P., Green, R.C., 2009. Direct to consumer genetic testing: avoiding a culture war. Genet Med 11, 1–2.

Fabsitz, R., McGuire, A., Sharp, R., 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.

Farrer, L.A., Cupples, L.A., Haines, J.L., et al., 1997. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer disease meta analysis consortium. JAMA 278, 1349–1356.

Feero, W., Guttmacher, A., Collins, F., 2010. Genomic medicine: an updated primer. N Engl J Med 362, 2001–2011.

Fitzgerald, R., Pirmohamed, M., 2011. Aspirin resistance: effect of clinical, biochemical and genetic factors. Pharmacol Ther 130, 213–225.

Flaherty, K.T., Puzanov, I., Kim, K.B., et al., 2010. Inhibition of mutated, activated BRAF in metastatic melanoma. N Engl J Med 363, 809–819.

Fokkema, I.F., Taschner, P.E., Schaafsma, G.C., Celli, J., Laros, J.F., Den Dunnen, J.T., 2011. LOVD v.2.0: the next generation in gene variant databases. Hum Mutat 32, 557–563.

Frueh, F.W., Greely, H.T., Green, R.C., Hogarth, S., Siegel, S., 2011. The future of direct-to-consumer clinical genetic tests. Nat Rev Genet 12, 511–515.

GeneTests, 2011. GeneTests: Medical Genetics Information Resource [online]. University of Washington, Seattle, WA, <http://www.genetests.org>.

Gilissen, C., Hoischen, A., Brunner, H.G., Veltman, J.A., 2011. Unlocking Mendelian disease using exome sequencing. Genome Biol 12, 228.

Gladding, P., White, H., Voss, J., et al., 2009. Pharmacogenetic testing for clopidogrel using the rapid INFINITI analyzer: a dose-escalation study. JACC Cardiovasc Interv 2, 1095–1101.

Gnirke, A., Melnikov, A., Maguire, J., et al., 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27, 182–189.

Goldenberg, A.J., Sharp, R.R., 2012. The ethical hazards and programmatic challenges of genomic newborn screening. JAMA 307, 461–462.

Goldsmith, L., Jackson, L., O'Connor, A., Skirton, H., 2012. Direct-to-consumer genomic testing: systematic review of the literature on user perspectives. Eur J Hum Genet doi: 10.1038/ejhg.2012.18 [Epub.].

Gollust, S.E., Gordon, E.S., Zayac, C., et al., 2012. Motivations and perceptions of early adopters of personalized genomics: perspectives from research participants. Publ Health Genomics 15, 22–30.

Gonzaga-Jauregui, C., Lupski, J.R., Gibbs, R., 2012. Human genome sequencing in health and disease. Ann Rev Med 63, 35–61.

Grant, R.W., Meigs, J.B., Florez, J.C., et al., 2011. Design of a randomized trial of diabetes genetic risk testing to motivate behavior change: the genetic counseling/lifestyle change (GC/LC) study for diabetes Prevention. Clin Trials 8, 609–615.

Grant, R.W., O'Brien, K.E., Waxler, J.L., et al., 2012. Personalized genetic risk counseling to motivate diabetes prevention. Diabetes Care. In Press.

Green, R.C., Roberts, J.S., Cupples, L.A., et al., 2009. Disclosure of APOE genotype for risk of Alzheimer's disease. N Engl J Med 361, 245–254.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

Green, R.C., Berg, J.S., Berry, G.T., et al., 2012. Exploring concordance and disordance for return of incidental findings from clinical sequencing. Genet Med. 14 (4), 405–410. doi: 10.1038/gim.2012.21 Epub 2012 Mar 15.

Guttmacher, A., McGuire, A., Ponder, B., Stefansson, K., 2010. Personalized genomic information: preparing for the future of genetic medicine. Nat Rev Genet 11, 161–165.

Guttmacher, A.E., Collins, F.S., 2003. Welcome to the genomic era. N Engl J Med 349, 996–998.

Hiraki, S., Chen, C., Roberts, J., Cupples, L., Green, R.C., 2009. Perceptions of familial risk in those seeking a genetic risk assessment for Alzheimer's disease. J Genet Couns 18, 130–136.

Hoischen, A., Van Bon, B., Gilissen, C., et al., 2010. De novo mutations of SETBP1 cause Schinzel–Giedion syndrome. Nat Genet 42, 483–485.

Holmes, M.V., Shah, T., Vickery, C., Smeeth, L., Hingorani, A.D., Casas, J.P., 2009. Fulfilling the promise of personalized medicine? Systematic review and field synopsis of pharmacogenetic studies. PLoS One 4, e7960.

Huang, M.E., Ye, Y.C., Chen, S.R., et al., 1988. Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. Blood 72, 567–572.

Hull, S.C., Sharp, R.R., Botkin, J.R., et al., 2008. Patients' views on identifiability of samples and informed consent for genetic research. Am J Bioeth 8, 62–70.

James, K., Cowl, C., Tilburt, J., et al., 2011. Impact of direct-to-consumer predictive genomic testing on risk perception and worry among patients receiving routine care in a preventive health clinic. Mayo Clin Proc 86 (10), 933–940.

Janssens, A., Van Duijn, C., 2009. Genome-based prediction of common diseases: methodological considerations for future research. Genome Med 1, 20.

Jonas, D.E., McLeod, H.L., 2009. Genetic and clinical factors relating to warfarin dosing. Trends Pharmacol Sci 30, 375–386.

Joneja, A., Huang, X., 2009. A device for automated hydrodynamic shearing of genomic DNA. Biotechniques 46, 553–556.

Jordan, D.M., Kiezum, A., Baxter, S.M., et al., 2011. Development and validation of a computational method for assessment of missense variants in hypertrophic cardiomyopathy. Am J Hum Genet 88, 183–192.

Kaufman, D.J., Bollinger, J.M., Dvoskin, R.L., Scott, J.A, 2012. Risky business: risk perception and the use of medical services among customers of DTC personal genetic testing. J Genet Couns 21 (3), 413–422.

Khor, B., Gardet, A., Xavier, R.J., 2011. Genetics and pathogenesis of inflammatory bowel disease. Nature 474, 307–317.

Khoury, M., Berg, A., Coates, R., Evans, J., Teutsch, S., Bradley, L., 2008. The evidence dilemma in genomic medicine. Health Aff 27, 1600–1611.

Kobelka, C.E., 2011. Sequencing: the next generation. Moving beyond population-based recessive disease carrier screening. Clin Genet 80, 25–30.

Kohane, I., Masys, D., Altman, R., 2006. The incidentalome: a threat to genomic medicine. JAMA 296, 212–215.

Kohane, I., Mandl, K., Taylor, P., Holm, I., Nigrin, D., Kunkel, L., 2007. Medicine. Reestablishing the researcher-patient compact. Science 316, 836–837.

Kohane, I.S., Hsing, M., Kong, S.W., 2012. Taxonomizing, sizing, and overcoming the incidentalome. Genet Med 14 (4), 399–404.

Kopits, I., Chen, C., Roberts, J., Uhlmann, W., Green, R., 2011. Willingness to pay for genetic testing for Alzheimer's disease: a measure of personal utility. Genet Test Mol Biomark 15, 871–875.

Kraft, P., Hunter, D.J., 2009. Genetic risk prediction – are we there yet? N Engl J Med 360, 1701–1703.

Kumar, P., Radhakrishnan, J., Chowdhary, M.A., Giampietro, P.F., 2001. Prevalence and patterns of presentation of genetic disorders in a pediatric emergency department. Mayo Clin Proc 76, 777–783.

Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4, 1073–1081.

Kwak, E.L., Bang, Y.J., Camidge, D.R., et al., 2010. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med 363, 1693–1703.

Lam, H.Y., Clark, M.J., Chen, R., et al., 2012. Performance comparison of whole-genome sequencing platforms. Nat Biotechnol 30, 78–82.

Lander, E.S., Linton, L.M., Birren, B., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

LaRusse, S., Roberts, J., Marteau, T., et al., 2005. Genetic susceptibility testing versus family history-based risk assessment: impact on perceived risk of Alzheimer disease. Genet Med 7, 48–53.

Lenfant, C., 2003. Shattuck lecture: clinical research to clinical practice – lost in translation? N Engl J Med 349, 868–874.

Li, H., Homer, N., 2010. A survey of sequence alignment algorithms for next-generation sequencing. Brief Bioinform 11, 473–483.

Linnenbringer, E., Roberts, J., Hiraki, S., Cupples, L., Green, R.C., 2010. "I know what you told me, but this is what I think:" perceived risk of Alzheimer disease among individuals who accurately recall their genetics-based risk estimate. Genet Med 12, 219–227.

Lo, S.S., Mumby, P.B., Norton, J., et al., 2010a. Prospective multicenter study of the impact of the 21-gene recurrence score assay on medical oncologist and patient adjuvant breast cancer treatment selection. J Clin Oncol 28, 1671–1676.

Lo, Y., Chan, K., Sun, H., et al., 2010b. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med 2 61ra91.

Lupski, J., Reid, J., Gonzaga-Jauregui, C., et al., 2010. Whole-genome sequencing in a patient with Charcot–Marie–Tooth Neuropathy. N Engl J Med 362, 1181–1191.

Lynch, T.J., Bell, D.W., Sordella, R., et al., 2004. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350, 2129–2139.

MacConaill, L.E., Garraway, L.A., 2010. Clinical implications of the cancer genome. J Clin Oncol 28, 5219–5228.

Malpas, P., 2005. The right to remain in ignorance about genetic information – can such a right be defended in the name of autonomy? NZ Med J 118, U1611.

Manolio, T.A., Collins, F.S., Cox, N.J., et al., 2009. Finding the missing heritability of complex diseases. Nature Review 461 (7265), 747–753.

Mardis, E.R., 2008. The impact of next-generation sequencing technology on genetics. Trends Genet 24, 133–141.

Margulies, M., Egholm, M., Altman, W.E., et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., et al., 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev. Genet. 9 (5), 356–369.

McCormack, M., Alfirevic, A., Bourgeois, S., et al., 2011. HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. N Engl J Med 364, 1134–1143.

McDermott, U., Downing, J., Stratton, M., 2011. Genomics and the continuum of cancer care. N Engl J Med 364, 340–350.

McGuire, A., Diaz, C., Wang, T., Hilsenbeck, S., 2009. Social networkers' attitudes toward direct-to-consumer personal genome testing. Am J Bioeth 9, 3–10.

McGuire, A.L., Burke, W., 2008. An unwelcome side effect of direct-to-consumer personal genome testing: raiding the medical commons. JAMA 300, 2669–2671.

McKernan, K.J., Peckham, H.E., Costa, G.L., et al., 2009. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. Genome Res 19, 1527–1541.

Mega, J.L., Close, S.L., Wiviott, S.D., et al., 2010a. Genetic variants in ABCB1 and CYP2C19 and cardiovascular outcomes after treatment with clopidogrel and prasugrel in the TRITON-TIMI 38 trial: a pharmacogenetic analysis. Lancet 376, 1312–1319.

Mega, J.L., Simon, T., Collet, J.P., et al., 2010b. Reduced-function CYP2C19 genotype and risk of adverse clinical outcomes among patients treated with clopidogrel predominantly for PCI: a meta-analysis. JAMA 304, 1821–1830.

Meyerson, M., Gabriel, S., Getz, G., 2010. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet 11, 685–696.

Mok, T.S., Wu, Y.L., Thongprasert, S., et al., 2009. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 361, 947–957.

Moussa, I., 2011. Coronary artery bifurcation interventions: the disconnect between randomized clinical trials and patient-centered decision-making. Catheter Cardiovasc Interv 77, 537–545.

Murphy, J., Scott, J., Kaufman, D., Geller, G., LeRoy, L., Hudson, K., 2008. Public expectations for return of results from large-cohort genetic research. Am J Bioeth 8, 36–43.

National Heart, Lung and Blood Institute, 2012. NHLBI GO Exome Sequencing Project (ESP): The Exome Variant Server. University of Washington. <http://evs.gs.washington.edu/EVS/>.

National Institute of Child Health and National Human Genome Research Institute. 2010. Newborn Screening in the genomic era: setting a research agenda, 13–14 December, 2010, Rockville, MD. Meeting Summary [online]. Newborn Screening Translational Research Network. <http://www.genome.gov/Pages/PolicyEthics/StaffArticles/Newborn_Screening_Meeting_Summary.pdf>.

Ng, P.C., Henikoff, S., 2003. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 31, 3812–3814.

Ng, S.B., Turner, E.H., Robertson, P.D., et al., 2009. Targeted capture and massively parallel sequencing of twelve human exomes. Nature 461, 272–276.

Ng, S., Buckingham, K., Lee, C., et al., 2010a. Exome sequencing identifies the cause of a Mendelian disorder. Nat Genet 42, 30–35.

Ng, S.B., Bigham, A.W., Buckingham, K.J., et al., 2010b. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 42, 790–793.

Offit, K., Groeger, E., Turner, S., Wadsworth, E., Weiser, M., 2004. The "duty to warn" a patient's family members about hereditary disease risks. JAMA 292, 1469–1472.

O'Connell, M.J., Lavery, I., Yothers, G., Paik, S., Clark-Langone, 2010. Relationship between tumor gene expression and recurrence in four independent studies of patients with stage II/III colon cancer treated with surgery alone or surgery plus adjuvant fluorouracil plus leucovorin. J. Clin. Oncol. 28 (5), 3937–3944.

O'Roak, B.J., Deriziotis, P., Lee, C., et al., 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 43, 585–589.

Paez, J.G., Janne, P.A., Lee, J.C., et al., 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304, 1497–1500.

Paik, S., Shak, S., Tang, G., et al., 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351, 2817–2826.

Palomaki, G.E., Kloza, E.M., Lambert-Messerlian, G.M., et al., 2011. DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. Genet Med 13, 913–920.

Palomaki, G.E., Deciu, C., Kloza, E.M., et al., 2012. DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study. Genet Med, 14. doi: 10.1038/gim.2011.73 [Epub.].

Pao, W., Wang, T.y., Riely, G.I., Miller, V.A., Pan, Q., Ladanyi, M., et al., 2005. KRAS mutations and primary resistence of lung adenocarcinomas to gefitinib or erlotinib. PLoS Med. 2 (1), e17.

Patwardhan, R.P., Lee, C., Litvin, O., Young, D.L., Pe'er, D., Shendure, J., 2009. High-resolution analysis of DNA regulatory elements by synthetic saturaton mutagenesis. Nat Biotechnol 27, 1173–1175.

Petitti, D., Teutsch, S., Barton, M., Sawaya, G., Ockene, J., DeWitt, T., 2009. Update on the methods of the U.S. Preventive Services Task Force: Insufficient evidence. Ann Intern Med 150, 199–205.

Pirmohamed, M., 2010. Pharmacogenetics of idiosyncratic adverse drug reactions. Handb Exp Pharmacol (196), 477–491.

Pirmohamed, M., 2011. Pharmacogenetics: past, present and future. Drug Discov Today 16 (19–20), 852–861.

Pleasance, E., Cheetham, R., Stephens, P., et al., 2010. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 463, 191–196.

Porreca, G.J., Zhang, K., Li, J.B., et al., 2007. Multiplex amplification of large sets of human exons. Nat Methods 4, 931–936.

Prenen, H., Guetens, G., De Boeck, G., et al., 2006. Cellular uptake of the tyrosine kinase inhibitors imatinib and AMN107 in gastrointestinal stromal tumor cell lines. Pharmacology 77, 11–16.

President's Council on Bioethics. 2008. The changing moral focus of newborn screening: an ethical analysis by the President's Council on Bioethics – Chapter 3: The future of newborn screening. <http://bioethics.georgetown.edu/pcbe/reports/newborn_screening/>.

Price, M.J., Berger, P.B., Teirstein, P.S., et al., 2011. Standard- vs high-dose clopidogrel based on platelet function testing after percutaneous coronary intervention: the GRAVITAS randomized trial. JAMA 305, 1097–1105.

Ratner, M., 2012. Genomics through the lens of precompetitive data sharing. IN VIVO 30, 46–50.

Relling, M., Klein, T.E., 2011. CPIC: clinical pharmacogenetics implementation consortium of the pharmacogenomics research network. Clin Pharmacol Ther 89, 464–467.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

Relling, M., Altman, R., Goetz, M., Evans, W., 2010. Clinical implementation of pharmacogenomics: overcoming genetic exceptionalism. Lancet Oncol 11, 507–509.

Relling, M.V., Hancock, M.L., Rivera, G.K., et al., 1999. Mercaptopurine therapy intolerance and heterozygosity at the thiopurine S-methyltransferase gene locus. J Natl Cancer Inst 91, 2001–2008.

Roberts, J.S., LaRusse, S.A., Katzen, H., et al., 2003. Reasons for seeking genetic susceptibility testing among first-degree relatives of people with Alzheimer's Disease. Alzheimer's Dis Assoc Dis 17, 86–93.

Roberts, J.S., Barber, M., Brown, T.M., et al., 2004. Who seeks genetic susceptibility testing for Alzheimer's disease? Findings from a multisite, randomized clinical trial. Genet Med 6, 197–203.

Rosenwald, A., Wright, G., Chan, W.C., et al., 2002. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 346, 1937–1947.

Rubenstein, W., Jiang, H., Dellafave, L., Rademaker, A., 2009. Cost-effectiveness of population based BRCA 1/2 testing and ovarian cancer prevention for the Ashkenazi Jews: a call for dialogue. Genet Med 11, 629–639.

Samuels, M.E., Rouleau, G.A., 2011. The case for locus-specific databases. Nat Rev Genet 12, 378–379.

Schaeffeler, E., Fischer, C., Brockmeier, D., et al., 2004. Comprehensive analysis of thiopurine S-methyltransferase phenotype-genotype correlation in a large population of German-Caucasians and identification of novel TPMT variants. Pharmacogenetics 14, 407–417.

Shalowitz, D.I., Miller, F.G., 2005. Disclosing individual results of clinical research: implications of respect for participants. JAMA 294, 737–740.

Sharp, R., 2011. Downsizing genomic medicine: approaching the ethical complexity of whole-genome sequencing. Genet Med 13, 191–194.

Stenson, P., Mort, M., Ball, E., et al., 2009. The human gene mutation database: 2008 update. Genome Med 1, 13.

Stratton, M., Campbell, P., Futreal, P., 2009. The cancer genome. Nature 458, 719–724.

Takala, T., 1999. The right to genetic ignorance confirmed. Bioethics 13, 288–293.

Taylor Jr., D.H., Cook-Deegan, R.M., Hiraki, S., Roberts, J.S., Blazer, D.G., Green, R.C., 2010. Genetic testing for Alzheimer's and long-term care insurance. Health Aff (Millwood) 29, 102–108.

Taylor, P., 2008. Personal genomes: when consent gets in the way. Nature 456, 32–33.

Tewhey, R., Warner, J.B., Nakano, M., et al., 2009. Microdroplet-based PCR enrichment for large-scale targeted sequencing. Nat Biotechnol 27, 1025–1031.

Tong, M.Y., Cassa, C.A., Kohane, I.S., 2011. Automated validation of genetic variants from large databases: ensuring that variant references refer to the same genomic locations. Bioinformatics 27, 891–893.

Travis, S., 2006. New thinking: theory vs practice. A case study illustrating evidence-based therapeutic decision making. Colorectal Dis 8 (Suppl. 1), 25–29.

Van't Veer, L.J., Dai, H., Van de Vijver, M.J., et al., 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415, 530–536.

Varmus, H., 2010. Ten years on: the human genome and medicine. N Engl J Med 362, 2028–2029.

Venter, J.C., Adams, M.D., Myers, E.W., et al., 2001. The sequence of the human genome. Science 291, 1304–1351.

Vernarelli, J.A., Roberts, J.S., Hiraki, S., Chen, C.A, Cupples, L.A., Green, R.C., 2010. Effect of Alzheimer's disease genetic risk disclosure on dietary supplement use. Am J Clin Nutr 91, 1402–1407.

Visscher, P.M., Brown, M.A., McCarthy, M.I., Yang, J., 2012. Five years of GWAS discovery. Am. J. Hum. Genet. 90 (1), 7–24.

Vogel, C.L., Cobleigh, M.A., Tripathy, D., et al., 2002. Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. J Clin Oncol 20, 719–726.

Waisbren, S.E., Albers, S., Amato, S., et al., 2003. Effect of expanded newborn screening for biochemical genetic disorders on child outcomes and parental stress. JAMA 290, 2564–2572.

Wang, Y., Klijn, J.G., Zhang, Y., et al., 2005. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 365, 671–679.

Weiss, S.T., McLeod, H.L., Flockhart, D.A., et al., 2008. Creating and evaluating genetic tests predictive of drug response. Nat Rev 7, 568–574.

Wilson, J., 2005. To know or not to know? Genetic ignorance, autonomy and paternalism. Bioethics 19, 492–504.

Wilson, J., Jungner, F., 1968. Principles and practice of screening for disease Public Health Papers. World Health Organization, Geneva.

Wolf, S.M., Lawrenz, F.P., Nelson, C.A., et al., 2008. Managing incidental findings in human subjects research: analysis and recommendations. J Law Med Ethics 36 219–248, 211.

Wolf, S.M., Crock, B.N., Van Ness, B., et al., 2012. Managing incidental findings and research results in genomic research involving biobanks and archived datasets. Genet Med 14 (4), 361–384.

Wood, H.M., Belvedere, O., Conway, C., et al., 2010. Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens. Nucleic Acids Res 38, e151.

Worthey, E.A., Mayer, A.N., Syverson, G.D., et al., 2011. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med 13, 255–262.

Yates, C.R., Krynetski, E.Y., Loennechen, T., et al., 1997. Molecular diagnosis of thiopurine S-methyltransferase deficiency: genetic basis for azathioprine and mercaptopurine intolerance. Ann Intern Med 126, 608–614.

Yin, T., Miyata, T., 2011. Pharmacogenomics of clopidogrel: evidence and perspectives. Thromb Res 128, 307–316.

Zick, C.D., Mathews, C.J., Roberts, J., Cook-Deegan, R., Pokorski, R.J., Green, R.C., 2005. Genetic testing for Alzheimer's disease and its impact on insurance purchasing behavior. Health Aff 24, 483–490.

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.

Abstract

p9000 Genomic medicine has arrived and is being rapidly integrated into medicine due to the reduced costs and increasing sophistication of next-generation sequencing. This chapter describes the laboratory sequencing, informatics analysis, and clinical applications that will face those using whole-exome and whole-genome sequencing in the practice of medicine. In anticipation of the widespread availability of genomes in the near future, particular emphasis is placed upon the challenges clinicians will face in dealing with incidental or secondary genomic findings and in communicating ambiguous risk information.

p9010 **Key Words**

p9015 direct-to-consumer, exome sequencing, general genome report, incidental findings, medical genomics, mutation, variant, whole-genome sequencing

To protect the rights of the author(s) and publisher we inform you that this PDF is an uncorrected proof for internal business use only by the author(s), editor(s), reviewer(s), Elsevier and typesetter MPS. It is not allowed to publish this proof online or in print. This proof copy is the copyright property of the publisher and is confidential until formal publication.