Chinese Journal of Cancer

Perspective

A proposal: a comprehensive platform to characterize tumors in Chinese and improve success in cancer drug discovery and development

Pearl S. Huang[1], Peter T. Ho[1] and Kang Zhang[2,3]

Abstract

Cancer is a collection of complex diseases in which cell proliferation and apoptosis are dysregulated due to the acquisition of genetic changes in cancer cells. These genetic changes, combined with the interrelated physiologic adaptations of neo-angiogenesis, recruitment of stromal support tissues, and suppression of immune recognition, are measurable characteristics in tumor gene expression profiles and biochemical pathways. These measures can lead to identification of disease drivers and, ultimately, can be used to assign therapy. With advances in RNA sequencing technologies, the ability to simultaneously measure all genetic and gene expression changes with a single technology is now possible. The ability to create a comprehensive catalog of genotypic and phenotypic changes in a collection of histologically similar but otherwise distinct tumors should allow for a more precise positioning of existing targeted therapies and identification of new targets for intervention.

Key words Tumor genetics, whole genome profiling, pathway signature, RNA sequencing, oncogenes

The incidence of cancer in China is high and increasing. The most prevalent cancer types worldwide are lung cancer, colorectal cancer, breast cancer, and prostate cancer. The prevalence of endothelial growth factor receptor (*EGFR*) mutations in the female non-smoking patients with lung cancer in Asia is higher from those in other continents[1]. However, the increasing incidence of gastric, esophageal, and hepatocellular cancers in China and other Asian countries are also distinct from those in western countries, and represent significant unmet medical needs[2]. In addition, small cohorts of Asian-specific cancers include oral cancer in India, bladder cancer in Taiwan, cholangiocarcinoma in Thailand, and nasopharyngeal carcinoma in southern China and Southeast Asia. Although progress has been made worldwide in treating specific types of cancer, current cancer therapies have variable and generally low success rates, with the majority of patients, especially adults having carcinomas, receiving treatment that remains ineffective.

On a population level, cancer is a heterogeneous disease, which has traditionally been defined by histological characteristics and the tissue site of the primary tumor. Collectively, approximately 200 subtypes of cancer exist based on the current definition of disease, and it is estimated that over 1000 additional subtypes remain to be defined.

Cancer is a disease class in which patients benefit from a personalized medicine approach. Gleevec[3] and Herceptin[4], which are both antagonists of specific oncogenes, are multi-billion dollar drugs which were developed in subpopulations of patients with histologically and genetically defined cancers. Oncogenes are defined as genes that are genetically deranged, either by mutation, amplification, or chromosomal translocation. The mutant genes or changes in gene expression create a gain of function, typically resulting in either constitutive activation of a protein (the drug target) or acquisition of

Authors' Affiliations: [1]Beigene Ltd., Beijing 102206, P. R. China; [2]Molecular Medicine Center, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, P. R. China; [3]Institute for Genomic Medicine and Shiley Eye Center, University of California at San Diego, La Jolla, CA 92093, USA.

Corresponding Author: Pearl S. Huang, 7 Life Science Park Road, ZhongGuanCun Life Science Park, Changping District, Beijing 102206, P. R. China; Email: pearl.huang@beigene.com.

neomorphic traits that drive tumorigenesis. The first gene identified to have these characteristics was the *myc* oncogene, followed shortly by the identification of mutant *K-Ras*. Recognition of this basic concept of cancer biology is reflected in the Nobel Prize in Physiology or Medicine, which was awarded to Bishop and Varmus in 1989[5]. Since then, the attempt to translate this concept into therapy has proceeded with prominence. This fundamental rule of cancer biology has been defined as "oncogene addiction"[6], and the differential efficacy of therapies on cancers with specific oncogene mutations has been referred to the "genetic therapeutic index"[7].

Using Knowledge of Oncogenes and Tumor Suppressor Genes for Cancer Drug Discovery

To date, over 300 candidate oncogenes and tumor suppressor genes have been proposed across multiple cancer types. Identification of new oncogenes, tumor suppressor genes, and pathways has been greatly enabled by whole genome or exon DNA sequencing and mRNA profiling technologies. Although either approach is quite useful, neither alone is sufficient to fully identify the causal or driver genes in a single tumor. As cancer is a disease of genomic instability, hundreds of genetic changes are observed in each tumor, with the majority being passenger rather than causal in nature. By definition, passenger mutations are neither diagnostic for the disease state, nor do they represent meaningful targets for drug discovery. Therefore, simply measuring gene expression levels or cataloging DNA mutations in a tumor is insufficient to understand the etiology of disease. A recent report describing the whole genome sequencing analysis of 11 breast and 11 colorectal cancers identified an average of 80 DNA mutations that result in amino acid changes in a tumor[8]. Further analysis suggested that < 15 of these mutations contribute to disease etiology. The identity and prevalence of individual mutations that drive critical pathways is variable from tumor to tumor. Thus, the significance of these changes needs to be verified from the same sample with an alternative method of measurement that captures the phenotypic changes accompanying driver mutations. In an integrated analysis, whole genome mRNA profiling and targeted DNA analysis for over 200 glioblastoma samples successfully stratified each histologically defined glioblastoma subtype on the basis of association with known oncogene drivers[9]. Although this study was not designed to identify new oncogenes for glioblastoma, the association of *PDGFRA*, *EGFR*, and *IDH1* mutations with specific subtypes suggests that inhibitors for these targets should be positioned within known subtypes to achieve greater clinical efficacy.

Ironically, although the oncogenes *myc* and *ras* have clearly been identified as drivers of cancer and are therefore excellent theoretical targets for therapeutic intervention, a therapy that directly targets either of these proteins has not been delivered to patients, despite considerable efforts over the last 20 years in industry and academia. An alternative approach to inhibiting "undruggable" oncogene targets is to target not the oncogene itself, but the gain-of-function pathways that consist of downstream effectors for oncogene function. A universal output of pathway perturbation is altered gene expression, which can be measured on well-established platforms of microchip arrays[10]. Therefore, identifying pathway signatures that represent oncogene activation in a tumor setting may allow for identification of new targets in otherwise untractable pathways. In the case where a tumor is complex and driven by multiple oncogenic pathways, a whole genome approach will enable identification of all critical genes that are driving tumorigenesis. In this latter case, well defined gene-based pathway signatures can be used to assign both rationally designed mono and combination therapies[11,12].

Pathway Signatures Capture Genetic Output and Epigenetic Features of Tumorigenesis

Although pathway signatures are complementary measures to genetic changes, they are also valuable tools in their own right. Altered gene expression is a universal output of pathway perturbation. Multiple methods for deriving pathway signatures from simultaneous whole genome mRNA measurements have been developed over the last decade, with many pathway signatures published and validated as representative measures of variant biological processes. Perhaps the disease most impacted by this approach is breast cancer, where multiple pathway signatures have been used to redefine the three major subclasses of disease (as defined by FISH status of Her2 and IHC measurement of ER and PR) into four new categories: Her2, luminal A, luminal B, and basal[13]. The latter three categories are defined by pathway signatures, which are currently being used as experimental biomarkers in clinical studies. The hope that these signatures will be used in clinical practice is encouraged by the 2007 Food and Drug Administration approval of an assay for another pathway signature, the prognostic Mammaprint assay[14]. This landmark achievement is evidence that a multi-analyte measurement of gene expression can be executed with rigor and reproducibility to meet clinical regulatory requirements.

Using Signatures to Translate Knowledge of Targets, Preclinical Functional Data, and Clinically Relevant Diseases

Perhaps the greatest value of pathway signatures is the ability to link complex preclinical biology with clinical biology. Although simple cancers can be identified by single analyte biomarkers, either those that are already known or those that remain to be discovered, the more common and complex cases will require multi-analyte tools. The universal nature of microarray platforms has allowed for comprehensive measurement of genome expression in digital formats that enable identification of coordinated gene expression changes associated with different biological states and development of techniques to assign a numerical value to a pathway signature. This powerful approach allows comparison of biological samples from many different sources. The Connectivity Map, a post-array digital analysis tool, uses pathway signatures derived from microarray measurements to explain the mechanism of action for novel drug candidates, and multiple drug response signatures have been derived from comparing pre- and post-treatment samples in preclinical models. These signatures can then be tested from array profiles of tumor samples to identify patients with relevant biology[15].

Another natural feature of measuring mRNA that is distinct from tumor gene mapping is that this approach measures the influence of oncogenes and tumor suppressor genes in specific context and also integrates signaling between the tumor and its microenvironment. Therefore, mRNA measurement has the potential to measure any biology that may override genetic determinants. For example, a pathway signature derived from overexpression of the *K-Ras* oncogene was used to identify tumors that have elevated pathway expression in the absence of the mutant oncogene, suggesting the presence of alternative and yet undiscovered drivers of disease [16]. Furthermore, as an example of measuring dominant signaling from the microenvironment, the mesenchymal subtype in the glioblastoma experiment mentioned earlier[9] was dominated by a pathway signature that had significant overlap with immune regulation and was distinct from one found in normal brain tissue, suggesting the presence of pro-inflammatory cells in this subtype.

The limited value of preclinical models of cancer is evidenced by the high response rate demonstrated in preclinical studies by clinical candidate molecules, followed by a high failure rate in the clinic due to lack of efficacy[17,18]. One way to improve these models is to better represent the heterogeneity of disease by moving a wide variety of tumor samples into preclinical platforms. Derivation of primary tumor samples into novel *in vivo* models is a well-described method of creating new preclinical models of cancer. Successful propagation of primary tumor samples as xenografts in immune-compromised mice varies from 10% to 30%, depending on the tumor type and technique used [19]. If the same tumor samples that are fully characterized in this proposed study can also be propagated *in vivo* or *ex vivo*, functional studies with novel candidate therapies on samples with full genomic and genetic characterization will enable discovery of selection biomarkers to enrich for responder populations for these novel therapies.

Gastric Cancer: An Example of Unmet Medical Need

An example of a cancer that is prevalent in China with unmet medical need is gastric or gastrointestinal cancer. The prevalence of this disease is far greater in Asia than in the West and is the second leading cause of cancer death in the world. Much work has been done to characterize this disease. Currently, gastric cancers are stratified into two histological groups, diffuse-type and intestinal-type adenocarcinoma, and both categories are associated with chronic inflammation by *H. pylori*, a common stomach infection[20]. Multiple studies with small sets of tumor samples reveal the changes in oncogenes that have already been identified from other cancer types, including *K-Ras*, *BRAF* and *PI3K* mutations, *PTEN* deletions, and *MET* and *HER2* amplifications, which have all be observed in some clinical samples. These data, in combination with the identification of point mutations in the *E-cadherin* gene within a family with inherited predisposition for this disease[21], suggest critical roles in growth factor signaling and the Wnt-β-catenin pathways. Interestingly, outer membrane proteins of *H. pylori* have been shown to interact with proteins from both of these pathways. For the majority of this disease, however, no oncogene or pathway driver is known. Treatment and response rates for gastric cancer in China are variable. Standard of care for treatment in the US includes the use of cytotoxic therapies (docetaxel, cisplatin, 5-fluorouracial) and radiation, where 5-year overall survival is less than 25% for all patients and less than 4% for those with advanced stage disease [22]. Therefore, the majority of patients who are diagnosed with mid to late stage gastric cancer carry a disease that has both an unknown etiology and a dismal prognosis. It can be argued, then, that a non-hypothesis driven characterization of gastric cancer samples can lead to the identification of the unknown drivers of the majority of this disease. The unusual role of *H. pylori* in this disease warrants an investigation of the host-microbe relationship. Genome-wide association studies to identify genomic vulnerabilities for this disease have been reported and are ongoing[23], but data sets composed of

significant numbers of fully characterized tumor samples from gastric cancer patients are not yet available.

The Proposal

Cancers in Chinese populations represent an unmet medical need. Initial clinical evaluation of new therapies in cancers that are prevalent in the Western population is the standard in pharmaceutical development and delays demonstration of their possible use in the Asian population. We propose to accelerate the treatment of cancers in Chinese populations by creating a comprehensive characterization of these diseases. The combination of both genotypic and phenotypic profiling should enable 1) identification of all candidate oncogene drivers and tumor suppressor genes by measuring all genetic changes in a single tumor; 2) identification of aberrantly activated pathways that correspond to these genetic changes by using phenotypic whole genome profiling as well as biological measurements and statistical methods to identify causality; and 3) discrimination between driver mutations and passenger mutations to identify new targets and biomarkers for drug discovery. To date, this approach has not be applied to cancers in Chinese populations, as the deployment of both genotypic and phenotypic profiling technologies on a single patient sample is limited by patient sample size, the cost of deploying two expensive technologies on a single patient sample, and unavailability of systematic analytical methods and software that enables analysis of data from two different kinds of data sets. More recently, the development of quantitative RNA sequencing, using second generation nucleic acid sequencing technology, offers the promise of improved dynamic range and fidelity of measurement for mRNA, and enables simultaneous detection of genetic changes and relevant mRNA gene expression levels[24]. Thus, the new development of accurate and quantitative RNA sequencing technology is an attractive alternative for characterizing a biological sample using a single measurement, revealing both the candidate driver genes and activated pathways for this disease.

Most cancer patients do not respond to therapy or eventually develop a cancer that is resistant to current therapies. It is therefore reasonable to assume that these patients have cancers that are complex and are unlikely to be defined by a single analyte biomarker. The application of genomic methods to reveal both DNA sequence and to calculate quantitative digital values for gene expression has the potential to improve our understanding of patient subpopulations with the most common and complex forms of these diseases.

Regardless of the technologies chosen to characterize patient samples, it is clear that an integrated whole genome and genomic approach, combining both DNA and RNA measurements, is critical to discovering the root causes of these complex and fatal disorders. These datasets will naturally attract rational drug discovery efforts, and derivation of new drug targets and clinical biomarkers will enable rapid demonstration of efficacy in a molecularly-defined subpopulation of responders to existing and future therapies. Once completed, these datasets will reside *in silico* and *in perpetuum*, for scientists and oncologists to interpret and use to guide the rational implementation of novel treatments into effective standards of care.

Received: 2011-04-06; accepted: 2011-04-08.

References

[1] Tsao AS, Tang XM, Sabloff B, et al. Clinicopathologic characteristics of the EGFR gene mutation in non-small cell lung cancer [J]. J Thorac Oncol, 2006,1(3):231-239.

[2] Yang L, Parkin DM, Ferlay J, et al. Estimates of cancer incidence in China for 2000 and projections for 2005 [J]. Cancer Epidemiol Biomarkers Prev, 2005,14(1):243-250.

[3] Kantarjian H, Sawyers C, Hochhaus A, et al. Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia [J]. N Engl J Med, 2002,346 (9):645-652.

[4] Slamon DL, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2 [J]. N Engl J Med, 2001,344(11):783-792.

[5] http://nobelprize.org/nobel_prizes/medicine/laureates/1989.

[6] Weinstein IB. Cancer. Addiction to oncogenes—the Achilles heal of cancer [J]. Science, 2002,297(5578):63-64.

[7] King AJ, Patrick DR, Batorsky RS, et al. Demonstration of a genetic therapeutic index for tumors expressing oncogenic BRAF by the kinase inhibitor SB-590885 [J]. Cancer Res, 2006,66(23):11100-11105.

[8] Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers [J]. Science, 2007,318(5853): 1108-1113.

[9] Verhaak RG, Hoadley KA, Purdom E, et al. Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1 [J]. Cancer Cell, 2010,17(1):98-110.

[10] Watters JW, Roberts CJ. Developing gene expression signatures of pathway deregulation in tumors [J]. Mol Cancer Ther, 2006,5(10):2444-2449.

[11] Jia J, Zhu F, Ma X, et al. Mechanisms of drug combinations: interaction and network perspectives [J]. Nat Rev Drug Discov, 2009,8(2):111-128.

[12] Hoeflich KP, O'Brien C, Boyd Z, et al. In vivo antitumor activity of MEK and phosphatidylinositol 3-kinase inhibitors in basal-like

breast cancer models [J]. Clin Cancer Res, 2009,15(14):4649–466.
[13] Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer [J]. N Engl J Med, 2009,360(8):790–800.
[14] van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer [J]. N Engl J Med, 2002,347(25):1999–2009.
[15] Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease [J]. Science, 2006,313(5795):1929–1935.
[16] Loboda A, Nebozhyn M, Klinghoffer R, et al. A gene expression signature of RAS pathway dependence predicts response to PI3K and RAS pathway inhibitors and expands the population of RAS pathway activated tumors [J]. BMC Med Genomics, 2010,3:26.
[17] Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? [J]. Nat Rev Drug Discov, 2004,3(8):711–716.
[18] Kamb A, Wee S, Lengauer C. Why is cancer drug discovery so difficult? [J]. Nat Rev Drug Discov, 2006,6(2):115–120.
[19] Morton CL, Houghton PJ. Establishment of human tumor xenografts in immunodeficient mice [J]. Nat Protoc, 2007,2(2):247–250.
[20] Polk DB, Peek RM Jr. Helicobacter pylori: gastric cancer and beyond [J]. Nat Rev Cancer, 2010,10(6):403–414.
[21] Cohen MM Jr. Molecular dimensions of gastrointestinal tumors: some thoughts for digestion [J]. Am J Med Genet A, 2003,122A(4):303–314.
[22] http://www.cancer.org/cancer/stomachcancer/overviewguide/stomach-cancer-overview-survival-rates.
[23] Study Group of Millennium Genome Project for Cancer, Sakamoto H, Yoshimura K, Saeki N, et al. Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer [J]. Nat Genet, 2008,40(6):730–740.
[24] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics [J]. Nat Rev Genet, 2009,10(1):57–63.