EMBO Molecular Medicine

Genomic biomarkers in predictive medicine. An interim analysis

Keywords: biomarkers; early detection; genomics; personalized medicine; translational research

Current genomics and biotechnology promise the development of biomarkers to predict individual disease risk, enable early detection of disease, and improve diagnostic classification to better inform individualized treatment. I discuss these objectives, commenting on progress to date and obstacles to future success. The discussion mainly uses examples from oncology where the nature of the disease has expedited genomic approaches for developing biomarkers. Many of the lessons being learned in oncology, however, should be applicable to other chronic diseases.

Biomarkers are biological measurements that can be used to predict risk of disease, to enable early detection of disease, to improve treatment selection and to monitor the outcome of therapeutic interventions. One major motivation of the Human Genome Project was the identification and development of such biomarkers for 'personalized, preventive and predictive medicine'. Although the sequencing of the human genome has had profound impacts on biomedical research in many other fields, and while it is still too early to fully assess its impact on biomarker development (Lander, 2011), I will provide an interim analysis and identify some of the roadblocks to progress.

One of the greatest problems in the development and validation of biomarkers is the ambiguity of the term and the failure to recognize that biomarker validity means fitness for intended use. An enormous amount of resources is simply wasted because researchers do not focus clearly on an intended use. This is seen, for example, in the gap between the enormous literature on prognostic biomarkers and the limited use of such markers outside research. Failure to focus also results in misleading claims for early detection biomarkers based on studies with inappropriate controls. Here, I shall discuss separately several broad categories of intended use as illustrated in Fig 1 (summarized in Box 1).

Biomarkers of Disease Risk

The development of genetic biomarkers for predicting risk of disease in individuals has had limited success to date. Numerous large whole-genome association studies (Ioannidis et al, 2010) involving thousands of patients have been conducted for many chronic diseases. These studies have genotyped cases and controls in order to identify germ-line polymorphisms that put individuals at higher risk for developing a specific disease. Many genetic loci have been identified as statistically significant and, in some cases, are providing valuable leads for understanding the biological basis of the diseases. However, the strength of the associations is often far too weak to provide much value for counselling individuals (Bloss et al, 2011). Ioannidis et al reviewed 56 GWAS reporting 92 statistically significant associations between cancer phenotypes and genetic variants and found a median per-allele odds ratio (OR) of 1.22 with an interquartile range of 1.15–1.36. The absolute risk of disease for a subject with a high-risk allele can be considered the 'positive predictive value' of the genetic test and can be expressed $PPV = RR^{*}\pi/(1+\gamma^{*}(RR-1))$, where π denotes the prevalence of the disease, γ denotes the prevalence of the high-risk allele and RR denotes the relative risk of disease for a subject with high-risk *versus* standard risk allele. Most diseases have sufficiently low prevalences that the OR and RR are essentially equal. From this formula one can show that the PPV is no greater than $RR^{*}\pi$. So if the RR is 1.22 and the disease prevalence is 5% ($\pi = 0.05$), then the absolute risk of developing the disease for a subject with a high-risk allele is no greater than 6.1%. If the RR were 5, then the absolute risk of developing disease for a subject with a high-risk allele could be as large as 10% for a disease with 5% prevalence in the population.

There are several possible explanations for the 'low penetrance' (low PPV) of the identified genetic loci. For oncology, a major reason is the genetic heterogeneity of most cancers. For example, estrogen receptor negative and estrogen receptor positive breast cancer are different in terms of the somatic mutations that characterize them, as well as with regard to natural history and responsiveness to treatment. From most perspectives they appear to be different diseases and lumping them together in searching for polymorphisms of disease susceptibility is problematic (Kraft and Halman, 2010) Indeed, the success of early genetic linkage studies that identified the highly penetrant BRCA1 locus owed, in large part, to restricting the studies to cases with early onset breast cancer. Many other chronic diseases are phenotypically and molecularly heterogeneous. They are also probably genetically heterogeneous and thus very difficult to study with broad genome-wide association studies.

BOX 1: Progress in genomic biomarker development

Biomarkers are biological measurements that can be used for a variety of purposes, including identifying individuals who are at high risk of developing a disease, detecting disease early at a stage when it is treatable and diagnostic classification for personalized treatment based on a biological characterization of the disease of each individual patient. The sequencing of the human genome has provided an important body of information for the development of biomarkers for all of the purposes mentioned. In this paper, I provide a short and personal assessment of the progress achieved in these areas of genomic biomarker development.

It is indicated that (i) progress has been slow in personalized risk prediction and in early detection; (ii) genome-wide association studies are more likely to provide leads for understanding the pathogenesis of diseases than useful information on personalized risk assessment; and (iii) development of biomarkers sufficiently sensitive and specific for early detection of diseases that will be life-threatening is very challenging and the validation of such biomarkers requires very large randomized screening trials. The development of biomarkers for personalizing treatment selection, particularly in oncology, has seen greater progress. Key bottlenecks that limit progress in the translation of discoveries in genomics to biomarkers and treatments that reduce mortality and morbidity from chronic diseases are also discussed.

Other possible explanations for the failure of finding highly penetrant susceptibility genes is the fact that chronic diseases are caused by the combined effects of multiple genetic polymorphisms and/or that chronic diseases are caused by a combination of genetic and environmental causes. The greatest potential value of genome-wide association studies is to shed light on the biological basis of the disease. For oncology, this may be less essential since cancers are in large part diseases of DNA modification and the tumour genomes can be directly evaluated. Of course interpreting tumour genomes to find the mutations, which are key to oncogenesis, is difficult (Ledford, 2010) and GWAS could provide additional useful information. Most cancers, however, result from complex sequences of somatic mutations, which interact with each other to influence tumour evolution (Ashworth et al, 2011). It seems unrealistic to expect that individual polymorphisms will have substantial explanatory or predictive power in elucidating these relationships. For other chronic diseases, however, GWAS may be more essential for generating leads concerning the biology of the disease. In most cases, however, these leads must be followed by fine mapping of the regions of the detected polymorphisms and then years of biological investigations to understand the relevance of the disease alleles. It is probably too early to evaluate the impact of GWAS on medical utility. It is clear, however, that the initial expectations of easy and direct translation of GWAS findings to patient benefit were unrealistic. GWAS studies in heterogeneous diseases such as cancer could be improved by evaluating associations with biologically meaningful subsets of patients with central review of cases to ensure accuracy of classification.

Early Detection Biomarkers

Many diseases can be more effectively treated at an early stage. Most solid tumours have a long sub-clinical course prior to diagnosis and hence, there should be substantial opportunity for early detection. There has been little success to date, however, in developing and validating early detection biomarkers with medical utility. Early detection research has been severely hindered by the use of poor study methodology. Many cancer biomarkers are 'discovered' by comparing levels of candidate proteins in tumour tissue, collected at diagnosis, to normal tissue. Numerous 'discoveries' have been published and publicized based on such evidence. Finding such a difference, however, is very weak evidence that the marker will be useful for early detection. A recent publication evaluated 28 candidate biomarkers using serum samples obtained from subjects in

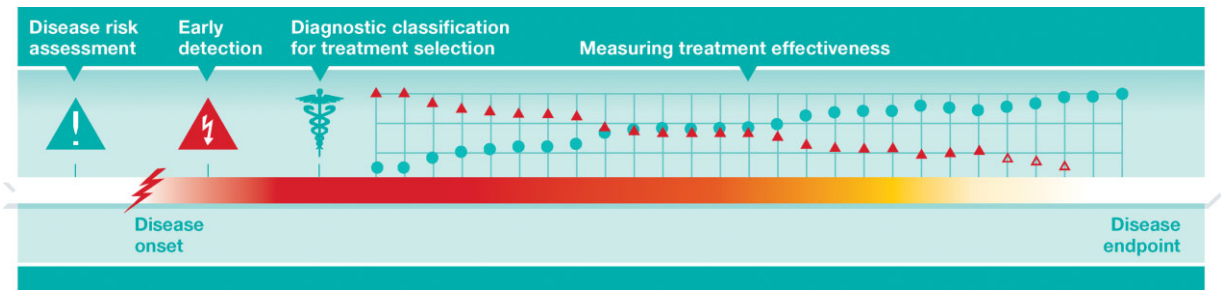

Figure 1. Broad categories of intended use of biomarkers.

a randomized screening trial. None of the 28 showed early detection performance either individually or in combination as good as or better than the traditional CA-125 (Mai et al, 2011). This prompted a commentary in Nature titled 'Missing the Mark'. (Bucher, 2011) One way to improve early detection research would be to perform genome-wide or proteome-wide discovery using serum samples archived from retrospective longitudinal cohorts rather than samples from patients at diagnosis. Such 'phase 3' studies have been generally reserved for validation of candidate markers discovered in the quick and easy phases 1 and 2 studies based on diagnostic samples.

If one of the candidates in the ovarian cancer study had been found elevated in serum samples prior to diagnosis based on the retrospective analysis of the screening trials, this would have given an indication of the lead-time achievable with that marker, but because the study was retrospective it would not provide information about whether such detection would reduce mortality or morbidity from the disease. One would not know whether the cases detected had localized disease at the time of detection. In general, although perhaps not the case for ovarian cancer, without a prospective randomized screening trial one would not know what proportion of the detected cases might not represent a tumour that would be life threatening within the lifetime of the patient (Etzioni et al, 2003).

The identification of early detection markers that are sufficiently specific for use in population screening is challenging. Some important aspects of the statistics of early detection screening is shown in Box 2. If the prevalence of the disease is denoted by π and the sensitivity and specificity are denoted by sens and spec, respectively, then the probability that a test positive case has the disease (positive predictive value) equals $\pi\ (\text{sens})/(\pi\ (\text{sens}) + (1-\pi)(1-\text{spec}))$. When sensitivity and specificity are both 0.95 and the prevalence of disease in the population is 1%, the positive predictive value is only about 0.16 – only 16% of the test positive individuals will actually have the disease. The remaining 84% will be possibly subjected to unnecessary and invasive follow-up procedures. In order to conduct effective population screening, we need a test with very high specificity and to restrict screening to high-risk populations.

There is also the issue of what do we mean by 'disease'. We not only need very high specificity, we need very high specificity for detecting the form of the disease which is life threatening. But, the earlier the point at which the disease is detected, the more difficult it may be to distinguish a life-threatening cancer from an nodule which may be indolent during the patient's lifetime, given that the early steps of oncogenesis are variable and stochastic. For example, a large proportion of individuals have a BCR-ABL fusion protein detectable in their blood, yet only a small percentage of those develop chronic myelogenous leukaemia (CML). Not all early lesions may progress to invasive cancer and yet early detection may encourage treatment with serious adverse effects.

In addition to showing the PPV resulting from values of prevalence, sensitivity and specificity, Box 2 also shows the expected number of lives saved by screening and the expected number of false positives for every 10,000 individuals screened. The expected number of lives saved is the product of the prevalence of lethal tumours in the population, the sensitivity of the test and the improvement in probability of cure by detecting disease T years before clinical diagnosis. The table uses hypothetical improvements in cure probabilities of 0.10 and 0.20 corresponding to lead times of 1 or 5 years. It also assumes that a longer lead time results in slightly lower sensitivity and specificity of the test. Using the assumptions shown in the table, a screening test that provides a 1-year lead-time for disease detection results in 9.5 lives saved and 495 false positives per 10,000 screened. With a test providing a 5-year lead-time, 18 lives are saved and there are 990 false positive tests per 10,000 patients screened. These numbers depend on the assumed 10 and 20% increases in cure rate with 1- and 5-year lead-times, respectively, and on the assumed 1% disease prevalence.

Progress in the identification of early detection biomarkers has been very limited. Most of the focus has been on funding the application of new technologies to identification of tumour markers. Each new round of technology development has generated overly optimistic claims based on poor research design and methodology. In order to expedite progress and to make more efficient use of limited resources, new strategies for biomarker discovery are required that make greater use of archived samples from longitudinal studies (Zhu et al, 2011). Improved policies for funding

BOX 2: Statistics of early detection

Sens = probability marker is positive at screening T years before diagnosis of lethal tumour.

Spec = probability marker is negative at screening in absence of lethal tumour.

π = prevalence of lethal tumours in population (0.01 means 1/100 patients screened).

ΔCure = improvement in cure proportion of lethal tumours detected T years before diagnosis.

PPV = probability of lethal tumour given marker is positive at screening

$$\text{PPV} = \frac{\pi\text{sens}}{\pi\text{sens} + (1-\pi)(1-\text{spec})}$$

NSaved = increase in lives saved per 10,000 individuals screened.

NSaved = 10,000 π sens ΔCure.

NFalse+ = number of false positive tests per 10,000 individuals screened.

NFalse+ = 10,000 $(1-\pi)$ $(1-\text{spec})$.

π	Sens	Spec	T	ΔCure	PPV	NSaved	NFlase+
0.01	0.95	0.95	1	0.10	**0.16**	**9.5**	**495**
0.01	0.90	0.90	5	0.20	**0.087**	**18**	**990**

early detection research are required that place greater emphasis on proper research design as well as use of state-of-the-art technology. Even with a candidate marker that has been demonstrated to enable earlier disease detection, however, demonstrating that it has medical utility for reducing mortality from the disease requires very large and expensive randomized screening trials.

Treatment Selection Biomarkers

In some areas of therapeutics, such as oncology, predictive and prognostic biomarkers have been effectively developed to help guide treatment decisions. For example, the OncotypeDx recurrence score and MammaPrint signature are used to determine whether a woman with node-negative hormone-receptor-positive breast cancer has a sufficiently good prognosis with local treatment and adjuvant hormonal treatment that she does not require cytotoxic chemotherapy (Paik et al, 2004; Van de Vijver et al, 2002). In oncology, there is an enormous literature of claims for improved prognostic factors that have never found clinical application. This gap between research and application has accelerated with the development of gene expression profiling. Identifying a prognostic impact of a gene mutation may suggest an important role of that gene product as a molecular target as was the case for HER2 in breast cancer. But the numerous gene expression-based signatures that have been developed based on prognosis or on clustering expression profiles (Perou et al, 2000) have had limited utility for either elucidating underlying biology or informing treatment decision-making.

Developing a prognostic gene expression signature is not likely to be useful unless the signature is developed with an intended use clearly in mind from the start. That intended use should drive the selection of cases and the interpretation of results. In order to identify a gene expression signature like the Oncotype DX recurrence score for use in determining which breast cancer patients with node-negative estrogen receptor positive disease have such good prognosis that they do not require chemotherapy, one needs to focus the study on node negative, hormone receptor positive patients who received hormonal treatment but no chemotherapy. Most prognostic studies use a convenience sample of heterogeneous patients and develop signatures that have no therapeutic relevance (Subramanian & Simon, 2010). The objectives of genomic prognostic studies are generally not clearly considered. The purpose of prognostic signatures like the Oncotype Dx recurrence score and the MammaPrint signature is to help patients and physicians in making informed therapeutic decisions. Useful tools like Oncotype Dx and MammaPrint should not be criticized because they do not provide biological insight into the disease.

'Predictive biomarkers' indicate which patients are most likely or unlikely to benefit from a specific treatment. For example, estrogen receptor expression levels have been used for many years to select patients for anti-estrogen hormonal treatment and HER2 over-expression or amplification is widely used to select patients for treatment with anti-HER2 drugs. EGFR mutation is used to select patients for small-molecule EGFR inhibitors in non-small-cell lung cancer and KRAS mutation to de-select patients from therapy with anti-EGFR antibodies in advanced colorectal cancer. Discovery of the BCR-ABL fusion protein in CML led to the development of imatinib and the analysis of mutations of that gene determines second-line treatments (Drucker et al, 2001). The identification of the EML4-ALK fusion gene in a small subset of patients with non-small-cell lung cancer led to using a kinase inhibitor that targets that gene with extremely promising results (Kwak et al, 2010). Similarly, the discovery of a single point mutation in the BRAF gene in 60% of patients with metastatic melanoma led to the development of an inhibitor with increased specificity for the mutated protein with extremely promising results (Flaherty et al, 2010).

In fact, most oncology drug development today is driven by molecular targets because cancers of most primary sites are heterogeneous with regard to oncogenesis and sensitivity to treatment. Consequently, the blockbuster strategy is unlikely to work for most cancers; testing new drugs in broad patient populations with molecularly uncharacterized tumours is no longer based on solid science and is unlikely to be successful. Development of cancer therapeutics with companion diagnostics is the dominant theme today. The predictive biomarkers that are used to guide treatment selection for molecularly targeted drugs are mostly based on the mutation or amplification of a single gene. The genes are usually either the target of the drug or a non-target gene with relevance to the pathways involving the target gene. Protein over-expression has tended to be a less reliable basis for credentialing gene targets or for development of predictive biomarkers. Some of the drugs being developed, such as kinase inhibitors, have multiple targets and development of a predictive biomarker is made difficult by uncertainty in the molecular basis for anti-tumour effect. The extreme clonal heterogeneity of most tumours is also a challenge for proper evaluation of predictive biomarkers in oncology. In the future, patients at major cancer centres will likely undergo whole exome sequencing of multiple samples from each accessible tumour site in order to develop an optimal therapeutic strategy.

Gene expression signatures have less frequently been used as predictive biomarkers for new drugs. Use of a genomic mutation or amplification of a gene related to the mechanism of action of the drug as a predictive biomarker is more scientifically satisfying. Gene expression profiling has been frequently used for developing prognostic signatures but much less frequently for developing predictive markers of benefit from specific treatments. This may be in part because frozen tissue samples have rarely been archived from patients in clinical trials suitable for the development of predictive signatures. When archived tissues are available from randomized trials, the tissues are generally formalin fixed and paraffin preserved (FFPP). Because of RNA degradation in FFPP tissues, such samples, until recently, were not suitable for microarray gene expression analysis. Development of prognostic signatures for a heterogeneous set of samples, which may be derived from different types of progenitor cells, also is easier than developing a predictive signature for a set of

patients homogeneous enough to have been included in a single clinical trial.

Unfortunately, progress in development of new therapeutics in oncology has primarily been restricted to monoclonal antibodies and small-molecule kinase inhibitors and improvements in therapeutics have only infrequently resulted in cures. Many of the most important molecular targets, such as the tumour suppressor genes p53 and Rb, are not effectively amenable to drug intervention. Using synthetic lethality approaches to target the effects of key mutations in such genes holds great promise and will likely be the dominant theme for future drug development in oncology (Ashworth et al, 2011; Haber et al, 2011).

Personalization of therapy is only effective if the therapeutic strategies for the identified subsets of patients are effective. For instance, the point mutation that causes sickle-cell anaemia was identified more than 60 years ago, but that discovery has not yet led to effective treatment for that disease (Pauling et al, 1949). The development of predictive biomarkers for guiding treatment for other diseases has in many cases lagged behind developments in oncology, which has the advantage that it is a disease of DNA and hence much information about tumour sensitivity to treatment options can be gained by using the plethora of new whole-genome technologies. Yet, there is substantial research in using high-throughput genomic and proteomic technologies to identify biomarkers for many other diseases.

Bridging the Gap between Basic Genomic Research and Patient Benefit

The gap between basic research and clinical benefit has been termed the 'valley of death' in the popular press (Begley, 2008; Butler, 2008). Much attention has been devoted to the numerous infrastructure and financial complexities of translational research including regulatory issues, human subject approvals, intellectual property issues, lack of funding, lack of patients, lack of training for physician-investigators and a fragmented research infrastructure (Hawk et al, 2008; Nathan and Varmus, 2000; Sung et al, 2003). The key scientific and structural roadblocks have received less attention, however, and they influence our ability to use genomic technologies for developing useful biomarkers as well as for developing effective therapeutics. Some of these roadblocks are discussed below and some suggested approaches for improving translational research are depicted in Fig 2.

First, basic research does not go far enough in identifying the key steps in the development and pathogenesis of most chronic diseases in order to enable translational research to proceed effectively. Even in oncology, our very limited understanding of the oncogenesis of cancer is a major hurdle to effective translational research (Simon, 2010). Once basic research identifies a key step of oncogenesis and a druggable molecular target, the pharmaceutical and biotechnology industries are often adept at developing potent inhibitors of that target.

We still do not fully understand the development and progression of any type of cancer even if, in rare cases such as CML, our knowledge of oncogenesis has been sufficient to develop effective treatments. Development of more effective treatments likely requires the characterization of key founder mutations that drive the pathogenesis of the individual tumour, understanding the networks in which these genes are involved and treating early enough with combinations of drugs to overcome resistant subclones. Deep single molecule sequencing of multiple samples from individual tumours will enable us to characterize the clonal heterogeneity of each tumour (Jones et al, 2008; Navin et al, 2011). With sufficient sequencing power, we can phylogenetically reconstruct the evolution of individual tumours and identify the founder mutations (Campbell et al, 2008). These founder mutations represent the initial rate-limiting genomic changes that enabled the developing tumour to grow to a size in which numerous subsequent mutations could develop in a non rate-limiting manner (Simon, 2010). Because these founder mutations are present in all sub-clones and because subsequent mutations developed in the context of these mutations, they may represent the key molecular targets for that individual tumour. Even with identification of tumour specific founder mutations, it may be necessary to treat early and with combinations of drugs selected based on knowledge of the networks in which the founder genes participate. Imatinib is highly effective in treating CML if treatment begins prior to transition towards blast crisis. The blast crisis of CML may represent a mutational meltdown that also occurs in solid tumours. In CML we have the benefit of detecting the disease before that meltdown occurs. The oncogene addiction to founder mutations that tumours sometimes exhibit can be dissipated by later mutations (Jonkers & Berns, 2004; Weinstein, 2002). Ashworth et al (Ashworth et al, 2011) provide a penetrating discussion of the possible

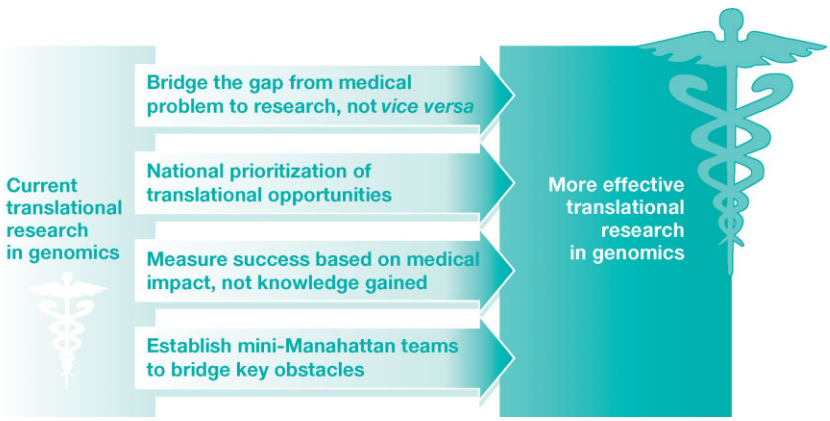

Figure 2. Suggested approaches to improve translational research.

basis of oncogene addiction and other kinds of gene interactions, which may be therapeutically exploitable but elucidation of such interactions is at an early stage of development.

For many diseases the challenges in understanding disease pathogenesis are even greater than for cancer. Germline polymorphisms may provide leads, but the process of elucidating the biology of the disease to find key molecular targets for treatment is often painstakingly slow.

A second area of scientific bottleneck is related to the fact that there is little focus outside of industry on identifying key breakthroughs in basic research and funding prioritized programs to translate those breakthroughs into products that benefit patients. Much of the public funding for translational research is devoted to making biological measurements on patient tissues in an attempt to understand the nature of the disease or providing infrastructure to help investigators bring their personal research to the clinic. After the V600E point mutation in the BRAF gene was discovered to be present in about 60% of patients with malignant melanoma, there was little drug discovery activity among NIH grantees to exploit this finding. Fortunately, BRAF was druggable with standard chemistry and two companies developed extremely promising specific inhibitors of the mutated form of the gene. When the translational challenges are more difficult, or when the financial incentives are either too limited or in conflict with industry concerns about market segmentation, however, such complete and uncoordinated dependence on industry for therapeutic development may not server the public well.

For example, mutations of the p53 and Rb tumour suppressor genes are prevalent and important in many types of cancer but their gene products are not easily druggable; neither industry nor academic research have developed successful approaches for exploiting these mutations. Developing feasible pharmacologic ways of interfering with mutated p53 or Rb in tumours are difficult, long-term, high-risk endeavours that are not adequately addressed either by industry or by the culture of the NIH investigator-initiated grant system. It is not just that these problems are scientifically difficult, it is that existing mechanisms for supporting research and most existing research organizations do not provide an effective framework for a concerted effort to tackle these problems. Consequently, the roadblocks remain, in some cases for decades as in the point mutation causing sickle-cell anaemia (Pauling et al, 1949).

Bridging the broad gaps between basic research and clinical benefit is likely to require major changes in the interactions between industry and academia, and more public funding of 'mini-Manhattan' project teams to overcome key roadblocks. The investigator-initiated framework is highly effective for basic research and has yielded major biomedical discoveries. It has not, however, elucidated the basic steps in the development and pathogenesis of many major chronic diseases nor has it provided adequate identification of key targets to enable effective translational research. It is not necessarily the most appropriate framework for bridging the 'valley of death'. The US National Science Foundation and Defense Advance Research Projects Agency have utilized strategies involving investigator-initiated approaches to publicly prioritized objectives in some of their programs.

To tackle the scientific bottlenecks to key translational opportunities, new horizontally integrated organizations of experts in basic, clinical and quantitative research are needed. Such organizations can play an important role in training the next generation of biomedical researchers to work in settings without silos or hierarchies of disciplines where creative basic scientists, clinicians and quantitative scientists can be part of a single team working together daily to focus on bridging the gap separating basic research from products that benefit patients.

The scientific challenges of understanding the pathogenesis of chronic diseases to the extent that we can effectively prevent, detect, diagnose and treat them are substantial. Nonetheless, there is an enormous amount of talent available to meet these challenges. To take advantage of the opportunities provided by the genomic, biotechnology and information revolutions, however, we need to better focus this talent on overcoming the key bottlenecks to progress. We need to treat the biomedical research enterprise as a system that needs to be optimized to achieve our objectives. This may require creating new kinds of organizations to better foster innovation, encourage transdisciplinary fertilization and ensure that resources are optimally allocated for overcoming the key obstacles to progress.

» The scientific challenges of understanding the pathogenesis of chronic diseases to the extent that we can effectively prevent, detect, diagnose and treat them are substantial. Nonetheless, there is an enormous amount of talent available to meet these challenges. «

Acknowledgements

The author is appreciative of the valuable comments of the referees.

The author declares that he has no conflict of interest.

Richard Simon

Richard Simon is Chief of the Biometric Research Branch in the US National Cancer Institute and head of the section on Molecular Statistics and Bioinformatics. This article represents his personal opinion on the current status and key roadblocks to progress in developing genomic biomarkers to reduce mortality and morbidity from cancer and other chronic diseases.

Biometric Research Branch, National Cancer Institute, Bethesda, MD, USA
E-mail: rsimon@nih.gov

References

Ashworth A, Lord CJ, Reis-Filho JS (2011) Genetic interactions in cancer progression and treatment. Cell 145: 30-38

Begley S (2008) We fought cancer . . . and cancer won. Newsweek, Sep 15, 43-66

Bloss CS, Schork NJ, Topol EJ (2011) Effect of direct-to-consumer genomewide profiling to assess disease risk. N Engl J Med 364: 524-534

Bucher L (2011) Missing the mark: Why is it so hard to find a test to predict cancer? Nature 471: 428-432

Bucher L (2011) Missing the mark: Why is it so hard to find a test to predict cancer? Nature 471: 428-432

Butler D (2008) Crossing the valley of death. Nature 453: 840-842

Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I, Follows GA, Green AR, Futreal PA, Stratton MR (2008) Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proc Natl Acad Sci 105: 13081-13086

Drucker B, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM, Lydon NB, Kantarjian H, Capdeville R, Ohno-Jones S, *et al* (2001) Efficacy and safety of a specific inhibitor of the BCRABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med 344: 1031-1037

Etzioni R, Urban N, Ramsey S, Mcintosh M, Schwartz S, Reid B, Radich J, Anderson G, Hartwell L (2003) The case for early detection. Nat Rev Cancer 3: 1-10

Flaherty KT, Puzanov I, Kim KB, Ribas A, Mcarthur GA, Sosman JA, O'dwyer PJ, Lee RJ, Grippo JF, Nolop K, *et al* (2010) Inhibition of mutated, activated BRAF in metastatic melanoma. N Engl J Med 363: 809-819

Haber DA, Gray NS, Baselga J (2011) The evolving war on cancer. Cell 145: 19-24

Hawk ET, Matrisian LM, Nelson WG, Dorfman GS, Stevens L, Kwok J, Viner J, Hautala J, Grad O (2008) The translational research working group developmental pathways: Introduction and overview. Clin Cancer Res 14: 5664-5671

Ioannidis JPA, Castaldi P, Evangelou E (2010) A compendium of genome-wide associations for cancer: critical synopsis and reappraisal. J Natl Cancer Inst 102: 846-858

Jones S, Chen WD, Parmigiani G, Diehl F, Beerenwinkel N, Antal T, Traulsen A, Nowak MA, Siegel C, Velculescu VE, *et al* (2008) Comparative lesion sequencing provides insights into tumor evolution. Proc Natl Acad Sci 105: 4283-4288

Jonkers J, Berns A (2004) Oncogene addiction. Cancer Cell 6: 535-538

Kraft P, Halman CA (2010) GWAS identifies a common breast cancer risk allele among BRCA1 carriers. Nat Genet 42: 819-820

Kwak EL, Bang YJ, Camidge R, Shaw AT, Solomon B, Maki RG, Ou SHI, Dezube BJ, Janne PA, Costa DB, *et al* (2010) Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med 363: 693-703

Lander ES (2011) Initial impact of the sequencing of the human genome. Nature 470: 187-213

Ledford H (2010) The cancer genome challenge. Nature 464: 972-974

Mai PL, Wentzensen N, Greene MH (2011) Challenges related to developing serum-based biomarkers for early ovarian cancer detection. Cancer Prev Res 4: 303-306

Nathan DG, Varmus HE (2000) The National Institutes of Health and clinical research: a progress report. Nat Med 6: 1201-1204

Navin N, Kendall J, Troge J, Andrews P, Rodgers L, mcindoo J, Cook K, Stepansky A, Levy D, Esposito D, *et al* (2011) Tumor evolution inferred by single-cell sequencing. Nature 472: 90-94

Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, *et al* (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351: 2817-2826

Pauling L, Itano HA, Singer SJ, Wells IC (1949) Sickle cell anemia, a molecular disease. Science 25: 543-548

Perou CM, Sorlie T, Eisen MB, van De Rijn M, Jeffrey SS, Reese CA, Pollak JR, Ross DT, Jonsen H, Akslen LA, *et al* (2000) Molecular portraits of human breast tumors. Nature 406: 747-752

Simon R (2010) Translational research in oncology: key bottlenecks and new paradigms. Expert Rev Mol Med 12: e32

Subramanian J, Simon R (2010) Gene expression based prognostic signatures in lung cancer: Ready for clinical use? J Natl Cancer Inst 102: 464-474

Sung NS, Crowley WF, Genel M, Salber P, Sandy L, Herwood LM, Johnson SB, Catanese V, Tilson H, Getz K (2003) Central challenges facing the national clinical research enterprise. J Am Med Assoc 289: 1278-1287

Van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AAM, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, *et al* (2002) A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347: 1999-2009

Weinstein IB (2002) Addiction to oncogenes—the Achilles heel of cancer. Science 297: 63-64

Zhu CS, Pinsky PF, Cramer DW, Ransohoff DF, Hartge P, Pfeiffer RF, Urban N, Mor G, Bast RC, Moor LE, *et al* (2011) A framework for evaluating biomarkers for early detction: validation of biomarker panels for ovarian cancer. Cancer Prev Res 4: 375-383

DOI 10.1002/emmm.201100153

