Translating the cancer genome into clinically useful tools and strategies

Federico Innocenti and Richard L. Schilsky

Advances in the field of genomics are opening unprecedented opportunities to revolutionize the practice of medicine. The acquisition of human genomic information from the Human Genome Project, the Haplotype Map Project and, more recently, the 1000 Genomes Project, has proceeded at a fast pace, thanks to large investments by governments and foundations, international cooperation, and technological improvements. The information contained in human germline DNA has now been decoded and extensively applied to decipher the genetic basis of diseases, dramatically improving our knowledge of human pathobiology. Similarly, primary tumor material used for the initial diagnosis of cancer can be used to interrogate the tumor DNA, which contains, in addition to the heritable germline information of the host, the acquired somatic DNA changes that occurred during the neoplastic transformation of non-cancerous cells. In parallel with the initiatives aimed at deciphering the code of the heritable human genome, several consortia have now focused their efforts to decode the map of somatic DNA changes in tumors.

According to recent data from the National Center for Health Statistics (http://www.cdc.gov/nchs/), cancer-related deaths have not decreased appreciably over the last 60 years, unlike deaths from heart disease and many other diseases. Given the enormous burden of cancer on public health and the cost of medical care, it is evident that new strategies are needed to reduce cancer incidence, improve early detection, and develop more effective and less toxic therapies. We believe that these goals can be achieved by using the results of the current efforts to catalogue the DNA aberrations in the human cancer genome. Sequencing of the cancer genome was mentioned first by Dulbecco in 1986 (Dulbecco, 1986) as a possible strategy to improve our understanding of cancer etiology and to reduce its incidence and mortality. Systematic approaches to decode the cancer genome have been launched, and we now have available resources and networks (International Cancer Genome Consortium, http://www.icgc.org/; Cancer Genome Anatomy Project, http://cgap.nci.nih.gov/; The Cancer Genome Atlas, http://cancergenome.nih.gov/; Cancer Genome Project, http://www.sanger.ac.uk/genetics/CGP/; Tumor Sequencing Project, http://www.genome.gov/19517442#1; and others) that use high-throughput sequencing to characterize DNA alterations across thousands of tumors.

The results of these early efforts were reviewed recently (Chin and Gray, 2008; Stratton et al., 2009) and can be summarized as follows: (1) most cancer types display substantial heterogeneity at the genetic level, with several distinct classes of DNA changes including point mutations, insertion/deletions, rearrangements, and copy number alterations; (2) there is high heterogeneity in the prevalence of point mutations and rearrangements across different tumor types, with some mutated genes that are frequent in diverse types of cancer but rare in others; (3) ‘positive controls’, that is, point-mutated genes known to be involved in cancer pathogenesis like \textit{BRAF}, \textit{PIK3CA}, \textit{EGFR} and others, have been identified through these unbiased genomic efforts; (4) point mutations have been classified as ‘drivers’ or ‘passengers’, the latter being more frequent; (5) the recent catalogue of driver mutations indicate that they are present in at least 1.6% of...
genes in the genome (Futreal et al., 2004), with 10% of them also found in the germline DNA of the host.

Further development of statistically rigorous methods, as well as experimental evidence of function, are needed to distinguish mutations that influence or drive disease, known as driver mutations, from the passenger mutations, which are not associated with disease but may still be detected with genetic testing. Moreover, the insights obtained from the whole genome sequencing of glioblastoma multiforme (Parsons et al., 2008) suggest that the genes that are altered frequently are the ‘usual suspects’ (i.e. those that are known previously to be functionally mutated), and that there is a plethora of less frequently altered genes that are difficult to detect. More than a one million-fold improvement in sequencing coverage has occurred in the last 30 years. In a few years, we expect to be able to sequence more than the 100 billion base pairs of DNA that are probably required to identify the full catalogue of somatic mutations, including rare mutations. The generation of cheaper and more efficient high-throughput technologies will continue to accelerate our understanding of tumor biology.

We do not yet have a complete compilation of the alterations that occur in the most common tumor types, as we have just started scratching the surface of the problem. Cancer-specific repositories (Cancer Gene Census, http://www.sanger.ac.uk/genetics/CGP/Census/; Catalogue of Somatic Mutations in Cancer, http://www.sanger.ac.uk/genetics/CGP/cosmic/; and others) aim to compile this type of information. The molecular data contained in these databases have great biological value but are not supported sufficiently by data on gene annotation and regulation, pathways of molecular interactions, and by clinical data, to be clinically useful. Data sharing and integration among different consortia will be the key to achieve this objective more rapidly. A comprehensive catalogue of cancer-associated mutations is expected to eventually provide genomic data that is useful to health practitioners and patients. However, without data on the clinical impact of the information carried by a certain tumor marker, the availability of public repositories of somatic mutations will not be sufficient to improve the practice of oncology. Hence, the utility of the databases of tumor mutations will be augmented dramatically by linking the information on the mutations with the evidence supporting their association with clinical outcome, thereby providing practicing physicians with the necessary elements to reach informed decisions about treatment strategies for their patients.

The results of the first whole-genome scan of a tumor sample, which came from acute myeloid leukemia, were released in 2008 (Ley et al., 2008). There is no doubt that similar results from other cancer genome-sequencing efforts can be applied to improve patient care through better detection measures and more effective targeted treatments. This goal can be achieved by using biomarkers for early detection, assessment of prognosis and selection of therapy. Indeed, we have already begun using molecular markers in the clinic, both from germline DNA and tumor RNA/DNA (Chau et al., 2008), to improve the outcome of patients and their survival. For example, using epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors selectively in patients with non-small-cell lung cancer whose tumors have sensitizing EGFR mutations, improves outcomes for these patients (Eberhard et al., 2008). We can also reduce the risk of life-threatening chemotherapy toxicity through knowledge of the germline genetic make-up of patients. For example, prospectively modifying the dose of irinotecan in patients with polymorphisms of the UGT1A1 gene, which are associated with impaired clearance and increased toxicity of this drug, may allow the drug to be given more safely (Kim and Innocenti, 2007). However, information has been (and is being) gathered from the cancer genome at a rate that is much faster than the process required for genomic
molecular markers to become validated and used to guide prevention, detection and treatment strategies.

Assume that a practicing physician receives the complete catalogue of tumor alterations that have been contrasted against the germline genomic information of a metastatic colorectal cancer patient. The physician would like to use these molecular data to reach an informed decision about the optimal treatment program for the patient. Which of the thousands of bits of molecular data will the physician will be able to use? Only a few validated markers have clinical utility, e.g. KRAS from the tumor genome can guide the prescription of anti-EGFR monoclonal antibodies (Walther et al., 2009). The remaining information from the sequenced genome requires a rigorous process of validation before markers are selected for clinical use. This process is divided into two phases of validation: experimental/biological and clinical (Chin and Gray, 2008).

The discoveries that emerge from sequencing efforts have been based mainly upon statistical association, but the ability of the identified mutations to cause cancer in experimental models needs to be subsequently determined. Mechanistic studies in preclinical model systems are necessary to validate the biological importance of the observed associations. Multidimensional data from several models, both in vitro and in vivo, can be compared in order to select markers for the next step of clinical validation. It is not clear yet whether this process should be run in parallel with the clinical validation or sequentially. If a novel candidate marker falls in a pathway that already has compelling evidence of biological significance in cancer, then clinical validation might run in parallel with experimental testing. Otherwise, experimental validation will need to precede clinical testing.

Clinical validation is a very lengthy and expensive process (Chau et al., 2008; Hodgson et al., 2009). In the past, the development of an effective therapeutic against a specifically mutated target took decades from its first identification as an oncogene. For example, in 2001, the U.S. Food and Drug Administration (FDA) approved imatinib for BCR/Abl mutations in chronic myelogenous leukemia despite the discovery and characterization of the causative Philadelphia chromosome in the 1960s and 1970s (Nowell and Hungerford, 1960; Rowley, 1973). Similarly, the FDA approved the first angiogenesis inhibitor in 2004, more than 30 years after the first evidence of the importance of the angiogenic process in tumor progression (Folkman, 1971). We can learn from past experience in the development of biomarkers to find solutions to accelerate this process. It is conceivable that the progress made in molecular biology techniques and the use of high-throughput library screening will reduce the time between oncogene discovery and the selection of chemical leads for preclinical and clinical testing. It is also expected that biomarker-drug co-development will shorten the lengthy process of validating biomarkers after drug marketing, as the likelihood of improving outcomes (compared with the standard of care) is increased in biomarker-enriched patient populations.

Before clinical testing, analytical validation of the performance of the assay should be obtained to guarantee reproducibility and accuracy, and to determine its sensitivity/specificity. As the assays for analytical validation are more complicated than the typical bioanalytical assay, the role of molecular pathology labs is crucial in selecting the appropriate sample matrices and assays, as well as to ensure the maintenance of sample integrity and the reproducibility of assay results.

The most pragmatic and rigorous approach for clinical validation is to design a prospective clinical trial where patients are stratified according to the biomarker status (positive/negative) and each arm is randomized to standard-of-care versus biomarker-driven targeted therapies; such a design would also determine if a biomarker has prognostic or predictive value or perhaps both. Alternatively, a somewhat less rigorous,
but often more convenient, approach is to use specimens and data collected previously from large controlled trials for prospective independent validation of the clinical performance of a biomarker. It is necessary to carefully determine the magnitude of the clinical improvement to be targeted, as this will dictate the sample size and the power to detect differences that are clinically meaningful. Consortia of different cancer centers are the ideal setting for biomarker validation, both retrospectively and prospectively. Large cancer consortia have a large volume of patients, an already-established infrastructure for data management, compliance with regulatory requirements, sample collection, and access to Clinical Laboratory Improvement Amendments (CLIA)-certified labs.

The current use of biomarkers is clearly changing the traditional paradigms of cancer treatment and drug development. The completion of the large-scale cancer genome sequencing projects will lead to the identification of novel molecular pathways of carcinogenesis, new classification of heterogeneous disease entities, and the elucidation of molecular mechanisms of tumor evolution. Novel targets can be selected for developing molecularly targeted anti-cancer agents and for tailoring individualized therapies to patients. We remain confident that we will be able to accelerate the translation of molecular markers into the clinic. We expect that the results of the cancer genome will have an even bigger impact in improving disease management, and that the number of clinically useful molecular markers will continue to expand in the near future. Although the use of tumor DNA sequencing itself as the tool for interrogation of a patient’s tumor is still premature (the cost of sequencing an entire genome should drop to at least US$1000), next-generation sequencing technologies will allow for more rapid analyses at lower cost, providing the signature of the molecular alterations in each patient’s tumor. We anticipate that the experience with gene expression profiling in breast cancer (Sotiriou and Pusztai, 2009) will pave the way for efficient clinical validation of tumor sequencing signatures in the future. Armed with this information, oncologists will achieve the vision of personalized medicine, to deliver the right treatment to the right person at the right time.

REFERENCES


