Review

50 Years of Preclinical Anticancer Drug Screening: Empirical to Target-Driven Approaches

Marie Suggitt and Michael C. Bibby
Tom Connors Cancer Research Centre, University of Bradford, Bradford, United Kingdom

ABSTRACT

The number of anticancer agents that fail in the clinic far outweighs those considered effective, suggesting that the selection procedure for progression of molecules into the clinic requires improvement. The value of any preclinical model will ultimately depend on its ability to accurately predict clinical response. This review focuses on the major contributions of preclinical screening models to anticancer drug development over the past 50 years. Over time, a general transition has been observed from the empirical drug screening of cytotoxic agents against uncharacterized tumor models to the target-orientated drug screening of agents with defined mechanisms of action. New approaches to anticancer drug development involve the molecular characterization of models along with an appreciation of the pharmacodynamic and pharmacokinetic properties of compounds [e.g., the US National Cancer Institute (NCI) in vitro 60-cell line panel, hollow fiber assay, and s.c. xenograft]. Contributions of other potentially more clinically relevant in vivo tumor models including orthotopic, metastatic, and genetically engineered mouse models are also reviewed. Although this review concentrates on the preclinical screening efforts of the NCI, European efforts are not overlooked. Europe has played a key role in the development of new anticancer agents. The two largest academic drug development groups, the European Organisation for Research and Treatment of Cancer and Cancer Research UK, have been collaborating with the NCI in the acquisition and screening of compounds since the 1970s. As with the drug development process internationally, rational pharmacodynamic approaches have more recently been adopted by these two groups.

TUMOR MODELS (1955-1985)

In 1955, following reports that the correlation between compound efficacy against transplanted tumors and clinical activity was substantially better than the previously used mammalian cell and bacterial cultures (1), the National Cancer Institute (NCI) began a large-scale anticancer drug screening program testing agents against a panel of three mouse tumor models: sarcoma 180, L1210 leukemia, and carcinoma 755 (2–6). In 1965, screening was limited to the use of only the L1210 and Walker 256 carcinosarcoma murine models, and by 1968, synthetic agents were screened in L1210 alone, whereas natural plant products were screened in L1210 and P388 leukemias. In 1972, B16 melanoma and the Lewis lung carcinoma mouse models were introduced (6–8). The reliance on the L1210 model in the first 25 years of anticancer drug screening raised concerns that screening against a rapidly growing animal leukemia may have resulted in preferential selection of drugs that were only active against rapidly growing tumors (9). During this time, clinical response of human leukemias and lymphomas improved (10, 11), whereas treatment response was less promising for most human solid tumors. As a consequence, in 1976 the Division of Cancer Treatment at the NCI introduced a new tumor panel incorporating transplantable solid “human” tumor models that were representative of the major histologic types of cancer prevalent in the United States at the time. This development followed the revolutionary discovery of the nude athymic (nu/nu) mouse (12) and the successful growth of human tumor xenografts (13). The panel consisted of matched animal and human tumors of the breast (CD85/MX-1), colon (colon 38/CX-1), and lung (Lewis/LX-1), along with the L1210 leukemia and B16 melanoma syngeneic models (6, 9). Syngeneic models involved inoculation i.p., s.c., or i.v., whereas human tumor xenografts were grown under the subrenal capsule (14). The subrenal capsule assay uses small tumor fragments growing under the renal capsule of athymic mice and normal immunocompetent mice (14). Although the subrenal capsule assay was labor-intensive, it provided a rapid means of evaluating new agents against human tumor xenografts at a time when longer-term s.c. assays were not feasible. The subrenal capsule assay has shown good predictive value of clinical response with an overall evaluable assay of 90% (15).

Prior to testing against this new “mouse-human” tumor panel, compounds were subjected to a relatively cost-effective in vivo “prescreen” using the P388 model. This model was sensitive to most classes of clinically effective drugs, yet was sufficiently discriminating to avoid overloading the panel (16). It was revealed that this mouse-human tumor panel (1976-1982) identified antitumor agents (e.g., taxol) that would have been missed by the L1210 model alone (9). Approximately 30% of compounds found to be active in at least one human tumor xenograft were missed by syngeneic models. Therefore, the
mouse-human tumor panel successfully achieved the goal of providing new agents for clinical trial, although the correlation between preclinical screening and clinical efficacy was found to be extremely low (17). This was thought to reflect either the poor predictivity of selection strategies or the low number of active compounds that actually existed.

Subsequently, in 1982 the NCI employed a new strategy involving a sequential process of “progressive selection” whereby a compound was presented with a progressively greater biological and pharmacologic challenge at each stage. Compounds were first subjected to the P388 prescreen model, and then a standard panel including tumor models known to produce a relatively high-yield of active compounds (MX-1, B16, and L1210). A new model, the M5076 sarcoma as well as advanced and multidrug-resistant tumor models, were incorporated in this standard panel (9). Agents found to be active in the standard tumor panel were subjected to secondary screening. The selection of secondary tumor models was “drug-orientated.” Agents were subjected to specific tumor models based on the known properties of each individual compound and previous experience in the standard tumor panel.

At this time, several retrospective preclinical-clinical correlation studies were reviewed (18), but no apparent positive correlation between preclinical and clinical efficacy based on tumor type was found. It was suggested that the lack of histologically based correlations may be a consequence of experimental design which limited tumors to one mouse and one human correlation for each of the three major types. It was also suggested that a “model system” composed of several tumors of the same type might, on the basis of percentage responders, predict for a reasonable clinical response rate against similar type tumors (18).

**HUMAN TUMOR STEM CELL ASSAY/CLONOGENIC ASSAY**

The development of the human tumor stem cell (HTSC) assay (19–21) offered an approach to determine whether using a model system composed of several tumors of the same type may be better at predicting a reasonable clinical response than previous in vivo tumor models (9). The HTSC assay was disease-orientated in concept and involved the growth on soft agar of colonies derived from freshly explanted human tissue. Compounds were tested against tumor colonies and activity defined by the growth inhibition of colonies. Salmon and colleagues (19) compared in vitro results to the clinical responses of myeloma and ovarian cancer patients and the study showed clear correlations and unique patterns of sensitivity and resistance. This study showed sufficient promise to warrant larger-scale testing to determine the efficacy of the HTSC assay for selecting clinically active agents and individualizing cancer treatment.

The applicability of the HTSC assay for drug screening purposes in terms of feasibility, validity, and potential to identify new antitumor agents was investigated (22). The testing of established standard chemotherapeutic agents in this pilot study revealed that most agents were found to be active with the exception of those requiring systemic activation. Clinically ineffective agents were confirmed to be true negatives with 97% accuracy. Other groups also showed the potential use of the assay in predicting clinical activity (23–29). Typically, an assay was shown to predict drug resistance with 90% accuracy and clinical drug sensitivity with between 40% and 70% accuracy. Additionally, of 79 compounds found previously to be inactive using the P388 prescreen, 14 were active in the HTSC assay. Reproducibility of survival values within assays and between laboratories was also revealed.

However, several limitations prevented the use of this assay for large-scale screening (29–31), the main criticism being that many tumor types have a low plating efficiency. Subsequently, only breast, colorectal, kidney, lung, melanomas, and ovarian tumors produce a sufficient yield of evaluable assays. Hence, the number of patients for whom treatment may be chosen by clonogenic assays was frequently <50%, although recently the growth rates of primary tumor tissues in the HTSC assay has significantly improved (70-80%; ref. 29). Additional problems encountered include labor intensity, automation, and difficulty in attaining a single cell suspension from human solid tumors. Clonogenic assays also include sources of substantial error, and quantification of data, cell survival curves, and colony size are often criticized (29–31).

To date, there are no phase III clinical studies of individualized therapy demonstrating a significant increase in survival compared with empirically determined standard treatment, therefore the clonogenic assay has not found a practical established role in the individualization of patient therapy (29). Clonogenic assays are still widely used as a secondary screen by independent researchers (29, 31, 32). At the Institute for Experimental Oncology in Freiburg, an in vitro/in vivo testing procedure is employed using target-defined tumor models. Patients’ tumors are tested directly in the in vitro clonogenic assay, or after being established as a permanent xenograft model. Agents are tested using both an empirical and target-orientated screening strategy. In addition to the routine end point of the clonogenic assay (inhibition of colony growth), pharmacodynamic assays are employed to determine compound activity (29).

**HUMAN TUMOR IN VITRO CELL LINE SCREEN, 1985 TO PRESENT DAY**

In 1985 the phase-down of the in vivo P388 prescreen, HTSC assay and the human/murine tumor panels began. The feasibility of employing human tumor cell lines for large-scale drug screening was investigated (33). At the time, it was appreciated that this new program could have the potential to rapidly evaluate a large number of anticancer agents, yet it was argued that “appropriate” transplantable mouse tumors should still have their place in the drug development program and involve a broader preclinical evaluation (i.e., pharmacokinetics, toxicology, metabolism, drug bioavailability, and therapeutic index; refs. 34, 35). With regard to therapeutic index, dose-limiting toxicity in human is sometimes difficult to predict from mouse studies (e.g., bone marrow toxicity and neurotoxicity).

This new in vitro human tumor cell line screen initiated in April 1990, shifted the NCI screening strategy from being “compound-orientated” to “disease-orientated.” The initial cell line panel incorporated a total of 60 different human tumor cell...
lines of diverse histologies derived from seven types of cancer (brain, colon, leukemia, lung, melanoma, ovarian, and renal), including drug-resistant cell lines. Thus, like the HTSC assay, the in vitro 60-cell line screen offered an approach to use several tumors of the same type. In December 1992, 10 of the original cell lines were replaced by a selection of breast and prostate cancer cell lines (36).

To date ~85,000 compounds have been screened against this in vitro cell line panel in a short-term assay (37), whereby the nonclonogenic protein stain sulforhodamine B assay (38) is used to determine cell viability. Each compound is tested over a 5-log concentration range against each of the 60 cell lines in a 2-day assay and generates 60 dose-response curves. These data generate a characteristic profile or “fingerprint” of cellular response, i.e. the “mean graph” (39). “COMPARE” is a computerized, pattern recognition algorithm used in evaluating and exploiting the fingerprint data in order to determine the degree of similarity between mean graph profiles generated by similar or different compounds (39).

Initial studies revealed that compounds matched by their mean-graph patterns often had related chemical structures. Closer examination of this phenomenon led to the realization that compounds of either related or unrelated structures, and matched by their mean-graph patterns, frequently shared the same or related biochemical mechanisms of action.

In contrast to fingerprints with pattern similarity to standard agents in the database, compounds have been detected which produce striking differential cytotoxicity fingerprints. These “COMPARE-negative” agents indicate that they have a unique mechanism of action.

Since the early 1990s, data on the expression of molecular characteristics within the 60-cell human tumor panel has been accumulated. The implication of this being that the sensitivity of a cell line, along with knowledge of its molecular characteristics may indicate that a compound’s cytotoxicity is mediated by its interaction with a known molecular target. Alternatively, differential expression in the form of a mean graph of particular molecular markers (e.g., MDR1/p-glycoprotein) may indicate why particular cell lines may be resistant to a test compound (40).

An extensive collection of molecular target data exists including kinases, phosphatases, genes associated with cell cycle control, apoptosis, DNA repair, signal transduction, oncogenes, and drug metabolizing reductase enzymes (http://dtp.nci.nih.gov) (ref. 37).

In 1999 an in vitro prescreen was introduced whereby compounds were screened in a three-cell line panel using three highly sensitive cell lines, MCF-7 (breast carcinoma), NCI-H460 (lung carcinoma), and SF-268 (glioma). The rationale for this prescreen was the observation that 85% of compounds screened in the past had shown no evidence of antiproliferative activity, and a three-line prescreen was shown to efficiently remove many of the inactive compounds from unnecessary and costly full-scale evaluation in the 60-line panel.

Three end points are used to determine compound activity and whether a compound will be considered for further evaluation. These are GI50 (concentration required to inhibit 50% of cells), total growth inhibition, and LC50 (concentration required to kill 50% of cells; ref. 37). Compounds possessing disease specificity or that are COMPARE-negative are also referred for secondary in vivo screening, initially using the hollow fiber assay (HFA; ref. 41). Compounds considered active in the HFA are prioritized for in vivo xenograft testing.

The present human tumor cell line in vitro screen is technically simple, relatively fast, cheap, reproducible, and provides valuable indicative data of mechanistic activity and target interaction. Yet it is not without its limitations. In vitro methods are susceptible to false-positive and false-negative results. It is also clear that factors other than the inherent chemosensitivity of tumor cells significantly influence the outcome of chemotherapy in vivo (e.g., pharmacokinetics, tumor microregions/pH, and pO2; refs. 34, 42). Such factors are not represented by the in vitro 60-cell line screening assay, yet it is appreciated that this assay was designed only to select compounds for a secondary, more comprehensive, in vivo testing. The original intention of the NCI/DTP was to produce a high-throughput in vitro screen that would be sufficiently discriminatory to ensure that only a relatively small number of compounds would be selected for further evaluation in human tumor xenograft models. This has not been the case and subsequently the in vivo HFA was implemented in 1995 in attempt to prioritize compounds for secondary xenograft screening and help reduce the large number compounds that were forming a bottleneck for entry into secondary xenograft testing.

THE HOLLOW FIBER ASSAY

Based on previous microencapsulation and hollow fiber culture systems (43–45), Hollingshead and colleagues (41) developed the HFA designed to identify in vivo activity of potential anticancer compounds. The HFA assesses the pharmacologic capacity of compounds to reach two physiologic compartments within the nude mouse and shows a practical means of quantifying viable tumor cell mass.

In 1995, due to the feasibility of growing over 50 human tumor cell lines within biocompatible hollow fibers and the relatively rapid and cost-effective demonstration of in vivo activity compared with xenograft models, the NCI employed the HFA as a routine preliminary in vivo screening assay. Although xenograft models have in the past proved invaluable in developing current chemotherapeutic agents, they are accompanied by various limitations including high costs associated with large-scale screening, time, and number of mice required. Additionally, the empirical dosing and development of pharmacokinetic assays for each compound evaluated in xenograft models would greatly reduce the rate at which compounds could progress to the clinic (46).

The current NCI HFA protocol involves the short-term in vitro culture (24-48 hours) of a panel of 12 cell lines inside hollow fibers, followed by in vivo implantation at both s.c. and i.p. sites of the nude mouse. The assay has the potential to simultaneously evaluate compound efficacy against a maximum of six cell lines (three cell lines/fibers per site). Mice are treated with test compound at two doses for up to 4 days, fibers excised and analyzed for cell viability using a modified 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide assay. Compounds are identified as active using a detailed scoring system and optimal or near-optimal treatment regimens are indicated for subsequent testing using xenograft models (41, 46).
In contrast to the use of the HFA as a rapid in vivo screening model, independent research groups have more recently used the HFA to investigate the pharmacodynamics of anticancer agents in vivo. Authors have taken advantage of the practical feasibility of pure cell population retrieval and have used a target-orientated approach for the evaluation of compound activity. Pharmacodynamic end points investigated include protein/gene/mRNA expression, tubulin/DNA damage, and cell cycle disruption (47–51). Other groups have also used the HFA to investigate specific areas of tumor biology (52, 53). Primary human tumor cells have also been cultivated within fibers in vivo (54). Such studies have been undertaken in an attempt to use this relatively rapid in vivo test system to predict patient response in the clinic.

Supportive evidence for the use of the HFA as an in vivo prescreen is derived from the good predictivity of in vivo xenograft activity (46, 47, 55, 56). Initial proof-of-concept, assay development, and validation studies, whereby compounds were tested in both in vivo xenograft and hollow fibers, have indicated that a xenograft-positive compound was unlikely to be missed by testing in the HFA (55).

Johnson and colleagues (56) showed that the likelihood of finding xenograft activity in at least 1/3 of in vivo models rose with increasing i.p. activity. This correlation was not evident with s.c. fibers. It has been suggested that there may be insufficient time for angiogenesis to occur at the s.c. site, thereby limiting drug delivery (57). Arguing against this hypothesis, Hollingshead and colleagues (58) showed the delivery of luciferin to the s.c. site within 6 minutes following i.p. administration 3 hours after in vivo implantation. In this novel study, hollow fibers were seeded with tumor cells transfected with the luciferase gene and implanted in nude mice. Bioluminescence permitted the monitoring of tumor growth with s.c. fibers. It has been suggested that there may be insufficient time for angiogenesis to occur at the s.c. site, thereby limiting drug delivery (57). Arguing against this hypothesis, Hollingshead and colleagues (58) showed the delivery of luciferin to the s.c. site within 6 minutes following i.p.

A strong correlation between potency in the 60-cell line screen and the HFA has also been reported (56), providing support for the argument that the HFA is unlikely to miss active agents at this preliminary filtering in vivo stage.

Initially (1995-1999), compounds defined as active in the HFA were tested in a series of xenograft models in a range of histologies encompassing the sensitive cell types determined in hollow fibers. However, in 1999, the development of agents based purely on empirical antitumor activity in human tumor xenografts without definition of an agent’s biological target was abandoned (59). New evolving strategies seemed to be in response to the advent of a new era of small molecule anticancer therapeutics that were replacing classic cytotoxic agents. Because this time positive hollow fiber results were used to prioritize compounds for further pharmacologic and mechanistic studies (46). Hollow fiber studies are accompanied by an effort to define the pharmacokinetics (peak concentration and area under the concentration × time curve) associated with drug activity.

The current NCI HFA does not specifically define any precise mechanisms of drug action that may have previously been indicated by the recent molecular characterization of the in vitro human tumor 60-cell line screen from which compounds have progressed. It would seem logical to follow up any indication of a compound’s mechanism of action using the relatively rapid and cost-effective hollow fiber model.

It is proposed that the continued use of the HFA will exploit its ability to define specific pharmacodynamic end points at this early stage in a drug’s evaluation (46) as has been shown previously (47–51). This may allow further refinement in lead selection through immunohistochemical, genomic, proteomic, and enzyme assays of cells extruded from fibers excised from drug-treated mice (46).

**HUMAN TUMOR XENOGRAPHS IN SECONDARY SCREENING**

Before the implementation of the HFA (1990-1995), each agent was evaluated in the murine P388 model and three s.c. human tumor xenograft models defined as the most sensitive in vitro (60).

Between 1995 and 1999, compounds defined as active in the HFA were evaluated in three s.c. xenograft models using the most sensitive tumors identified by the HFA (60). Approximately 40% of compounds showed sufficient activity in the HFA and were selected for further in vivo testing. Therefore, a significant number of compounds were filtered out at this preliminary in vivo stage which would otherwise have used a significant amount of more costly in vivo resources. At the same time, specific assays were carried out to determine the pharmacologic/biological properties of a compound (60).

Xenograft tumors are generally established by the s.c. inoculation of tumor cells into nude mice (1.0 × 10^7 cells per mouse). Growth of solid tumors is monitored using in situ caliper measurements and models may be advanced stage or early stage (6, 61). Generally, activity is defined by tumor growth delay, optimal % T/C (T/C - median treated tumor mass/median control tumor mass) or net log cell kill. Drug-related deaths and body weight loss are used as parameters of toxicity.

Many compounds have shown promising activity in s.c. xenograft models, progressed to the clinic and revealed disappointing results. As a result, there has been considerable debate regarding the value of the xenograft model.

Despite enormous efforts to discover new chemotherapeutic drugs for treating the most common cancers, the conventional murine and xenograft test systems have identified only a limited number of useful agents that are clinically active at well-tolerated doses. Despite this, chemotherapeutic agents used routinely in the clinic today have played a significant role in reducing the mortality/morbidity, increasing the survival, and improving the quality of life of cancer patients. Chemotherapy may be given before, after, or in combination with surgery or radiotherapy. Indeed some agents have had a great beneficial impact on the survival of cancer patients, such as tamoxifen (UK-based pharmaceutical company ICI, now AstraZeneca, Macclesfield, United Kingdom) in the treatment of breast cancer. Moreover, testicular cancer has proven to be curable using chemotherapy (bleomycin, etoposide, and cisplatin).

Several groups have detailed studies supporting the value of the s.c. xenograft model for predicting clinical activity. Fiebig and colleagues (29, 62) established a large panel of xenografts...
derived from patient biopsies, and activity in xenografts was compared with clinical response. A correct prediction of clinical outcome was observed for both tumor resistance (97%) and tumor sensitivity (90%). These results also revealed that the xenograft model was more predictive of clinical activity than the clonogenic assay. Thus, it seems likely that well-characterized in vivo models which are more biologically representative of patient tumors may be more predictive of clinical response than uncharacterized in vivo models (29).

These results corroborate early studies comparing the activity of standard agents against xenograft models and clinical response (63). During the 1980s, several groups established disease-specific panels of xenografts from patient biopsies and correlated xenograft activity with patient response (61).

The NCI reported results of a retrospective study whereby 39 agents with both xenograft and phase II clinical trial results were compared (56). In vivo xenograft activity of a particular histology did not closely correlate with activity in the same human cancer histology. However, for compounds with activity in at least 1/3 of xenografts, there was a correlation with ultimate activity in at least some clinical trials.

Many of the agents tested in this study are standard chemotherapeutic agents used today (e.g., paclitaxel and doxorubicin; ref. 56).

With respect to the development of cytotoxic agents, the s.c. xenograft model does seem to have value as a predictive in vivo preclinical model (29, 56, 61, 63). However, the view that xenograft tumors (uncharacterized at the molecular level) are poorly predictive of the same histologic type of tumor in patients has influenced the initiation of the current strategy employed by the NCI’s drug discovery and development program (59).

Specifically, the in vitro 60-cell human tumor line screen is used to indicate possible mechanisms of drug action. Further development occurs only after an effort is made to have a molecular end point of the compound’s action, and defining a subset of tumors likely to respond to the agent that possesses the intended target. Pharmacologic scheduling and toxicology studies are done prior to target-directed phase I studies (59).

Many agents undergoing anticancer drug development at present are not cytotoxic agents but are small molecules rationally designed to inhibit fundamental processes known to be involved in the initiation and/or progression of human malignancy. Common target molecules include Bcr/Abl kinase in chronic myelogenous leukemia, c-kit in gastrointestinal stromal tumors, and erbB2/HER-2 oncogenic receptor tyrosine kinase in breast cancer. A comprehensive review of current molecular drug targets is provided elsewhere (64). The discovery of such molecular targets has been accelerated by recent technical innovations such as high throughput genomic/proteomic technology.

It is anticipated that the s.c. xenograft model will still be of value in this modern era of target-driven anticancer drug discovery if used appropriately. Specifically, it is becoming more and more appreciated that xenograft models should be characterized to ensure that the molecular drug target is expressed (46, 59, 61, 65, 66) and that xenograft studies should integrate both pharmacokinetic and pharmacodynamic investigation (34, 46, 59, 61, 67, 68).

The comparative analysis done by the NCI on clinical and xenograft activity found that only non–small cell lung xenografts were predictive of response in this histology in patients (56). In this study, xenografts were derived from cell lines that had been cultured over a long period, and it is likely that such xenografts no longer retained the original molecular characteristics of the patient tumor. In contrast Fichtner and colleagues (69) recently established human tumor xenografts directly from patient tumors that were characterized for specific molecular markers of the original patient tumor. The response to chemotherapy of these patients coincided very closely with the response of the individual xenografts. These data support previous work whereby fresh tumor explants were grown as xenografts and activity was found to show good correlation with clinical outcome (29). Together, these studies show that xenograft models which closely mimic the clinical situation are valuable models in predicting clinical outcome, and it is also emphasized that such molecularly characterized xenograft models may serve as valuable tools in the target-oriented drug development of specific rationally designed small molecules (69).

In line with this approach, a recent initiative by the NCI/ Cancer Treatment Evaluation Program and the Children’s Oncology Group aims to characterize available models of childhood tumors through proteomic and genomic profiling (Pediatric Oncology Preclinical Protein Tissue Array Project). These data will be of value in selecting xenografts for preclinical testing of molecular targeted therapies (68).

Retrospective analysis of pharmacokinetic and pharmacodynamic parameters in preclinical and clinical studies can often logically explain the failure of compounds in the clinic (68, 70). For instance, plasma concentrations required for antitumor activity in mice may be in excess of that achievable in man, and may even cause toxicity at lower doses in man. It has been suggested that the failure of conventional cytotoxic drugs in the clinic may be due to such inappropriate drug dosing (67).

Human tumor xenografts have been shown to be remarkably predictive of clinical cytotoxic therapy for a given type of cancer provided that clinically relevant pharmacokinetic parameters (i.e., dosing) are employed (67, 68, 70). For instance, Nomura, Inaba and colleagues at the Cancer Chemotherapy Centre, Japanese Foundation for Cancer Research, Tokyo, Japan, and Houghton and colleagues at St. Jude’s Children’s Hospital, Memphis, USA, gave clinically relevant or rational doses in preclinical s.c. xenograft models and found that the pattern of response in mice was similar to the activity in the respective human cancer with the same drugs (67). Such studies emphasize the need for the determination of exposure levels required to cause an antitumor effect in xenografts so that clinical trials may be coordinated accordingly and unnecessary toxicity avoided.

**ORTHOTOPIC AND METASTASIS TUMOR MODELS**

As described previously, compounds are usually screened against a panel of poorly characterized human tumor xenografts implanted s.c. in nude mice. S.c. tumor models do not represent the primary site of common human cancers or sites of metastasis (66). It has been suggested that the disparity between preclinical and clinical activity is related to the treatment of advanced...
metastatic disease in the clinic, whereas conventional s.c. xenograft models do not represent advanced metastatic disease nor is the orthotopic site represented (67). Orthotopic transplantation models attempt to mimic the morphology and growth characteristics of clinical disease (71–75) and are thought to represent a more clinically relevant tumor model with respect to tumor site and metastasis (76). One of the most obvious advantages of orthotopic systems is that attempts to target processes involved in local invasion (e.g., angiogenesis) can be carried out in a more clinically relevant site (66). Several other models are also used to assess the angiogenic properties of novel agents (e.g., corneal micropocket assay; ref. 77). Since the early studies showing orthotopic transplantation of colon tumors and metastasis to the liver (78), tumor material has been grown orthotopically in mice at most common sites of human cancer. Whether preclinical models representative of clinical disease (e.g., orthotopic/metastatic models) should be employed as a replacement for traditional s.c. nonmetastatic xenografts (67, 74) is an interesting question.

There are several key preclinical studies describing the disparity of activity in s.c. and orthotopic models in relation to the success of compounds in the clinic (66, 79–81). Although these studies provide convincing argument that orthotopic models may be the more appropriate model for the prediction of clinical response, this remains to be seen. Orthotopic models have also been observed to falsely predict clinical activity (e.g., batimastat; ref. 82). It has been suggested that further studies using current chemotherapeutic agents against s.c. and orthotopic models of common types of cancer may help confirm this (66).

Despite the obvious clinical relevance of orthotopic models, their application is hindered by several disadvantages. In contrast to conventional s.c. tumor xenografts, limitations include technical skill, time, and cost. Therapeutic efficacy is also more difficult to assess in contrast to the relative ease of s.c. tumor measurements (66).

The characterization of metastasis models may be evaluated more easily with the use of noninvasive microimaging research tools (66, 83, 84). Recent technological advances have had a major impact on research using orthotopic models for studying the process of cancer metastasis. Magnetic resonance imaging and positron emission tomography are being used to visualize tumor and metastasis progression. Reporter genes with specific fluorescence properties have been developed including the stable green fluorescent protein (83), galactosidase lacZ gene (85), and the luciferase gene (86). Although such technology is not widely used, it has permitted the visualization of tumor growth using ethically more acceptable noninvasive techniques and has reduced the numbers of animals required for orthotopic and metastasis studies.

**AUTOCHTHONOUS MODELS**

Autochthonous tumors include spontaneously occurring tumors and induced tumor growth (e.g. by chemical, viral, or physical carcinogens). It is thought that autochthonous tumors may mimic human tumors more closely than transplanted tumors (i.e., s.c./orthotopic). Advantageous properties include orthotopical growth, tumor histology devoid of changes introduced by transplantation, and a route of metastasis through lymph and blood vessels that surrounded early tumor growth (87).

Despite such properties, the use of autochthonous tumor models has not been widespread due to several limitations. A large variability in take rate and growth exists, the large number of animals needed, time frames of several months to years exist for a single experiment due to long tumor latencies as opposed to weeks in transplanted xenograft models, and lack of spontaneous metastasis (84, 87). Autochthonous models are used occasionally as tools for advanced or phase II screening (87), but more recently in this “postgenome era” autochthonous models have largely been replaced by genetically engineered mouse (GEM) models.

**GENETICALLY ENGINEERED CANCER MODELS**

Over the past 20 years, GEM models have made a significant contribution to the field of cancer research. GEM models have increased our understanding of the molecular pathways responsible for the initiation and progression of human cancer, and have highlighted the importance of specific oncogenes and tumor suppressor genes (TSG) in particular types of cancer.

GEM models possess well-validated molecular/genetic characteristics (e.g., gene mutations) which ultimately facilitate the rational design of small molecule therapeutics (88).

The main aim of GEM models is to recapitulate genetic/molecular changes in human cancer and use these to test novel anticancer therapeutics in an attempt to accurately predict clinical response.

The first strains of genetically engineered mice predisposed to cancer were transgenic mouse models whereby cellular/viral oncogenes were introduced to the mouse germ line. One of the first transgenic cancer models involved the constitutive expression of the c-myc oncogene under the control of the mouse mammary tumor virus promoter leading to the development of mammary tumors (89). Many transgenic experiments have followed and clearly shown that the manipulation of the mouse germ line could predispose the mice to cancer (90).

Upon the discovery that the progression to a malignant phenotype often involves the loss of TSG function (91), transgenic mouse models were developed which involved introducing a mutant TSG to the mouse germ line. TSG function can be impaired by either targeted gene knockout (92) or the transgenic expression of a dominant-negative form of the TSG (93). One of the first TSG mutants was the Rb “knockout” mouse (94). Mice heterozygous for a null Rb allele developed tumors (pituitary adenocarcinomas, medullary thyroid carcinoma, and/or phaeochromocytomas). Since the Rb knockout, many mutant TSG or knockout cancer-prone mouse models have been developed including p53 (94–96), Apc (97), Nf1 (98), and many more reviewed elsewhere (88, 99–101).

Alternatively, mouse models have been developed by inducing germ line mutagenesis. For instance the Min (multiple intestinal neoplasia) mouse was created by germ line mutagenesis using N-ethyl-N-nitrosourea, which causes a point mutation in the Apc TSG (102). The most common human cancer, basal cell carcinoma, has also been modeled by the exposure of Pich
heterozygous mice to UV light (103). Etiologic factors such as diet are also being modified and have been shown to affect tumor development in GEM models (104, 105).

It is well known that cancer is a multistep process involving several sequential mutations (106), and in the light of this finding, different mouse strains of known genetic background have been interbred in an attempt to identify any cooperativity between oncogenes and TSGs in the progression towards the malignant phenotype (88). Sinn and colleagues (107) clearly showed that when two transgenic mouse strains individually expressing c-myc and v-ras were crossed, resultant progeny developed mammary gland carcinoma at a much greater rate (synergistically) than transgenic mice expressing only one oncogene. This study clearly indicated that c-myc cooperated with v-ras in the progression towards malignancy in mammary gland tumorigenesis in vivo.

There are several existing mouse models of multistep tumorigenesis. The RIP-Tag mouse expresses the SV40 antigen under the control of the insulin promoter leading to the development of pancreatic islet cell carcinoma (108). The analysis of the genetic/histologic changes occurring during tumor progression in the RIP-Tag model has identified specific molecular markers associated with apoptosis, angiogenesis, and cell adhesion (e.g., Bcl-XL and E-cadherin) at specific stages (88, 109–113).

Transgenic and knockout approaches can also be used to evaluate the role of specific components of the tumor microenvironment (e.g., matrix metalloproteinase 9) in tumor progression. The chemically induced skin carcinoma model (114) is another well-explored model that mimics more than one stage of the progression towards malignancy in mammary tumor progression with defined genetic/molecular characteristics [H-ras activation, up-regulation of cyclin D1, loss of p53, and up-regulation of transforming growth factor-β1 (88)].

Transgenic and knockout mouse models involving manipulation of the mouse germ line are often limited by embryonic lethality. In an attempt to overcome this, several new approaches to create GEM strains have been introduced, including transient conditional gene targeting, latent oncogene expression and the use of avian sarcoma leucosis viruses, and chromosome engineering (88, 115–118). These novel methods have been facilitated by recent technological advances [e.g., cytogenetic (spectral karyotyping), genomic comparative genomic hybridization, comparative genomic hybridization microarrays, and gene expression profiling; refs. (119, 120)].

Conditional transgenic/knockout models involve spatial control over the initiation of oncogene expression and TSG inactivation, respectively, and have been used to create models of several types of cancer (115). The Cre-Lox system is the most widely used for both transient conditional knockout (121) and oncogene expression (122). Another new approach in creating mouse models involves latent mutant alleles that become expressed following somatic recombination in vivo (123). Inducible oncogene expression involves the tissue-specific expression of oncogenes in response to stimulation by small molecules (e.g., doxycycline and tamoxifen; ref. 115). Avian sarcoma leucosis viruses are also employed to deliver oncogenes and dominant-negative forms of TSGs to cells in vivo that express the avian sarcoma leucosis virus retroviral receptor (124).

 Unlike xenograft models, GEM models possess well-validated drug targets and may potentially offer a more appropriate preclinical model in which to test modern small molecule therapeutics. Additionally, GEM tumors develop autochthonously/in situ and therefore may be more biologically representative of a particular tumor type in humans than transplanted xenografts. Despite such promise, GEM models are not without limitation. Compared to the traditional xenograft model, GEM models are expensive and time-consuming. Their use is often restricted by intellectual property rights and patents (125). In addition to embryonic lethality, mice often do not develop the expected tumor type as they may die prematurely from a different tumor type caused by the constitutive expression of oncogene/TSG. Species-specific differences also exist in the role of different genes in different cell types, which can lead to different mutant phenotypes in both man and mouse (126).

It is not usually a primary tumor which kills a patient, but metastatic disease, and unfortunately this advanced stage is not represented by most GEM models (101, 116). In comparison to traditional xenograft models, very few studies have shown the therapeutically efficacy of anticancer agents using GEM models.

Small molecule inhibitors have been used to target farnesyl transferase, epidermal growth factor receptor, and FLT3, using GEM models and have been predominantly shown to block tumor development or regress established malignancy (116, 128–130). RIP-Tag (pancreatic) and TRAMP (prostate) tumor progression models have been used to test the efficacy of angiogenesis/matrix metalloproteinase inhibitors (116, 131, 132), and vascular endothelial growth factor receptor inhibitors, respectively (133).

Additionally, very few studies have tested known clinically effective agents using GEM models (134–136). Such GEM studies using mice (whereby the equivalent mutation found in human malignancy is validated and therapeutic efficacy is observed) provides optimism that GEM models may indeed be of value in predicting clinical response. Despite such promise, the value of GEM mouse models in anticancer drug discovery is yet to be determined. Only time will tell if GEM models will be any better at predicting clinical activity than currently used xenograft models.

Compared to xenograft models, relatively few studies have documented the use of orthotopic (66, 79–81), transgenic (134–136), and autochthonous models (87) in cancer therapy and moreover in predicting clinical response. This is largely due to the fact that there are relatively few laboratories using these test systems in drug development programs. Additionally, unlike xenograft models (56), there are no studies collectively analyzing large bodies of data from these models, and therefore no definite conclusions can be made and the relative predictive value of orthotopic/autochthonous/transgenic models remains largely speculative. In order to assess the predictive value of these models, preclinical studies testing currently used chemotherapeutic agents are required.
With time it is anticipated that the clinical use of small molecule therapeutics will outweigh the use of classic cytotoxic agents. As the identification of specific pathways driving the development of human cancer increases, it is imperative that transgenic models are developed that truly reflect clinical disease. New drug molecules will have been identified through a rational preclinical cascade culminating in the demonstration of in vivo proof of principle of efficacy, in appropriate preclinical mouse models. It is essential for the validation of transgenic models that such small molecule therapeutics are evaluated in patients that have been appropriately selected on the basis of the expression of the molecular target.

SUMMARY

As we have discovered more about the underlying mechanisms responsible for the initiation and progression of human cancers, we have experienced a move away from the development of classic cytotoxic agents to the rational design of small molecule anticancer therapeutics. This has prompted a transition from empirical compound-oriented preclinical screening to target-oriented drug screening. The use of uncharacterized tumor models (e.g., xenografts/syngeneic models) has been continuously replaced by more clinically relevant and molecularly characterized models along with the integration of pharmacodynamic and pharmacokinetic approaches.

The value of any preclinical tumor model will ultimately depend on its ability to accurately predict clinical response. In this modern era of target-driven anticancer drug discovery, we believe that the full potential of any tumor model can only be met if it is used appropriately, that is, if it is fully characterized to ensure that the molecular target of interest is expressed and that the model is used to confirm drug-target interaction. In addition to determining quantitative antitumor activity, it is believed that preclinical tumor models should be used to gain a broad range of information (i.e., pharmacokinetics/metabolism/pharmacodynamics). It is emphasized that this preclinical information needs to be used appropriately in clinical trials. In particular, if a novel target-directed agent is to be used to treat a patient, the presence of its respective target must initially be confirmed.

If preclinical models are used routinely to this extent and a closer relationship exists between the clinician and the laboratory scientist, it is anticipated that the clinical use of small molecule therapeutics will improve on the present clinical reality of classic cytotoxic chemotherapy identified using traditional uncharacterized models.

REFERENCES


49. Suggitt M, Bradshaw TD, Stevens MF, Bibby MC. 2-(4-amino-3-methylphenyl)benzothiazole prodrug (“Phortress”) causes selective DNA damage in breast carcinoma cells cultured in the in vivo hollow fibre assay. AACR-NCI-EORTC international Conference, Molecular Targets and Cancer Therapeutics [abstract]. 2003 Nov 17–21; Boston, MA, USA.


50 Years of Preclinical Anticancer Drug Screening: Empirical to Target-Driven Approaches

Marie Suggitt and Michael C. Bibby


Updated version  Access the most recent version of this article at: http://clincancerres.aacrjournals.org/content/11/3/971

Cited articles  This article cites 103 articles, 34 of which you can access for free at: http://clincancerres.aacrjournals.org/content/11/3/971.full.html#ref-list-1

Citing articles  This article has been cited by 20 HighWire-hosted articles. Access the articles at: /content/11/3/971.full.html#related-urls

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.