Oncogenomics and the development of new cancer therapies

Robert L. Strausberg[1], Andrew J. G. Simpson[2], Lloyd J. Old[2] & Gregory J. Riggins[3]

[1]*Department of Mammalian Genomics, The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 2085, USA (e-mail: RLS@tigr.org)*
[2]*Ludwig Institute for Cancer Research, 605 Third Avenue, New York, New York 10158, USA*
[3]*Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, Maryland 21224, USA*

Scientists have sequenced the human genome and identified most of its genes. Now it is time to use these genomic data, and the high-throughput technology developed to generate them, to tackle major health problems such as cancer. To accelerate our understanding of this disease and to produce targeted therapies, further basic mutational and functional genomic information is required. A systematic and coordinated approach, with the results freely available, should speed up progress. This will best be accomplished through an international academic and pharmaceutical oncogenomics initiative.

Cancer can result from a multitude of genetic alterations. These range from point mutations to insertions and deletions to chromosomal translocations in the tumour cell's DNA[1–3]. The completion of the human genome sequence[4,5], rapid improvements in our understanding of the molecular basis of cancer, and the introduction of new technologies for assessing genomic, transcriptional and proteomic changes provide new opportunities for addressing the practical challenges of cancer. In particular, prospects for the development of new therapeutics appear to be excellent.

Important cancer-causing mutations, genomic rearrangements and transcriptional changes could be fully catalogued during the next decade, judging by the success of recent oncogenomics initiatives. Coordinated international research efforts analogous to the Human Genome Project (HGP) have the potential to produce an integrated database of cancer-causing mutations, and of the subsequent transcriptional and translational alterations. Such a database could directly identify targets for therapeutic agents.

The clinical responses seen with the first generation of targeted therapeutic agents, such as the kinase inhibitor Gleevec[6] and the monoclonal antibodies Rituxan[7], Herceptin[8] and Avastin[9], have raised expectations that in future there will be an arsenal of targeted cancer therapies of high efficacy and specificity at our disposal[10,11]. Before us is an extraordinary opportunity to contribute to the development of new therapies, but how best to direct and coordinate our efforts is a question we need to investigate further and is the subject of this review.

The importance of genomic technology

The use of automated sequencing technology and the completion of the human genome have accelerated the identification of mutations and alterations in cancer genomes that might be 'targetable' by various therapies. The most easily 'banked' data — and the most definitive of molecular biology data forms — are DNA sequence information. But significant challenges remain: finding small alterations (often just a difference of a single base pair) in the genome is extremely difficult, as is making such studies comprehensive across the many types of cancer and ensuring that large amounts of data are reported in a meaningful fashion. Nevertheless, we are optimistic that a comprehensive mutational database is possible for most human cancers. Many parts of the genome are now being resequenced to identify somatic alterations in exons, and several important cancer mutations have been identified using this strategy[12–14]. Furthermore, advances in DNA sequencing technologies may further reduce the cost and increase the efficiency of obtaining cancer genome sequences.

The availability of the human genome sequence has made alterations such as chromosomal aberrations, translocations, deletions, amplifications and methylation easier to locate[2,15–20]. The new sequence-based approaches, such as digital karyotyping[21,22], a method for detecting DNA copy number on a genomic scale, allow sequence amplifications and deletions to be discovered in a high-throughput manner. And the recently introduced bacterial artificial chromosome (BAC)-end sequencing[23] identifies chromosomal breakpoints in a comprehensive, high-resolution manner. The challenge is to gain the best from each technology so that an efficient and comprehensive analysis of each cancer can be achieved.

Although much effort is focused on individual cancer types, the sequence-based nature of many approaches, with the genome as a point of reference, allows ready database compilation. Databases have already been established for mutations in individual genes such as p53 (ref. 24). Such initiatives could eventually serve as a focal point or model for coordinated community-wide compilations of cancer-related mutations[12,13,25] and the prospect of a more comprehensive approach.

It is not just genomic technology that has improved — methods for detecting messenger RNA have too. Microarrays make major contributions to this because they provide a cost-effective means of assessing and comparing mRNA levels in multiple samples[26–29]. Indeed, studies using this technology have suggested that transcription profiles allow the molecular classification of cancers, as well as insight into the biology of tumour progression[30–32]. This raises the possibility of novel approaches for diagnosing cancers and for predicting response to therapy. In addition, sequence-based approaches for gene tagging — for example, serial analysis of gene expression (SAGE)[33–35], massively parallel signature sequencing[36] and expressed sequence tags (including ORESTES)[37] — provide data sets that facilitate and complement the microarray approaches.

The large-scale transcript sequencing projects, such as those of the United States and Brazil, have resulted in the establishment of and presentation of data on the Cancer

Genome Anatomy Project website (http://cgap.nci.nih.gov), a key reference for the definition of human gene expression in normal tissues and tumours[38]. Although sequence-based approaches such as SAGE are currently less well suited to the analysis of large numbers of samples, sequencing facilitates the discovery of new genes, thereby serving as a platform for the design of microarrays. In addition, the precise digital nature of DNA sequencing is particularly well suited to transcript quantification and the development of community-wide databases[34,38].

The continued improvements in DNA sequencing have increased the cost-effectiveness of transcript tagging, and have made it easier to obtain more information about rarer transcripts. In addition, the use of longer SAGE tags has improved the linkage of tags to specific genes and within the genome[39]. Although still expensive, full-length complementary DNA sequencing remains the 'gold standard' for both gene identification and the definition of gene structure[40,41]. Full-length cDNA sequences will be crucial for the eventual comprehensive documentation of alternative splice forms[42–44] in cancer, and for the discovery of their biological roles in the initiation and progression of the disease. Although efforts have begun[42], these transcripts have not been comprehensively catalogued in normal cells and cancer cells, and a complete picture has yet to emerge. Increasing emphasis on full-length cDNA sequencing promises to greatly increase our knowledge of splice variants in cancer and offers the potential to identify new cancer-specific targets.

An important goal for oncogenomics is the integration of complete genome analyses and transcriptome analyses so that accurate models of the molecular basis of cancer can be built. This step will depend first and foremost on the further decrease in cost and increase in scale of the analytical technologies. But it will also depend on how well these activities are organized: the new databases and analytical tools must be aimed at providing integrated views of the changes in cancer within cells and across diverse cancers.

The need for enhanced informatics tools

Database development and analysis tools were vital for the HGP. They will be equally important for the development and population of a database of the molecular biology of cancer. The goal is to incorporate various types of genomic data (such as mutations, transcriptional changes and cytogenetic events) in a common database with standard terms (for example, using a common scheme for tumour classification). Already, cancer genome projects have been established that provide a starting point for this[12,34,38,45]. Given the existing template provided by the human genome sequence, it is feasible to display diverse data sets (especially those with sequence-based information) in a common framework that is accessible to all. To provide a complete view of cancer, it is important that the database contains not only information about molecular changes in cancers, but also data documenting the absence of changes in cancer, such as resequenced genes in cell lines or tumours that are found to be wild type. This will provide a more complete view of tumours and will help to avoid duplication of resequencing efforts. Similar to the HGP, the cancer genome project would need a set of standards for the data.

Discovering cancer targets

Central to expectations of accelerated target discovery is the perception that genome, transcriptome and proteome analyses will lead to the identification of molecules against which cancer therapeutics might be targeted (Fig. 1). Among many possible approaches, the three that are currently receiving the most attention are: (1) small-molecule inhibitors of oncogenic signals, (2) antibodies to surface components and intercellular communicating factors, and (3) molecularly defined vaccines.

For the first of these options, obvious targets are oncogenes, such as kinases[46,47], because of their causal involvement in tumorigenesis. The development of targeted chemical and biological agents that reduce their oncogenic function has been successful[48,49]. Oncogenes can be putatively identified by the amplification of the gene that encodes them, the over-representation of their transcript or protein in tumours, chromosomal translocation, and/or the cancer-related occurrence of activating missense mutations in their gene sequences[2]. The best example of a targeted, small-molecule oncogene inhibitor that interrupts a crucial signalling pathway is imatinib (better known by its commercial name, Gleevec), which is already in clinical use.

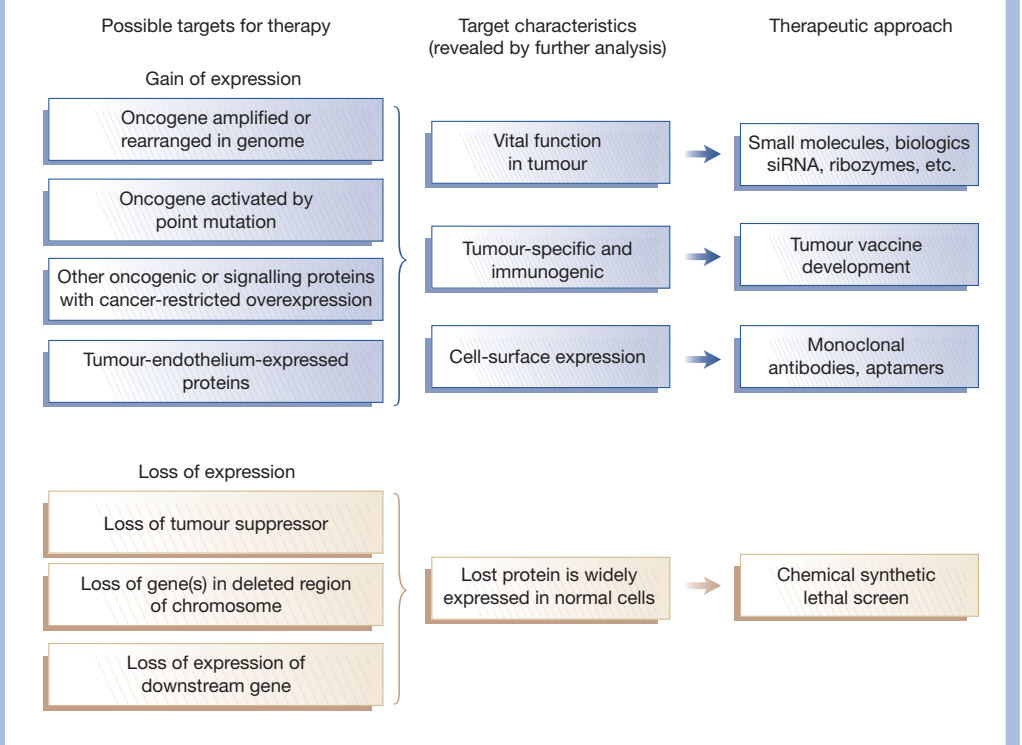

Figure 1 Types of target identified through comprehensive genome and transcriptome analysis of cancer cells. Most genetic targets will result from a 'gain of expression', as in the case of genes that are activated by either a point mutation or a chromosomal rearrangement that creates a new protein sequence. Targets can also be proteins with a normal sequence that are pathologically overexpressed by genomic amplifications, promoter mutations or upstream pathway activation. An example is the discovery of epidermal growth factor receptor (*EGFR*) genomic amplification, leading to its overexpression and increased kinase activity, which can be targeted by small molecules. Designing compounds that are toxic to cells on the basis of 'loss of expression' might be an alternative way to target cancer cells; this approach could be adopted in the case of missing tumour-suppressor proteins, a genetic event that is common in tumour development.

It was the observation of a ubiquitous chromosomal translocation in chronic myelogenous leukaemia that first prompted the development of Gleevec. Although Gleevec was initially developed as an inhibitor of the bcr-abl kinase[6], it is now known to inhibit a select family of tyrosine kinases, including KIT and PDGFR, in a variety of tumour types[50]. The application of Gleevec to a specific, oncogene-dependent form of sarcoma, known as gastrointestinal stromal tumour (GIST), had spectacular anti-tumour effects and provided proof that both haematopoietic and solid tumours are amenable to targeted therapy[51,52].

The frequency of relapse because of resistance currently limits the effectiveness of this treatment. The need for other treatments is stimulating the search for similarly effective molecules that might be simultaneously applied, as in the anti-HIV cocktail[53,54].

Part of the success of Gleevec can be attributed to the intensity and productivity of the academia–industry interaction that took place throughout its development as an anti-cancer agent. This new and effective cancer treatment would not be available today without this cooperative interaction[6,52], nor would its wider application to include the treatment of the solid tumour GIST have been achieved[55,56]. Furthermore, the first treatment for Gleevec-resistant patients is now emerging from a similar academia–industry interaction with the development of the multi-kinase inhibitor SU11248 (ref. 57).

Transcriptome and genome data also identify targets that are potentially suitable for immunotherapeutic intervention. Among the genes that are most actively sought are those that have their expression restricted to one or more non-essential tissues or cells, such as the breast, the prostate or melanocytes, irrespective of disease state[58–61]. These genes are potential targets for vaccines or antibodies, because the targeted destruction of the remnants of an organ, such as the prostate, following its surgical removal is clinically acceptable irrespective of the disease state of the individual cells. As potential targets, these molecules have the great advantage of typically being present at high levels in tumours. In addition, new genes are being discovered that are expressed in a range of tumours but are restricted in their normal expression to one or more non-essential or embryonic tissues distinct from that in which the tumour is found. A particularly striking example is the growing number of cancer genes in the testis that are expressed only in tumours and in human gametes[62–64]. Although the contribution of these genes to the process of tumorigenesis is not yet known, they are important targets of potential cancer vaccines[65].

There are not, as yet, any molecular, therapeutic cancer vaccines for clinical use, although clinical trials are underway[66,67]. However, antibodies are already making a substantial contribution to therapy[68–70]. The monoclonal antibody Herceptin, for example, is targeted to the epidermal-growth-factor-like receptor encoded by the *HER-2/neu* gene, which is overexpressed in 20–30% of breast cancer patients. Rituxan is targeted to CD20 (which is selectively expressed on B cells) and is used for the treatment of patients with non-Hodgkin's lymphoma both as single-agent therapy and in combination with chemotherapy[71]. Providing further optimism for the application of antibodies to cancer intervention is the recent introduction of Avastin[9], the first targeted anti-angiogenesis agent, now approved as a treatment for metastatic colon cancer[72].

Herceptin, Rituxan and Avastin all function as naked antibodies. But antibodies can also be used to deliver toxic compounds to the tumour. Examples of this approach have already reached the clinic. Zevalin, a derivative of Rituxan ligated to the radioisotope ^{90}Y, is the first radiolabelled monoclonal antibody to be approved for the treatment of cancer[73]. Another example is Mylotarg, a monoclonal antibody used for the treatment of acute myeloid leukaemia. It targets CD33 and is coupled to the toxin calicheamicin, which initiates apoptosis[74].

From genomic discovery to clinical utility

Mutations in certain tumour suppressors, such as p53 and APC, occur in many cancer types at high frequency. This information has raised hopes that targeted therapies could work on a broad spectrum of tumours. Although exploitation of these targets has proved to be quite challenging, several creative approaches are being explored, such as the use of selective replication of viruses in cancer cells, RNA interference or ribozymes[75–78]. For example, viruses have been engineered that selectively replicate in cells that have a loss of p53 function[75].

In the meantime, the most commonly used treatment involves the targeting of oncogenes, in particular kinases. But alterations in specific kinase genes are seen only in a subset of tumours. For example, colon cancer has a broad spectrum of low-frequency mutations in kinases, each of which could be a potential target[13,14]. This molecular heterogeneity complicates the development of universally applicable therapies. The development of vaccines and antibody therapy against an array of cancers is similarly hampered by potential targets being found only in subsets of any tumour type. An individual target is rarely present in as many as half of the tumours that arise in a given tissue[62].

Although a piecemeal approach to therapy may be unavoidable at present, a comprehensive genomic approach to the discovery of molecular alterations may identify unifying features that could result in broader-range therapies. Such common features could provide a means of targeting rare tumours or tumours more common in distinct geographical areas in a cost-effective manner. For example, in the case of kinase mutations in colon cancer, although no single mutation has been found in all of these cancers, a high proportion of cancers do have a kinase mutation, sometimes at the corresponding amino-acid position in different genes — for example, in the autoinhibitory activation loop[13].

It is highly likely that targetable mutations (or other molecular changes) common to many cancers will be found and will open the way for the targeting of different types of tumour (including rare tumours that otherwise might not receive sufficient attention) that can be inhibited by the same agent, as in the case of Gleevec. A key to progress in this area is the establishment of assays that can indicate potential success in the clinic.

Tools and assays for targeting molecular changes in cancer

To discover useful small-molecule inhibitors once mutations or other alterations have been identified in cancers, several approaches (Fig. 2) are available. One clever strategy is to use two cell lines that are identical (isogenic) except for a single cancer-causing mutation for screening. For example, Kinzler and colleagues, who developed this approach, compared cells with and without a *K-ras* mutation[79]. The two cell lines were labelled with different fluorescent markers, then mixed together and allowed to grow in the presence of small-molecule libraries. The relative growth rates of the cells were monitored using fluorescence spectroscopy. If any of the small molecules added inhibited the growth of the cancer-causing cell line, but not the other line, then its properties were explored further.

To be able to screen for inhibitors, scientists in research laboratories of even moderate size need to have access to materials such as diverse small-molecule libraries both for drug development and for the analysis of gene function in model systems or cell lines (Fig. 3).

Although at an early stage, ways of accessing chemical libraries are being introduced, and there are strong indications that this process will continue. One such programme is the National Cancer Institute (NCI) Initiative for Chemical Genetics (ICG), which includes support for ChemBank (http://chembank.med.harvard.edu)[80]. This programme aims to systematically identify perturbational small molecules for each cancer-related protein coded in the human genome. This should help to better define regulatory pathway networks in cancer, thereby providing valuable information for the development of cancer treatments[81]. Coupled with the National Institutes of Health (NIH) newly announced goal — part of its Roadmap initiative (http://nihroadmap.nih.gov) — to apply small-molecule drug discovery on a genome scale[82], as well as the increasing role of the National Human Genome Research Institute in translational

research[83], the prospects of greater access to molecules for screening look good. Other established efforts, such as the NCI's Rapid Access to Investigational Drugs (RAID) programme (http://dtp.nci.nih.gov/docs/raid/raid_index.html), are providing opportunities for academic (and industrial) researchers to have an active role in clinical development. Moreover, the NIH Roadmap initiative is now strongly advocating major academic participation in many areas of translational research.

Although much of the emphasis has been placed by researchers on the screening of synthetic molecules created by combinatorial chemistry, it should be remembered that the majority of clinically successful small molecules for the treatment of cancer are based on natural compounds[84]. Most of the world's biodiversity exists in developing countries, which highlights the need to work internationally to improve cancer therapy[85].

The availability of small-molecule libraries is only a starting point. Functional data are needed for each target, testing in relevant animal models is required, and potential treatments (leads) have to be improved, re-tested and refined by medicinal chemists. Thereafter, the problems of good manufacturing practice (GMP) production and regulatory approval have to be addressed before beginning to negotiate issues such as toxicology, pharmokinetics and access to tumour tissue.

Similar difficulties need to be overcome for therapeutic antibodies and vaccines to be developed. For either form of immunotherapy, the distribution of the potential target protein throughout the body will need to be determined, and for this, monoclonal antibodies will be used as investigative reagents. Although such antibodies can be raised using DNA or peptides, the use of whole protein has historically proved to be most reliable. But this requires the routine production of at least small quantities of such proteins, a process that is currently low throughput, although projects aimed at the large-scale generation (within the academic domain) of comprehensive sets of human proteins have been planned (FLEXGene)[86].

In addition, the analysis of naturally occurring immune responses to the potential cancer-causing targets is important in identifying those that are most likely to be immunogenic and protective. Even after the suitability of a target has been confirmed, further challenges lie ahead: the availability of experimental therapeutic reagents, in the form of antibodies or recombinant proteins, is even more limited than it is for small-molecule inhibitors. At present, antibodies are produced in mice and then adapted for human use ('humanized'). Alternative strategies include the use of phage display libraries for antibody generation[87,88], as well as selectively binding peptides[89,90]. The production of the quantity (grams) required for initial human trials is a major undertaking. And the production of gram quantities of highly purified recombinant proteins for vaccines, under GMP conditions, can at present only be pursued on a one-by-one basis in purpose-built facilities.

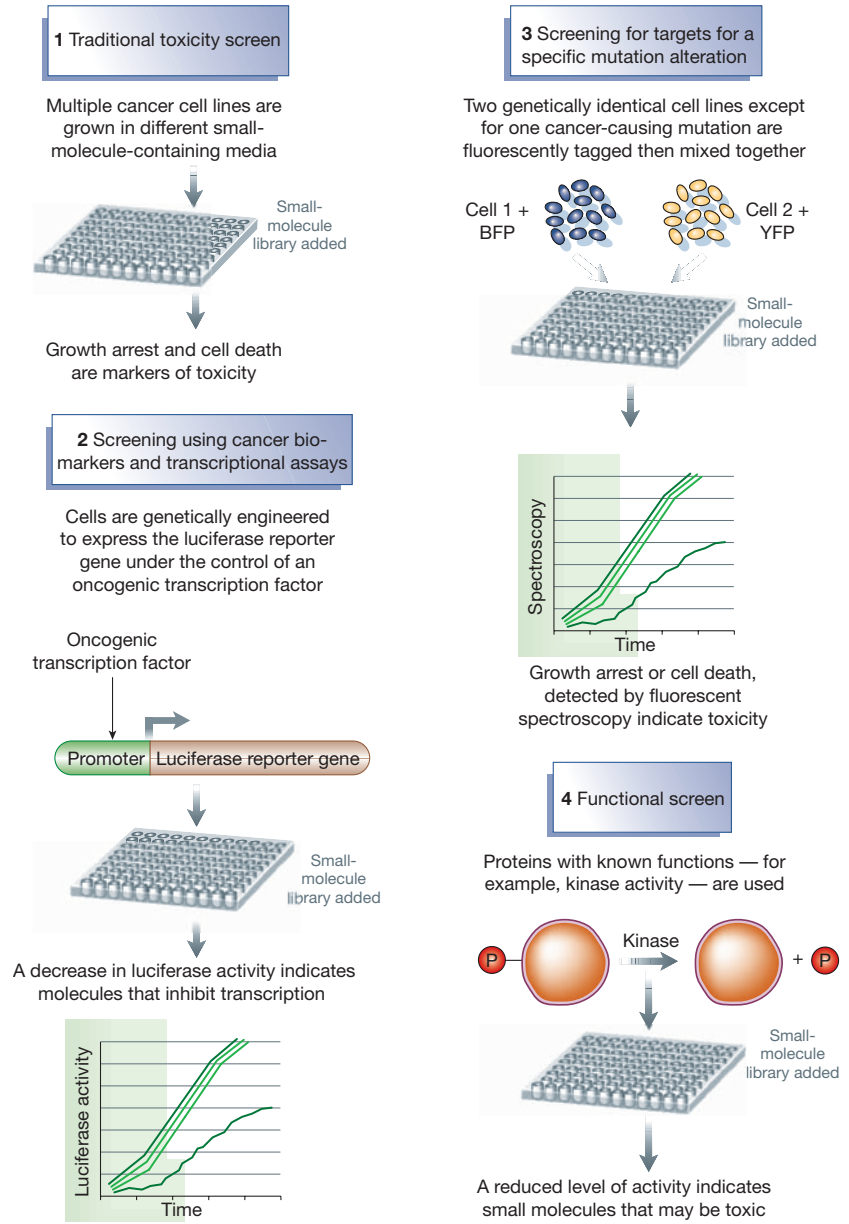

Figure 2 Examples of general drug-screening approaches. Recently developed screens (shown in panels 2–4) take advantage of the knowledge generated by oncogenomics and move beyond the traditional toxicity screens portrayed in panel 1. Known mutations can be used to develop a controlled screen for a single alteration within isogenic cell lines[79]. BFP, blue fluorescent protein; YFP, yellow fluorescent protein. Pathways or profiles of malignant transcription can be used as biomarkers for transcriptional screens[97]. Defined oncogenic mutations can form the basis of functional screens.

For much of the academic research community, the barriers preventing the development of immunotherapy reagents are formidable. However, they are not insurmountable. For example, the Ludwig Institute for Cancer Research has taken several potentially therapeutic antibodies, discovered by its own scientists, into clinical trials using reagents produced either in collaboration with industry or within its own biological production facilities[91–93]. Moreover, numerous vaccine trials have been undertaken by the Institute on the basis of two important and widely expressed antigens, MAGE and NY-ESO-1, both of which were first discovered by scientists at the institute[94,95]. Very recently, St Jude Children's Research Hospital announced the construction of a large in-house production facility

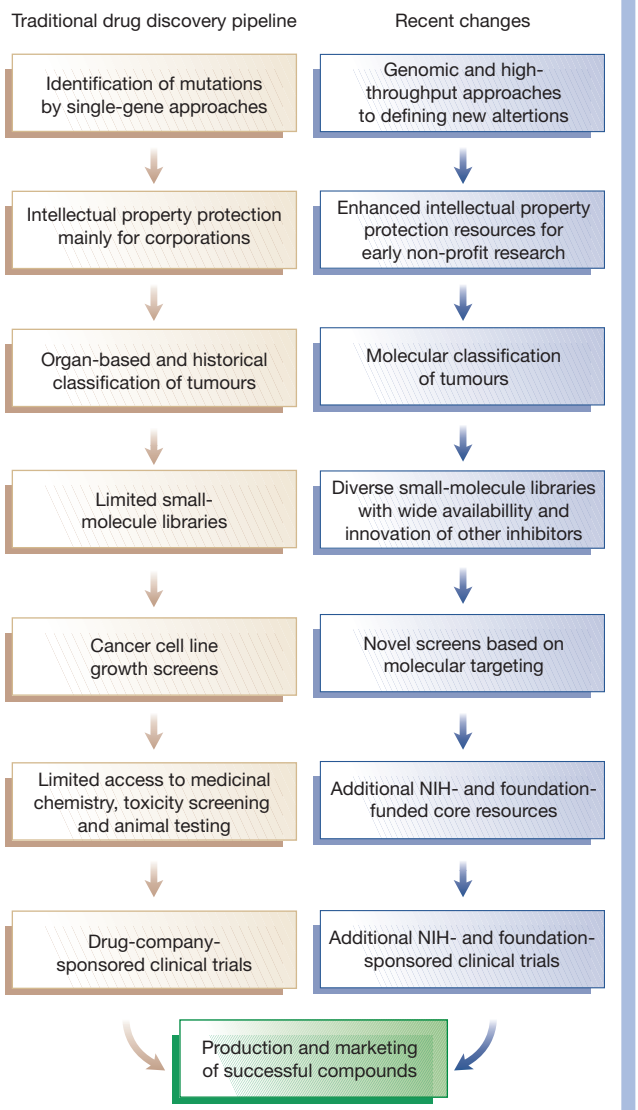

Figure 3 A comparison of how cancer drug development has occurred traditionally and more recently.

for generating novel therapeutic agents, and its own staff will take forward any leads into clinical trials.

Involving the entire research community

Over the past few decades, the complexity and difficulty in developing cancer interventions has become increasingly apparent. One informative indication of this challenge is the recent estimate that the total cost to develop a novel therapeutic is over US$800 million — and much of this figure is accounted for by the number of failed leads[96]. It is hoped that increased industry–academia collaboration in the early assessment of potential therapeutics will significantly reduce the number of failures, and thus improve the rate of development and bring down costs. Among the important advances of the HGP is a change in thinking about research organization based on the establishment of international teams of academic and industrial scientists. We think that such collaborations will be vital for the generation of cancer therapeutics. For this to occur, academic researchers will need to adopt a more prominent role in the leadership of early-stage clinical trials (Fig. 3).

We are at a crossroads where the community as a whole is accepting the responsibility for taking forward the advances that have been made in basic genomics, not only into their focused application in identifying new therapeutic targets, but also in the development of new therapies. The range of talent, approaches and organizational structures that this will bring together will certainly increase the probability of significant successes and justify the hopes that launched the HGP when it was first initiated some 15 years ago. Indeed, it is only when we can convincingly argue that the cancer victim is the real beneficiary of this endeavour that success can be claimed. □

doi:10.1038/nature02627

1. Vogelstein, B. & Kinzler, K. W. *The Genetic Basis of Human Cancer* (McGraw-Hill, New York, 2002).
2. Balmain, A., Gray, J. & Ponder, B. The genetics and genomics of cancer. *Nature Genet.* **33** (suppl.), 238–244 (2003).
3. Popescu, N. C. Comprehensive genetic analysis of cancer cells. *J. Cell. Mol. Med.* **4,** 151–163 (2000).
4. International Human Genome Sequencing Consotium. Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
5. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
6. Druker, B. J. STI571 (Gleevec) as a paradigm for cancer therapy. *Trends Mol. Med.* **8,** S14–S18 (2002).
7. Kim, J. A. Targeted therapies for the treatment of cancer. *Am. J. Surg.* **186,** 264–268 (2003).
8. Smith, I. E. Efficacy and safety of Herceptin in women with metastatic breast cancer: results from pivotal clinical studies. *Anticancer Drugs* (suppl. 4), S3–S10 (2001).
9. Salgaller, M. Technology evaluation: bevacizumab, Genentech/Roche. *Curr. Opin. Mol. Ther.* **5,** 657–667 (2003).
10. Schadt, E. E., Monks, S. A. & Friend, S. H. A new paradigm for drug discovery: integrating clinical, genetic, genomic and molecular phenotype data to identify drug targets. *Biochem. Soc. Trans.* **31,** 437–443 (2003).
11. Fischer, O. M., Streit, S., Hart, S. & Ullrich, A. Beyond Herceptin and Gleevec. *Curr. Opin. Chem. Biol.* **7,** 490–495 (2003).
12. Davies, H. *et al.* Mutations of the *BRAF* gene in human cancer. *Nature* **417,** 949–954 (2002).
13. Bardelli, A. *et al.* Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* **300,** 949 (2003).
14. Samuels, Y. *et al.* High frequency of mutations of the *PIK3CA* gene in human cancers. *Science* **304,** 554 (2004).
15. Hilgenfeld, E. *et al.* Spectral karyotyping in cancer cytogenetics. *Methods Mol. Med.* **68,** 29–44 (2002).
16. Lucito, R. *et al.* Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res.* **13,** 2291–2305 (2003).
17. Cowell, J. K. & Nowak, N. J. High-resolution analysis of genetic events in cancer cells using bacterial artificial chromosome arrays and comparative genome hybridization. *Adv. Cancer Res.* **90,** 91–125 (2003).
18. Liang, G. *et al.* DNA methylation differences associated with tumor tissues identified by genome scanning analysis. *Genomics* **53,** 260–268 (1998).
19. Cottrell, S. E. *et al.* A real-time PCR assay for DNA-methylation using methylation-specific blockers. *Nucleic Acids Res.* **32,** e10 (2004).
20. Greshock, J. *et al.* 1-Mb resolution array-based comparative genomic hybridization using a BAC clone set optimized for cancer gene analysis. *Genome Res.* **14,** 179–187 (2004).
21. Wang, T. L. *et al.* Digital karyotyping. *Proc. Natl Acad. Sci. USA* **99,** 16156–16161 (2002).
22. Wang, T. L. *et al.* Digital karyotyping identifies thymidylate synthase amplification as a mechanism of resistance to 5-fluorouracil in metastatic colorectal cancer patients. *Proc. Natl Acad. Sci. USA* **101,** 3089–3094 (2004).
23. Volik, S. *et al.* End-sequence profiling: sequence-based analysis of aberrant genomes. *Proc. Natl Acad. Sci. USA* **100,** 7696–7701 (2003).
24. Beroud, C. & Soussi, T. The UMD-p53 database: New mutations and analysis tools. *Hum. Mutat.* **21,** 176–181 (2003).
25. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4,** 177–183 (2004).
26. Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nature Genet.* **33,** 49–54 (2003).
27. Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl Acad. Sci. USA* **100,** 8418–8423 (2003).
28. Rosenwald, A. *et al.* Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. *J. Exp. Med.* **198,** 851–862 (2003).
29. van 't Veer, L. J. *et al.* Expression profiling predicts outcome in breast cancer. *Breast Cancer Res.* **5,** 57–58 (2003).
30. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347,** 1999–2009 (2002).
31. Nutt, C. L. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res.* **63,** 1602–1607 (2003).
32. Dyrskjot, L. *et al.* Identifying distinct classes of bladder carcinoma using microarrays. *Nature Genet.* **33,** 90–96 (2003).
33. Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. Serial analysis of gene-expression. *Science* **270,** 484–487 (1995).
34. Lal, A. *et al.* A public database for gene expression in human cancers. *Cancer Res.* **59,** 5403–5407 (1999).
35. Buckhaults, P. *et al.* Identifying tumor origin using a gene expression-based classification map. *Cancer Res.* **63,** 4144–4149 (2003).
36. Jongeneel, C. V. *et al.* Comprehensive sampling of gene expression in human cell lines with massively parallel signature sequencing. *Proc. Natl Acad. Sci. USA* **100,** 4702–4705 (2003).
37. Brentani, H. *et al.* The generation and utilization of a cancer-oriented representation of the human transcriptome by using expressed sequence tags *Proc. Natl Acad. Sci. USA* **100,** 13418–13423 (2003).
38. Strausberg, R. L. *et al.* An international database and integrated analysis tools for the study of cancer gene expression. *Pharmacogenomics J.* **2,** 156–164 (2002).
39. Saha, S. *et al.* Using the transcriptome to annotate the genome. *Nature Biotechnol.* **20,** 508–512 (2002).

40. Ota, T. *et al.* Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nature Genet.* **36,** 40–45 (2004).
41. Strausberg, R. L. *et al.* Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl Acad. Sci. USA* **99,** 16899–16903 (2002).
42. Xu, Q. & Lee, C. Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. *Nucleic Acids Res.* **31,** 5635–5643 (2003).
43. Wang, Z. *et al.* Computational analysis and experimental validation of tumor-associated alternative RNA splicing in human cancer. *Cancer Res.* **63,** 655–657 (2003).
44. Kriventseva, E. V. *et al.* Increase of functional diversity by alternative splicing. *Trends Genet.* **19,** 124–128 (2003).
45. Strausberg, R. L., Simpson, A. J. & Wooster, R. Sequence-based cancer genomics: progress, lessons and opportunities. *Nature Rev. Genet.* **4,** 409–418 (2003).
46. Drevs, J., Medinger, M., Schmidt-Gersbach, C., Weber, R. & Unger, C. Receptor tyrosine kinases: the main targets for new anticancer therapy. *Curr. Drug Targets* **4,** 113–121 (2003).
47. Sausville, E. A., Elsayed, Y., Monga, M. & Kim, G. Signal transduction — directed cancer treatments. *Annu. Rev. Pharmacol. Toxicol.* **43,** 199–231 (2003).
48. Cockerill, G. S. & Lackey, K. E. Small molecule inhibitors of the class 1 receptor tyrosine kinase family. *Curr. Top. Med. Chem.* **2,** 1001–1010 (2002).
49. Wilhelm, S. & Chien, D. S. BAY 43-9006: preclinical data. *Curr. Pharm. Des.* **8,** 2255–2257 (2002).
50. Joensuu, H. & Dimitrijevic, S. Tyrosine kinase inhibitor imatinib (STI571) as an anticancer agent for solid tumours. *Ann. Med.* **33,** 451–455 (2001).
51. Demetri, G. D. Targeting the molecular pathophysiology of gastrointestinal stromal tumors with imatinib. Mechanisms, successes, and challenges to rational drug development. *Hematol. Oncol. Clin. North Am.* **16,** 1115–1124 (2002).
52. Druker, B. Imatinib mesylate in the treatment of chronic myeloid leukaemia. *Expert Opin. Pharmacother.* **4,** 963–971 (2003).
53. Tipping, A. J. & Melo, J. V. Imatinib mesylate in combination with other chemotherapeutic drugs: *In vitro* studies. *Semin. Hematol.* **40,** 83–91 (2003).
54. Druker, B. Imatinib alone and in combination for chronic myeloid leukemia. *Semin. Hematol.* **40,** 50–58 (2003).
55. Joensuu, H. *et al.* Management of malignant gastrointestinal stromal tumours. *Lancet Oncol.* **3,** 655–664 (2002).
56. Joensuu, H. Effect of the tyrosine kinase inhibitor STI571 in a patient with a metastatic gastrointestinal stromal tumor. *N. Engl. J. Med.* **344,** 1052–1056 (2001).
57. GIST SU11248 Study Group. Clinical activity and tolerability of the multi-targeted tyrosine kinase inhibitor SU11248 in patients (pts) with metastatic gastrointestinal stromal tumor (GIST) refractory to imatinib mesylate. *Proc. Am. Soc. Clin. Oncol.* **22,** 814 (abstr. 3273) (2003).
58. Egland, K. A., Vincent, J. J., Strausberg, R., Lee, B. & Pastan, I. Discovery of the breast cancer gene *BASE* using a molecular approach to enrich for genes encoding membrane and secreted proteins. *Proc. Natl Acad. Sci. USA* **100,** 1099–1104 (2003).
59. Olsson, P., Motegi, A., Bera, T. K., Lee, B. & Pastan, I. *PRAC2*: a new gene expressed in human prostate and prostate cancer. *Prostate* **56,** 123–130 (2003).
60. Jager, D. Identification of a tissue-specific putative transcription factor in breast tissue by serological screening of a breast cancer library. *Cancer Res.* **61,** 2055–2061 (2001).
61. Bera, T. K. *NGEP*, a gene encoding a membrane protein detected only in prostate cancer and normal prostate. *Proc. Natl Acad. Sci. USA* **101,** 3059–3064 (2004).
62. Scanlan, M. J., Gure, A. O., Jungbluth, A. A., Old, L. J. & Chen, Y. T. Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. *Immunol. Rev.* **188,** 22–32 (2002).
63. Scanlan, M. J., Simpson, A. J. & Old, L. J. The cancer/testis genes: review, standardization, and commentary. *Cancer Immun.* **4,** 1 (2004).
64. Scanlan, M. J. *et al.* Identification of cancer/testis genes by database mining and mRNA expression analysis. *Int. J. Cancer* **98,** 485–492 (2002).
65. Old, L. J. Cancer/testis (CT) antigens — a new link between gametogenesis and cancer. *Cancer Immun.* **1,** 1 (2001).
66. Durrant, L. G. & Spendlove, I. Cancer vaccines entering Phase III clinical trials. *Expert Opin. Emerg. Drugs* **8,** 489–500 (2003).
67. Jager, E., Jager, D. & Knuth, A. Antigen-specific immunotherapy and cancer vaccines. *Int. J. Cancer* **106,** 817–820 (2003).
68. Albanell, J., Codony, J., Rovira ,A., Mellado, B. & Gascon, P. Mechanism of action of anti-HER2 monoclonal antibodies: scientific update on trastuzumab and 2C4. *Adv. Exp. Med. Biol.* **532,** 253–268 (2003).
69. Blackledge, G. & Averbusch, S. Gefitinib ('Iressa', ZD1839) and new epidermal growth factor receptor inhibitors. *Br. J. Cancer* **90,** 566–572 (2004).
70. Coiffier, B. Immunochemotherapy: the new standard in aggressive non-Hodgkin's lymphoma in the elderly. *Semin. Oncol.* **30,** 21–27 (2003).
71. Wannesson, L. & Ghielmini, M. Overview of antibody therapy in B-cell non-Hodgkin's lymphoma. *Clin. Lymphoma* **4** (suppl. 1), S5–S12 (2003).
72. Jain, R. K. Tumor angiogenesis and accessibility: role of vascular endothelial growth factor. *Semin. Oncol.* **29,** 3–9 (2002).
73. Milenic, D. E. & Brechbiel, M. W. Targeting of radio-isotopes for cancer therapy. *Cancer Biol. Ther.* **3** (2004).
74. Damle, N. K. & Frost, P. Antibody-targeted chemotherapy with immunoconjugates of calicheamicin. *Curr. Opin. Pharmacol.* **3,** 386–390 (2003).
75. McCormick, F. Cancer-specific viruses and the development of ONYX-015. *Cancer Biol. Ther.* **2** (suppl. 1), s157–s160 (2003).
76. Tong, A. W., Zhang, Y. A., Cunningham, C., Maples, P. & Nemunaitis, J. Potential clinical application of antioncogene ribozymes for human lung cancer. *Clin. Lung Cancer* **2,** 220–226 (2001).
77. Moon, C., Oh, Y. & Roth, J. A. Current status of gene therapy for lung cancer and head and neck cancer. *Clin. Cancer Res.* **9,** 5055–5067 (2003).
78. McNeish, I. A., Bell, S. J. & Lemoine, N. R. Gene therapy progress and prospects: cancer gene therapy using tumour suppressor genes. *Gene Ther.* **11,** 497–503 (2004).
79. Torrance, C. J., Agrawal, V., Vogelstein, B. & Kinzler, K. W. Use of isogenic human cancer cells for high-throughput screening and drug discovery. *Nature Biotechnol.* **19,** 940–945 (2001).
80. Strausberg, R. L. & Schreiber, S. L. From knowing to controlling: a path from genomics to drugs using small molecule probes. *Science* **300,** 294–295 (2003).
81. Koehler, A. N., Shamji, A. F. & Schreiber, S. L. Discovery of an inhibitor of a transcription factor using small molecule microarrays and diversity-oriented synthesis. *J. Am. Chem. Soc.* **125,** 8420–8421 (2003).
82. Zerhouni, E. Medicine. The NIH Roadmap. *Science* **302,** 63–72 (2003).
83. Collins, F. S., Morgan, M. & Patrinos, A. The Human Genome Project: lessons from large-scale biology. *Science* **300,** 286–290 (2003).
84. Newman, D. J., Cragg, G. M. & Snader, K. M. Natural products as sources of new drugs over the period 1981–2002. *J. Nat. Prod.* **66,** 1022–1037 (2003).
85. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403,** 853–858 (2000).
86. Brizuela, L., Richardson, A., Marsischky, G. & Labaer, J. The FLEXGene repository: Exploiting the fruits of the genome projects by creating a needed resource to face the challenges of the post-genomic era. *Arch. Med. Res.* **33,** 318–324 (2002).
87. Kretzschmar, T. & von Ruden, T. Antibody discovery: phage display. *Curr. Opin. Biotechnol.* **13,** 598–602 (2002)
88. Brekke, O. H. & Loset, G. A. New technologies in therapeutic antibody development. *Curr. Opin. Pharmacol.* **3,** 544–550 (2003).
89. Pasqualini, R. & Ruoslahti, E. Organ targeting *in vivo* using phage display peptide libraries. *Nature* **380,** 364–366 (1996).
90. Ellerby, H. M. *et al.* Anti-cancer activity of targeted pro-apoptotic peptides. *Nature Med.* **5,** 1032–1038 (1999).
91. Scott, A. M. *et al.* A Phase I dose-escalation study of sibrotuzumab in patients with advanced or metastatic fibroblast activation protein-positive cancer. *Clin. Cancer Res.* **9,** 1639–1647 (2003).
92. Welt, S. *et al.* Phase I study of anticolon cancer humanized antibody A33. *Clin. Cancer Res.* **9,** 1338–1346 (2003).
93. Scott, A. M. *et al.* Specific targeting, biodistribution, and lack of immunogenicity of chimeric anti-GD3 monoclonal antibody KM871 in patients with metastatic melanoma: results of a phase I trial. *J. Clin. Oncol.* **19,** 3967–3987 (2001).
94. Jager, E. *et al.* Induction of primary NY-ESO-1 immunity: CD8+ T lymphocyte and antibody responses in peptide-vaccinated patients with NY-ESO-1+ cancers. *Proc. Natl Acad. Sci. USA* **97,** 12198–12203 (2000).
95. Atanackovic, D. Vaccine-induced CD4+ T cell responses to MAGE-3 protein in lung cancer patients. *J. Immunol.* **172,** 3289–3296 (2004).
96. Duyk, G. Attrition and translation. *Science* **302,** 603–605 (2003).
97. Rapisarda, A. *et al.* Identification of small molecule inhibitors of hypoxia-inducible factor 1 transcriptional activation pathway. *Cancer Res.* **62,** 4316–4324 (2002).

Acknowledgements We thank G. Demetri for discussions and advice during the preparation of this manuscript.

Competing interests statement The authors declare that they have no competing financial interests.