

## Knocking on the clinic door

High-throughput sequencing for clinical purposes faces technical and quality challenges, but it's worth it.

The \$3-billion price tag for the first human genome would now buy not one but a million human genome sequences, each completed in just a few weeks. Personal genome sequencing is becoming a reality, and targeted or whole exome sequencing is being explored to facilitate diagnosis and guide treatment, in some conditions and for some patients. The problem is that extracting clinically actionable information from genome data is currently hit or miss, time intensive and dependent on access to knowledgeable specialists. What's more, much of the IT infrastructure and decision support systems necessary to deliver genome information to physicians has yet to be put in place.

This issue of *Nature Biotechnology* summarizes the current status of high-throughput sequencing as applied to life sciences and medical research. The notion of simultaneously accessing all genes associated with all diseases through a patient's genome is both tantalizing and daunting: tantalizing because it could potentially provide a path to patient-centered medicine; daunting because of the sheer scale of data involved and also the numerous changes required to clinical research laboratory practice, accreditation and standards, regulatory oversight, and issues relating to ethics, privacy, consent and legal protections.

The requirements for bench sequencing and clinical sequencing differ markedly in at least two respects. First, test results must be delivered rapidly (because patients cannot wait months for results). Second, the data must be distilled to facilitate clinical decision making (physicians already wrestle with information overload). Indeed, one reason why pharmacogenetic testing has been narrowly adopted is that doctors won't order tests when the patient is in front of them waiting for a prescription. To be useful, pharmacogenetic testing must be undertaken *a priori*, enabling patients' genetically predicted drug responses to guide prescription.

Adapting sequencing assays to clinical work will also require higher levels of sensitivity and specificity than research. In most research applications, a 30% false-negative rate (missed variants) and an even higher false-positive rate (erroneous variants called) is the norm. To approach the 99.9% raw base-calling accuracy needed for many medical applications, however, an average sequence depth would have to approach 30× coverage (~10× coverage is acceptable in research), with >95% of all calls made at a depth of ten reads or more.

To ensure such quality of sequence coverage, laboratories conducting sequencing services will have to be accredited under the Clinical Laboratory Improvement Amendments (CLIA)—and to equivalent standards outside the United States. This explains the relationship between 23andMe and LabCorp, the acquisition of Navigenics by Life Technologies and in September the acquisition of Complete Genomics by BGI. At the same time, proficiency testing standards, such as those released by the College of American Pathologists in July and those presented on p. 1034, will be essential to assure quality sequencing.

But assuring the quality and proficiency of the platforms is not enough. Most of the algorithms for aligning reads, detecting variants and calling

genotypes are imperfect. Large copy number deletions and trinucleotide repeats in short read data are fundamentally difficult to detect. Biological variability introduced from samples can also confound analysis and interpretation (e.g., formalin fixation can crosslink cytosines and somatic mosaicism can result in misleading calls).

But an even more fundamental problem lies in determining which of the millions of variants in an individual's genome sequence are of clinical validity (that is, give information about the patient) and, more importantly, which are also of clinical utility (that is, aid treatment decisions). According to current estimates, only 10–30% of the variants identified by sequencing have clinical validity, and fewer still are actionable.

Identifying variants of clinical significance is very complex technically and far from perfect. ENCODE has revealed to us prolific transcription of noncoding regions. But we struggle to filter even just the thousands of variants revealed by exome sequencing against reference sets, such as common variants identified in the HapMap project, the US National Heart, Lung and Brain Institute exome variant set or the Complete Genomics 69-sample set. This yields hundreds of variants altering protein structure that still need to be screened for likely candidates by painstaking literature research. But the sifting process is imperfect because the reference sets do not truly represent the spectrum of variation in different ethnic groups (1000 Genomes attempts to address this). Even then, the multigenic complexity of many disorders complicates analysis.

In short, apart from applications in cancer diagnosis and therapy, the major immediate clinical benefit from sequencing will not arise from personal genomes, but from an increased rate of disease-gene discovery, particularly in undiagnosed patients with putative monogenic disorders. The unserved market today lies in turnkey genome analysis and genome interpretation for the clinic, a niche that companies such as Silicon Valley Biosystems, Knome, Cypher Genomics, Personalis, Genomatix, Omicia, Ingenuity Systems and Station X seek to occupy.

In diagnostics, CLIA labs will increasingly use sequencing to test multiple gene panels rapidly and simultaneously—and diagnostic companies will begin to develop multiplexed gene test kits.

But the rate of progress will depend on payers' willingness to embrace the technology. Already, a whole-genome sequence costs the same as a handful of single-gene tests. But unlike a gene test, the value of a genome increases with time (as research identifies more causal variants). There are immediate savings for patients who currently confound traditional diagnosis, going from clinical center to center receiving lots of costly and futile treatments. But what insurers really need to see is that one outlay also opens up cost-savings down the line. Truly, a patient's genome is a gift that keeps on giving.

*Nature Biotechnology* is grateful to sponsor Life Technologies, whose support enables this focus to be freely available for the next 6 months.