The Drug Development Crisis: Efficiency and Safety

C. Thomas Caskey

Brown Foundation Institute of Molecular Medicine for the Prevention of Human Diseases, The University of Texas Health Science Center at Houston, Houston, Texas 77030; email: C.Thomas.Caskey@uth.tmc.edu

Annu. Rev. Med. 2007. 58:1–16

First published online as a Review in Advance on October 20, 2006

The *Annual Review of Medicine* is online at http://med.annualreviews.org

This article's doi: 10.1146/annurev.med.58.042705.124037

0066-4219/07/0218-0001$20.00

Key Words

targets, screening, antibodies, NCEs, phase IV

Abstract

Despite advancements in genetics, chemistry, and protein engineering, recent years have seen fewer approvals of new drugs, increases in development costs, and high-profile drug withdrawals. This article focuses on technologic methods for improving drug development efficiency. These technologies include high-content cell screening, expression profiling, mass spectroscopy, mouse models of disease, and a post-launch screening program that enables investigations of adverse drug effects. Implementation of these new technologies promises to improve performance in drug development and safety.

INTRODUCTION

Drug discovery, approval, safety, and marketing are a current focus of attention by academic, industrial, and legislative leaders. At a time when the promise for new therapeutics for untreated or poorly treated diseases is at its highest, public opinion of the drug industry's performance is low. This article identifies choke points in drug development where innovation offers solutions.

The development of recombinant DNA technology (1), the polymerase chain reaction (2), high-throughput DNA sequencing, and related molecular biology techniques have enabled gene-specific drug development. This represents a major paradigm shift from older drug development strategies, which utilized whole-animal physiology-based testing (3).

The Human Genome Project (HGP), driven by both industrial and government sponsors, reigns highest in scientific impact (4–6). The HGP and related genome sequencing projects have provided sequence information for all genes in man, nonhuman primates, mouse, rat, *Drosophila*, yeast, bacteria, and other organisms. The previous trickle of genetic information was replaced by a fire hose for the study of gene function and disease causation. When human genetic mapping linked a heritable disease to a gene, that gene quickly became a target for gene-related therapeutics. Examples of diseases in which human genetic disease discoveries spawned drug development efforts include Duchenne muscular dystrophy, cystic fibrosis, myotonic dystrophy, Alzheimer's disease, breast and colon cancer, and coronary artery disease (7).

The advent of a wide variety of research tools have enabled the functionation of human genes in the wake of genome sequencing. A plethora of bioinformatics algorithms now exist for gene function predictions. Mouse genetics has profoundly influenced these new approaches because of the ability to genetically manipulate the laboratory mouse (through both gene knockout and transgenic approaches) to study disease mechanisms and therapeutic agents (8). Developments in combinatorial chemistry and robotic high-throughput screening allow screening of large chemical compound libraries against genetic targets. In parallel with and enabled by the DNA technologies, novel therapeutics emerged that expanded the opportunities for disease intervention. These new products included recombinantly produced vaccines against hepatitis B, human papillomavirus, and HIV; recombinant reproduction of proteins (growth hormone, insulin, erythropoietin, and granulocyte stimulating factor); and the development of monoclonal antibodies (CD20, herceptin, anti-TNF, and integrins) (9).

In the United States, this convergence stimulated the biotechnology industry, which focused on novel early-stage therapeutics. Biotech companies were distinguished from pharmaceutical companies, which excel in development of therapeutics. The biotech industry now accounts for ~35% of new drug candidates (10).

How could one not be excited by these developments and promise of transformation of an industry? Yet there are significant challenges in harnessing these promising elements. The flat rate of drug approvals by the U.S. Food and Drug Administration (FDA) is shown in **Figure 1**. Estimates using the historic costs of drug development forecast 70–80 approvals per year at this level of biopharma expenditure. However, drug development costs have risen to an estimated average of $800,000,000 per approved drug (e.g., $1 billion for Taxol and $250 million for human growth hormone), and the development timeline has stretched to 10–15 years (11). The high cost is predominantly due to failed drug initiatives (12). The FDA reports that only 8% of phase I trial drugs are ultimately approved. In addition to this cost escalation, 4% of approved drugs are withdrawn (including Bextra, Vioxx, Baycol, Rezulin, and Tysabri; **http://www.fda.gov**) leading to lost income and high-profile and expensive product litigation.

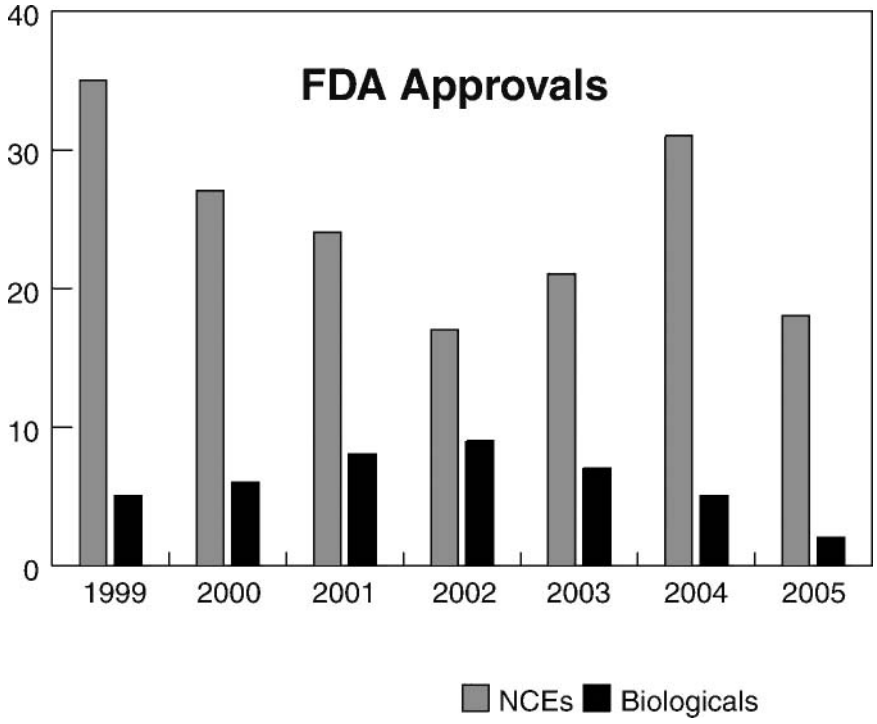

Figure 1

In reviewing the trends over the past five years for FDA approvals, the Institute of Medicine study noted that the number of new chemical entities (NCEs) approved in 2005, namely 29, was the lowest number in 10 years. This however could easily change, as IMS noted growth in the actual numbers of compounds known to be in phase 1, phase 2, and preclinical study. It was also highlighted that the U.S. prescription growth (year-on-year sales growth) for 2005 was positive at 5.4%. Although this has tended toward slower growth over the past five years (down from 11%), it has been shown to be cyclical as longer periods of time are examined. Further, IMS noted that this growth would have been roughly 6.9% were it not for the withdrawal of Vioxx and Bextra and the impact these withdrawals had on Celebrex. (Source: Pink Sheet, January 2006, Lehman Brothers.)

In the following sections, suggestions are put forward to solve these challenges by using new technologies. Little attention is given to the regulatory challenges except where new technology interdigitates.

TARGET SELECTION

It is well established that incorrect target selection accounts for the failures of some drug candidates. Experience from the biopharmaceutical industry indicates that currently only 5% of new targets will lead to FDA-approved products (13). Thus, innovative approaches to identify a "validated" target would boost productivity. The HGP has identified 20,000–25,000 genes as the repertoire for target selection (5, 14). Several groups have used bioinformatic methods to hypothesize that 5000–6000 genes (15) may be drug targets. This estimate is independent of disease and pathway understanding and is based purely on putative gene function.

The challenge is to match gene with disease and validate the target for drug development. Three industrial approaches to this challenge are gaining acceptance: human genetic association, mouse models, and RNA profiling/RNAi technologies. Man is certainly the best-studied mammal, so abundant

disease phenotyping is available via the medical community, which describes disease using clinical and laboratory methods. New genetic mapping methods, including whole-genome association, promise to link diseases and genes more efficiently (16).

The rate of discovery of human disease–gene associations has increased to ~30 per month. Examples of successful new drug development based on such genetic information include Proscar, ReoPro, and Gleevec (9). Proscar is an inhibitor of 5α reductase-2. This enzyme was found to be deficient in patients with male hypospadius (17) and small prostates, thus validating it as a target for a drug that treats benign prostatic hyperplasia. Glanzmann thrombasthenia, a heritable platelet dysfunction disease (18), was found to be due to an inactivating mutation of integrin structure. A monoclonal antibody and small molecule antagonists of glycoprotein GPIIb/IIIa were developed, which mimicked the gene mutation effect and thus function as anticoagulants (9, 19, 20). Chronic myelogenous leukemia is associated with the presence of the Philadelphia chromosome (20), a de novo translocation mutation, found by molecular methods to result in a junctional Bcr-Abl kinase (21). Gleevec (22) was developed as a specific inhibitor of this disease-causative mutant kinase. All three drugs were remarkably successful (9). Human gene associations for obesity, Alzheimer's disease, Parkinson's disease, diabetes, and LDL regulation have also validated new targets previously unknown.

The mouse is the leading mammalian model system for target validation because it is genetically similar to man and can be genetically manipulated, selectively bred, and challenged by diet, infectious agents, and drug candidates. Lexicon Genetics (15) has been a leader in the use of mouse models of human genetic disease by industrializing the processes of gene knockout production (random insertion and targeted gene mutation) and medical diagnosis of the mutant mice. The concept is simple and elegant: The gene knockout mouse can be studied to identify targets for drug development and concomitantly identify unwanted target-related effects. The new "clean" drug ideally should match the genetic knockout phenotype to be free of adverse effects. For example, a "clean" knockout mutant used in osteoporosis research has increased bone density only (**Figure 2**). Because the loss of gene function leads to increased bone density, a drug that inhibits the gene function can be proposed as an osteoporosis therapeutic. If new effects are found with the drug, they are considered "off-target" effects, and variants of the compound are made via medicinal chemistry until the knockout and drug effect match. The goal is a gene target–specific drug with no off-target effects. A recent publication has validated the strategy for the 100 best-selling drugs on the market (23). Lexicon pursues their 5000 drug candidate targets by bioinformatic selection of target genes coupled with extensive medical testing (phenotyping) of each knockout. The not-for-profit Texas Institute for Genomic Medicine now makes the Lexicon mice available for research. The Jackson Laboratory is another source of research mice.

Another example of the success of this approach is illustrated by the discovery of leptin as a cause of obesity in the ob/ob mouse. Subsequently, recombinantly produced leptin was clinically successful in treating rare leptin-deficient obese human children (24). Additional knockout mouse models of obesity have identified new gene targets and fostered the development of new obesity drugs. It is clear that the mouse knockouts give knowledge of new, validated drug targets and also identify "dirty" targets—genes whose suppression will cause complications. The use of knockout mice is routine in virtually all drug development today.

RNA interference (RNAi) has had a major impact on target validation using cultured cells (25). Short, sequence-specific, structured nucleotides (siRNAs or shRNAs) complementary to target gene transcripts trigger mRNA degradation and/or translational arrest, resulting in a decrease in target gene

expression. Commercially available sets of RNAi reagents directed against ~5500 genes in the "drugable genome" are widely available. Using a combination of three siRNA/shRNAs for a specific mRNA, the specificity is reported to exceed 90% for cellular knockouts. One can scan efficiently and specifically for these effects using high-throughput screening. The first of these RNAi-validated targets are now in drug development. The technology benefits from speed, flexibility, and specificity.

Many have placed high value on comparing mRNA expression levels in diseased and control tissues to validate targets. The mRNA levels can be quantified using microarrays (26) and/or quantitative PCR (27). Although it lacks the disease-gene-identification specificity of the above-mentioned methods, this technique can identify disease-gene candidates that can be validated by independent methods. Affymetrix, Illumina, and others have developed commercial gene-array products. A recent example of the power of this approach is the identification of a gene key to metabolic energy source shifts during hibernation, and lipid regulation in humans, by comparison of liver gene expression in circadian rhythm mouse mutants (28). A limitation of this approach is the difficulty of distinguishing causative from correlative ("bystander") transript changes. For example, mRNA expression profiling of a mouse with a single gene knocked out resulted in 20–200 significant quantitative mRNA changes in addition to the absence of the knocked-out gene's mRNA (T. Caskey, unpublished data).

PRECLINICAL SAFETY

Many drug toxicities are only recognized at the clinical trial stage, after enormous resources have been invested in the drug candidate. The objective of new technologies is to provide improved predictions of human drug safety in the preclinical development phase.

The previous section discussed the utility of the genetically modified mouse for target identification. Mouse models of disease also offer the opportunity for preclinical safety and drug efficacy measurements. Curative and undesirable effects of a new compound can be examined by therapy testing in the engineered mouse model of that disease. Both safety and efficacy can be determined in an engineered mouse model prior to study in man.

Ideally, drugs should be potent and specific. Preclinical drug candidate screens against "like" targets are useful for minimizing the risk of off-target effects. For example, new drug candidates in the class of G protein–coupled receptors (GPCR) can be screened against a panel of like receptors for activity that determines in vitro specificity. More recently, kinase panels have been developed which allow in vitro screening for specificity. Such screening is critical for safety as regulation of cell growth and apoptotic pathways is targeted for cancer and immune modulation.

Technology now enables evaluation in living cells, so unanticipated protein complex interactions are taken into account. Automated microscopic scanning delineates both the localization and translocation of proteins. Tsien and colleagues (30) recently reviewed a wide variety of fluorescent cell-signaling strategies. These methods are referred to as high information content screening (HICS) because they measure both quantitative and qualitative cellular events and intracellular sites of drug action. This detection technology, when combined with the general pathway detection method of protein-fragment complementation assay (PCA) (31), enables survey of multistep pathways. PCA is now coupled to a high-throughput screening format, thus providing economical data acquisition (**Figure 3**). This technology in its current state of development permits evaluation of a single drug candidate for 2500 cellular events in 250 pathways, representing a true "systems biology" approach. Thus, in a single HICS cellular study, one may generate data on the specificity and projected safety of a drug candidate preclinically.

It is traditional in the industry to have several structural series of a drug candidate. One can select the candidate with the fewest and mildest off-target effects in order to minimize clinical adverse reactions. The pathway commonality of multiple approved drugs in a target class, as well as distinctive aberrant activities of individual drugs such as statins and COX inhibitors are readily identified by PCA/HICS. As expected, the approved drugs with the smallest off-target effects have enjoyed the widest safety margins in the clinic. A schematic representation of pathway safety analysis is shown in **Figure 4**. The technology is being applied to kinase inhibitors in cancer drug candidates with the objective of identifying a compound with minimal off-target effects and high potency (31).

A second strategy utilizes the measurement of expression for large numbers of genes and drugs. Affymetrix, Illumina, and Applied Biosystems' quantitative PCR are all used. The FDA now encourages expression data sharing among biotechnology and pharmaceutical companies to improve preclinical safety. The broad acceptance of gene expression technology by the biotechnology and pharmaceutical industries should accelerate its utility for preclinical safety applications. There are already numerous examples of expression profiling being applied to the clinical classification of lymphomas (32), breast cancer, prostate cancer, and other cancers. It is reasonable to expect drug response profiling will be carried out in readily accessible cancer tissues such as leukemias and solid tumor biopsies (33). Thus, profiling may improve the safety and specificity of drugs.

Mass spectroscopy (34) is a powerful scanning tool applicable to proteins, peptides, hormones, and small-molecular-weight compounds in tissue or sera. Matching mass of specific biologic fractions with reference compounds and genomic predicted mass of molecules identifies both the molecule and its quantity. Given the wide genetic diversity of man and the genetic heterogeneity of diseases (both mechanism and pathology), mass spectroscopy would appear to have a technical advantage over other methods because it measures specific molecules in specific patients at differing times of pathology. In one example of mass spectroscopy's scope of detection (**http://www.metabolon.com**), 700 biologic chemical compounds are measured from sera in a quantitative manner, thus exceeding by 30–50-fold current compound surveys (SMA-20 and multiplex ELISA). Using this technology to measure chemical analytes, HIV protease inhibitors with high and low incidence of adverse side effects were identified.

Serum fractionation methods enable the quantitation of proteins and peptides (36). Improved fractionation protocols are needed to access abundant and rare proteins. Advantages of fractionation include a wide range of sensitivity and clear visualization of low-abundance proteins and peptides ("bioactive" regulatory molecules) whose abundance fluctuates in health and disease. Many have achieved reproducible measurement of multiple proteins and peptides by employing computational algorithms (37). In an early report of mass spectroscopy success, drug-induced liver injury (38) was detected earlier and with more sensitivity than with serum enzyme measurements. Measurement of proteins by mass spectroscopy has also been successful in cancer diagnostics. The proteins CA-125 and PSA (prostate-specific antigen) have been in clinical use for early detection of ovarian cancer and prostate cancer, respectively, and using ELISA methods to detect these proteins is a standard of practice. Mass spectroscopy protein profiling for ovarian cancer (37) and prostate cancer (37) is now entering clinical practice using a constellation of protein quantitations. Mass spectroscopy is also applicable to the early detection of hepatic, renal, and cardiac injury in both laboratory animals and humans (38). Given the broad scanning capacity of mass spectroscopy, it is likely to be a major discovery tool for development of more specific protein/peptide assays of drug toxicity for in vivo screening. Mass spectroscopy is widely used in drug development today for

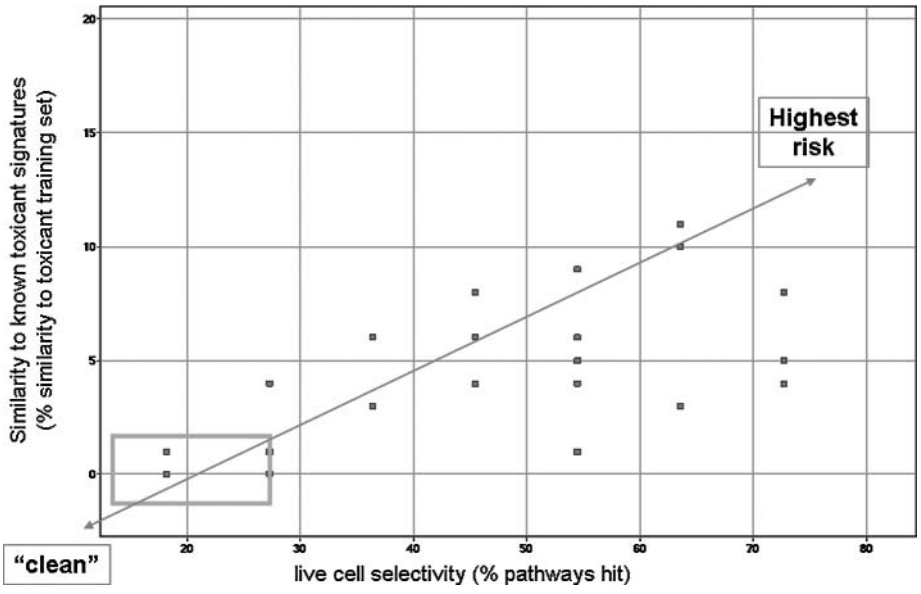

Figure 4

Multidimensional profiles enable identification of the safest leads. Source: Odyssey Thera (John Westwick, CEO).

the study of drug metabolism in laboratory animals and humans where variation is known to influence safety. This variation can be genetically based, influenced by stage of disease and drug interactions. Individuals' differential responses to codeine and wafarin, which have a clear genetic basis, illustrate the challenge (39). Mass spectroscopy can provide data on individual patient drug pharmacokinetics and possibly identify early those susceptible to adverse reactions. It is not unreasonable to incorporate mass spectroscopy analysis in all first-in-man (FIM), phase I, and phase II studies for safety.

FIRST-IN-MAN AND PHASE I SAFETY

The need for caution and improved safety measures is increasing as new genetically identified targets and new therapeutic modalities, such as aptamers (40), RNAi (25), protein activators, and mAbs are brought to the clinic. Although the promise of new drugs is encouraging, detection of unexpected adverse clinical effects must be carefully planned.

The recently approved P450 allele diagnostic kit (Abbott) (41) is an example of a new entry in this arena. It is well known that certain P450s have genetic variations in the population that either extend or shorten the half-life of drugs, potentially enhancing toxicity or efficacy. Therefore, with any new drug known to be metabolized by these pathways, patient screening for mutant alleles affecting pharmacokinetics seem prudent in FIM and phase I studies. Methods of P450 drug induction that can be used routinely and reliably are still needed. As pointed out in the previous section, mass spectroscopy can determine the variation in pharmacokinetics on a patient-by-patient basis. Both prescreening of patients and actual testing of drug levels should be standards of practice in early phase testing. It is premature to identify the biomarkers evolving out of research initiatives that more accurately evaluate drug efficacy. These will, however, evolve rapidly and should be rapidly adapted as surrogate markers of efficacy and injury. Following discovery, assay methods that are highly validated for specificity, sensitivity, and economy can be developed.

Many of the above comments have addressed NCE safety, but we have had a recent wake-up call regarding biologic preclinical

safety. Dramatic adverse outcomes of a FIM clinical study of a CD28 binding protein (a pivotal receptor in immune response) were recently reported (42). Within minutes of first human administration, six patients were *in extremis* with multiple organ injury—a totally unexpected outcome based on preclinical safety studies and product quality assurances. Given the mechanism of a CD28 activator, it is speculated a cytokine "storm" was induced. This event will undoubtedly lead to new safety regulations and FIM trial revisions once the mechanism is fully understood. For example, quality control is a bigger issue for biologics than for NCEs because they are produced cellularly and must be purified of cellular contaminants, correctly posttranslationally modified, structurally folded, and stabilized for storage.

Why must the safety concerns and methods for biologics differ so markedly from those for NCEs? It is well known that NCE specificity and potency reside in small changes in the molecules. Furthermore, because binding interaction sites can be small, cross reactivity of other proteins—i.e., off-target effects—can occur. Biologics differ significantly in that they gain high specificity by the summation of scattered contact sites to the target. This specificity is the beauty of biologics for use in humans and also the challenge for safety determination. Comparative genomic sequence information on humans, rats, mice, and nonhuman primates is helpful (43). Although mammals share largely similar sequences of specific genes, we differ in the details of gene structure and very much in their regulation. These differences should affect our safety evaluation methods. For example, to determine that a mAb has no adverse effect on a monkey is useless if the tested mAb is not signaling as it would in a human. Changes in the preclinical safety measurements need to accommodate differences at the in vitro cell signaling level. Mouse gene knock-in models (mice with a human target gene) should be considered to determine the most effective and reproducible safety strategy. A possible solution to this challenge may be the in vitro use of human whole blood samples, where mAb signaling could be quantitated prior to in vivo safety trials (44). The mAbs in development today in both the biotechnology and pharmaceutical industries will be markedly delayed in their approval unless we resolve this critical chokepoint in safety.

A second unexpected complication of biologics is illustrated by two therapeutics, erythropoietin (9) and thrombopoietin (45). In the case of erythropoietin, some patients were found to develop antibodies to recombinant and endogenously produced erythropoietin, resulting in profound suppression of erythropoiesis (46). The drug that caused the adverse outcome was traced to a manufacturing site. It was suspected that the recombinantly produced erythropoietin was immunogenic because of incorrect folding or aggregation, and production was corrected. In the case of thrombopoietin, the problem has not been corrected thus far. Thrombopoietin was found to also stimulate antibody production and thus exacerbate thrombocytopenia (47) in some patients. Using sensitive antibody detection methods, patients were found to have preexisting antibodies to endogenous thrombopoietin. Exogenous thrombopoietin amplified the antibody titers and enhanced platelet depression with therapy.

These two examples have the common thread of endogenous antibody development provoking a significant adverse outcome. The problem is an old one and has been studied extensively for insulin (48). Many patients have antibodies to insulin but continue to benefit from insulin administration. Some need immune suppression to receive insulin because their antibodies affect insulin utility. How do we detect the emergence of antibodies that impact the utility of recombinant protein and its equivalent endogenous protein? We may need to call upon the technologies emerging from vaccine development research programs that detect immune response genes in blood specimens early in immunization (**http://www.canvac.ca**). The

greatest success to date is in multiplex cytokine quantitation. With such technology one can detect immunologic response to the therapeutic, either preclinically in nonhuman primates or in FIM or phase I studies. Furthermore, detection of an antibody response that blocks exogenous (recombinant product) and endogenous ligand levels is needed. Both will require research efforts in cooperation with these emerging technologies to identify a reliable detection format.

PHASES II AND III

Diabetes mellitus II, atherosclerosis, cancer, and dementia have multiple genetic bases. Such diseases are likely to require specific therapeutics discovered through pharmacogenomics, or perhaps cocktails of drugs for individual patients. It is rapidly becoming feasible to match a therapeutic to a genetic marker of disease in a patient. This advance has been facilitated by the technology of PCR and a plethora of nucleic acid detection methods now used clinically.

On the leading edge of this transition was the development of AIDS therapeutics (50). Investigators could rapidly determine the efficacy of a new therapeutic (blockers of protease, reverse transcriptase, integrase, or receptor-mediated uptake) by simply quantitating viral replication inhibition in vivo by measuring serum viral titers. As a result, the HIV viral titer rapidly changed from a biomarker to a surrogate endpoint measurement. In post-launch application of these drugs, it is common to measure drug resistance in individual patients before shifting them to new drugs.

The stunningly efficient development of Gleevec, the Bcr-Abl kinase inhibitor, for chronic myelogenous leukemia (CML), was enabled by genetic diagnosis prior to drug trial (51). Only patients with the genetic diagnosis of CML caused by the Bcr-Abl mutation were selected. Had the broad category of leukemia been the criterion, drug utility would have been diluted and possibly missed. In the post-launch application of Gleevec, drug resistance emerged and was explained at a molecular level via the identification of Bcr-Abl variants with a mutant Gleevec binding site. Target mutations must therefore be rapidly detected and therapy shifted to new agents (52) as resistance emerges.

In a second cancer example, the gene amplification of the Her2 growth receptor was used to select subjects for a mAb inhibitor trial (53). Without this genetic identification of the trial group, the clinical endpoints would have been missed and thus an effective therapeutic lost. These experiences and the encouragement of the FDA to make greater use of biomarkers and patient genetic identification have been outlined in the FDA's Critical Path document (54). The pharmaceutical industry is increasingly incorporating these strategies into clinical trials, thereby reducing costs and strengthening the likelihood of success. The opportunities for genetically matching patients to drug trials are growing as disease genes and "risk" genes are discovered in mice and humans.

Imaging as a biomarker or surrogate endpoint has played an increasingly important role in clinical trials. Investigators of central nervous system (CNS) disease have been particularly aware of the need to demonstrate brain penetration and on-target localization of the drug candidate. Positron-emission tomography (PET) ligand binding has been most successful in estimating CNS penetration and dosage requirements (55). PET has recently become easier with the availability of improved chemistry for synthesis of specific reagents. The application of PET scanning for utilization of fluorodeoxyglucose (FDG) in CNS disease and cancers has received widening acceptance (56). The diminishment of glucose uptake by a tumor in response to a therapeutic is now under study as a biomarker of response and is supported by Centers for Medicare and Medicaid Services (CMS) payment. The cancer trials monitor tumor size by CT, MRI, and radiographs, and metabolism via PET. Cognition, vision, and hearing are

also studied by FDG metabolism PET imaging, and may be applied to the development of drugs for cognition and neurodegenerative disorders (56).

Ongoing substantial research with in vivo nanoparticle imaging has the potential for highly specific imaging of disease states and gene products using mAbs. This could be a paradigm shift in improving in vivo biomarkers for more efficient trial outcome measurements.

PHASE III AND LAUNCH

The medical-value endpoint of a successful drug trial will be increasingly important in the future. Health economists are anticipated to have an increasing impact on new drug usage and formulary acceptance. CMS is now studying the concept of "pay for performance" related to medical professional care. Risk can be mitigated in such circumstances by matching patients to therapies using genetic and other biomarkers, as well as physician adherence to standard of practice. Advocacy groups can also play a role in this arena by asking practical questions on patient selection, biomarkers, and cost/risk/benefit considerations, not viewing these as mechanisms to withhold therapeutics, but rather ways to improve the success of new drugs.

PHASE IV

The recent examples of drug withdrawals with substantial litigation can be divided into three categories: (*a*) totally unexpected specific complication, (*b*) the subtle elevation of risk for a common disease in the population not related to the drug's indicated usage, and (*c*) drug interactions with existing therapeutics likely to be used in combination therapy.

The first category is well-illustrated by Tysabri, a mAb used in multiple sclerosis. It has remarkably positive results in multiple sclerosis but, at a frequency of 1 in 1000, causes severe encephalopathy (57). The risk/benefit of this therapeutic had to be weighed in a context where therapeutic options are limited and disease progression severe—not unlike cancer chemotherapy. The latest FDA decision returned Tysabri to the market with changes in risk description. Similarly, interleukin-2 gene therapy for X-linked Severe Combined Immunodeficiency (X-SCID) resulted in improvement in symptoms but caused leukemia (58) in 3 of 9 patients treated, apparently due to integration of the virus vector into a growth-regulating area of the genome. Use of this therapy has been suspended in most centers, but resumption of its use must balance its complication rate against the efficacy and complications of the alternative treatment, which is bone marrow transplantation. Given the knowledge of the mechanism of vector-induced leukemia, a logical means of improving the vector to avoid leukemia is under development. In the vaccine arena, the makers of the recently approved rotavirus vaccine (59) were required to study 10,000 patients for intussusception (1 in 1000 incidence) to rule out the vaccine risk. In each of these examples, a specific, unexpected complication occurred after the launch of the drug. Each was identified by the reporting of an adverse outcome.

In the second category, patients taking the therapeutic have been claimed to be at increased risk for common diseases. These risk increases were not anticipated in the approval trials and were identified post-launch or by new clinical trials focusing on the potential risk. The most successful weight-reduction therapeutic developed to date, fenfluramine-phentermine (Fen-Phen), has been highly publicized for its alleged increase in mitral valve disease prolapse (MVP) (60). MVP is a common cardiac disease frequently occurring in women at an incidence of 1 in 750. Following identification of index MVP cases, Fen-Phen was withdrawn and individual and class action litigation initiated. An estimated $20–30 billion has been awarded to plaintiffs—comparable to the cost of 30–40 new drugs. The ethics of this litigation was recently covered in a *Forbes* (61) article suggesting the

awards most benefited lawyers and doctors. The withdrawal of the COX-2 inhibitor rofecoxib (Vioxx) was in response to an increase in cardiovascular events in those who took the drug for more than 18 months (62, 63). This adverse effect has been a focus of the national media, and over 14,000 cases are now in litigation. The financial risk to rofecoxib's manufacturer has been estimated to be substantial, perhaps in excess of the Fen-Phen awards. Trial outcomes have thus far ranged from no award to $127 million. Of note, rofecoxib was re-reviewed by the FDA and approved for relaunch with additions to the warning label, but the manufacturer has not relaunched the drug.

The complex issue of drug interaction risk can be mentioned only briefly in this limited space. Determining and minimizing this risk are important problems with no good solutions. Patients with multiple or complex medical problems, such as patients post-transplantation, are prescribed a plethora of drugs for various complications without a full grasp of their interactions. In addition, patients often take over-the-counter or nutriceutical agents, which can influence the metabolism of prescription therapeutics. The future of therapy will prominently feature combination therapies, in keeping with the success of those for tuberculosis, AIDS, cancer, and cardiovascular diseases. Patient safety is overseen by the patient's physician in these settings.

The need for a technology change post-launch is evident. Geneticists understand that a 1000-patient study is inadequate to determine risk for common gene variations in humans. A very successful new drug for a common condition, such as Alzheimer's disease, hypertension, cardiovascular disease, diabetes, pain, or sleep disorders, may be used by 100,000 patients within a year after launch. Thus, clinical trials have limited capacity to detect our varied drug responses. Furthermore, clinical trials are frequently designed for patients with "clean" disease and not for patients with multiple diseases, such as cardiovascular disease/diabetes, Alzheimer's/obesity, asthma/autoimmune diseases.

I propose a post-launch plan, "phase IV," which would archive samples in anticipation of new technology that will sort out adverse outcomes. Phase IV incorporates concepts from pediatrics (newborn screening), new planning in biobanking, and application of emerging technologies with probative ability.

Newborn screening for childhood disease is now the standard of practice at the international level (60). The original technology, developed for the microbial Guthrie test, utilized dried heel blood spots on Whatman paper for the patient sample. Newborn diagnosis of phenylketonuria and galactosemia, with the consequent changes in diet, virtually eliminated mental retardation in these children. Private and government initiatives now target 45 diseases for newborn screening in the United States (64). The technologies for detection include ELISA, recombinant DNA, protein analysis, mass spectroscopy, and microbiology, all measured from paper-stored blood spots, which have been proven stable for >20 years.

We need to initiate a drug safety screening program for patients treated by new therapeutics. Some leaders in the pharmaceutical industry have already implemented private archiving for their clinical trials in phases I–III. The phase IV program would initiate at drug launch, an area that is frequently understudied (with occasional painful outcomes, as discussed above). This plan is not intended to replace the adverse events reporting system but rather to facilitate the investigation of these events for understanding what went wrong, and whether the events were associated with the therapeutic or not. The repository and its oversight would best be administered jointly by the FDA, the pharmaceutical industry, and the National Institutes of Health, with advocacy representation. Advocacy groups are needed for understanding of the safety objective, assuring protection of

personal information, and judging potential value.

The history and advantages of the blood spot, with its ease of collection, shipment, and storage, argue for it to be the first but not the only method to be considered. Collection should occur at the point of drug pickup and the collector should be compensated (**Figure 5**). There would be electronic entry of all a patient's drugs and dosages at the pharmacy. The physician's electronic records would archive disease diagnoses (CPT codes), progression, drug sensitivity, and prescriptions. Medical epidemiologists, geneticists, and physicians would provide advice on the time intervals of blood collection and size of population, as well as control of the population requirements. The FDA's electronic adverse outcome reporting would link to the program.

The use of samples for adverse outcomes study is of the highest priority. An initial

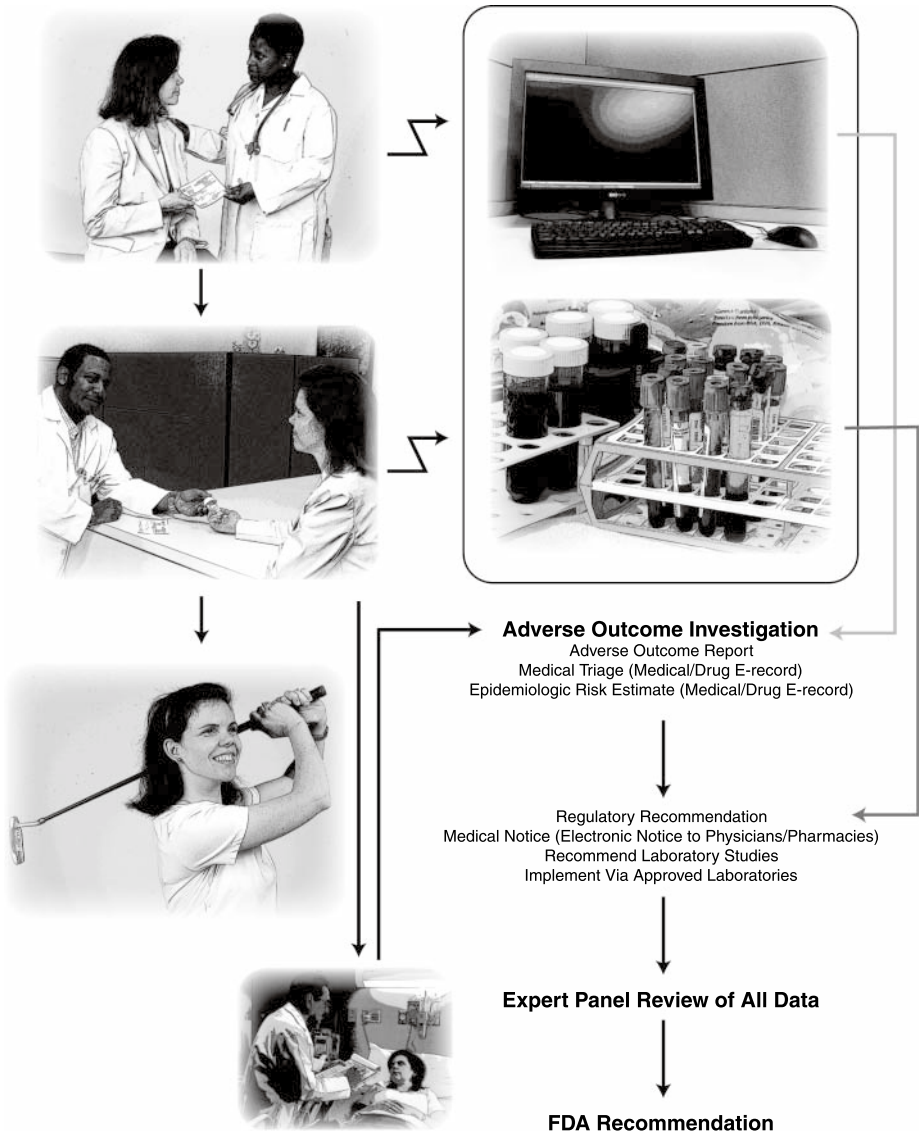

Figure 5

Adult health screening. Flow diagram for identification of adverse outcomes regarding the decision to study archived blood samples. Such a program enables the laboratory investigation of alleged adverse outcomes by a variety of recognized analytic methods and science-based decisions regarding the cause of suspected drug-related illness.

set of analytic tests should be preselected for adverse biomarker analysis upon clinical identification of a potential iatrogenic adverse outcome. Such tests might include mass spectroscopy for drug levels; SMA-20 surveys for metabolic, liver, and renal biomarkers; ELISA for cytokine quantitation; and detection of infectious agents, drug metabolism alleles, proteins/antibodies, and specialized biomarkers. Following a suitable time, the samples might be made available for research, as are samples used in newborn screening. In such applications, samples would likely be anonymized, and rules governing recontact would be determined in advance. Technologies for improved safety are powerless without the archived samples to study and a method to report the results publicly.

Through this "pharmaceutical biobank," drug safety would begin to match the safety we give children by detecting their genetic variations. We would protect the public as we provide new therapies for tragic diseases. Much has been written of the importance of genetic/environmental interactions; I propose the environmental addition of a single drug is most easily studied.

It should be noted that international research initiatives are proceeding with large biobanks (65), which store patient biologic materials with links to their medical and lifestyle features. Both the newborn screening and the biobank experiences will contribute to the ease of the pharmaceutical biobank's implementation. The establishment of the pharmaceutical biobank has the potential to protect the public, understand the mechanism(s) of drug toxicity, and improve regulatory guidance more rapidly. [*Disclosure Statement]

LITERATURE CITED

1. Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular Cloning, a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Lab. 2nd ed.
2. Shampo MA, Kyle RA. 2002. Kary B. Mullis—Nobel Laureate for procedure to replicate DNA. *Mayo Clin. Proc.* 77:606
3. Drews J. 1996. Genomic sciences and the medicine of tomorrow. *Nat. Biotechnol.* 14:1516–18
4. Sachidanandam R, Weissman D, Schmidt SC, et al. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928–33
5. Venter JC, Adams MD, Myers EW, et al. 2001. The sequence of the human genome. *Science* 291:1304–51
6. Waterston RH, Lander ES, Sulston JE. 2002. On the sequencing of the human genome. *Proc. Natl. Acad. Sci. USA* 99:3712–16
7. McKusick VA. 1994. *Mendelian Inheritance in Man*. Baltimore/London: Johns Hopkins Univ. Press
8. Capecchi MR. 1989. Altering the genome by homologous recombination. *Science* 244:1288–92
9. Thomson Healthcare. 2005. *Physicians' Desk Reference*. Montvale, NJ: Thomson PDR. 59th ed.
10. Ommaya A. Workshop of the Forum on Drug Discovery, Development, and Translation. 2006. *New approaches for preclinical safety and pharmacovigilance*. Presented at Natl. Acad. Sci., Inst. Med., Mar. 28–29, Washington, DC
11. Ommaya A. Workshop of the Forum on Drug Discovery, Development, and Translation. 2005. *The role of consumers and healthcare professionals in adverse drug event reporting*. Presented at Natl. Acad. Sci., Inst. Med., Nov. 3–5, Washington, DC

*This PDF ammended on (6 November 2007): See explanation at http://arjournals.annualreviews.org/errata/med

12. Singh A, Gilbert JK, Henske P. 2003. Rebuilding big pharma's business model. In *In Vivo: The Business and Medicine Report*, p. 73, Windhover Information
13. Caskey CT. 2006. Drug development crisis: efficiency and safety. **http://www.itmat.upenn.edu/symposium_materials/ppts/sat/Caskey.pdf**
14. International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:291–305
15. Zambrowicz BP, Abuin A, Ramirez-Solis R, et al. 2003. Wnk1 kinase deficiency lowers blood pressure in mice: a gene trap screen to identify potential targets for therapeutic intervention. *Proc. Natl. Acad. Sci. USA* 100:14109–14
16. Crawford DC, Nickerson DA. 2005. Definition and clinical importance of haplotypes. *Annu. Rev. Med.* 56:303–20
17. Walsh PC, Madden JD, Harrod MJ, et al. 1974. Familial incomplete male pseudohermaphroditism, type 2. Decreased dihydrotestosterone formation in pseudovaginal perineoscrotal hypospadias. *N. Engl. J. Med.* 291:944–49
18. Nurden AT, Rosa JP, Fournier D, et al. 1987. A variant of Glanzmann's thrombasthenia with abnormal glycoprotein IIb-IIIa complexes in the platelet membrane. *J. Clin. Invest.* 79:962–69
19. Hamm CW. 2003. Anti-integrin therapy. *Annu. Rev. Med.* 54:425–35
20. Nowell PC. 2002. Progress with chronic myelogenous leukemia: a personal perspective over four decades. *Annu. Rev. Med.* 53:1–3
21. Kurzrock R, Shtalrid M, Romero P, et al. 1987. A novel c-abl protein product in Philadelphia-positive acute lymphoblastic leukemia. *Nature* 325:631–35
22. O'Dwyer ME, Mauro MJ, Druker BJ. 2002. Recent advancements in the treatment of chronic myelogenous leukemia. *Annu. Rev. Med.* 53:369–81
23. Zambrowicz BP, Sands AT. 2003. Knockouts model the 100 best-selling drugs—will they model the next 100? *Nat. Rev. Drug Discov.* 2:38–51
24. Farooqi IS, O'Rahilly S. 2005. Monogenic obesity in humans. *Annu. Rev. Med.* 56:443–58
25. Dykxhoorn DM, Lieberman J. 2005. The silent revolution: RNA interference as basic biology, research tool, and therapeutic. *Annu. Rev. Med.* 56:401–23
26. Schena M, Shalon D, Davis RW, et al. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–70
27. Wang Y, Barbacioru C, Hyland F, et al. 2006. Large scale real-time PCR validation on gene expression measurements from two commercial long-oligonucleotide microarrays. *BMC Genomics* 7:59
28. Zhang J, Kaasik K, Blackburn MR, et al. 2006. Constant darkness is a circadian metabolic signal in mammals. *Nature* 439:340–43
29. Deleted in proof
30. Giepmans BN, Adams SR, Ellisman MH, et al. 2006. The fluorescent toolbox for assessing protein location and function. *Science* 312:217–24
31. MacDonald ML, Lamerdin J, Owens S, et al. 2006. Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nat. Chem. Biol.* 2:329–37
32. Staudt LM. 2002. Gene expression profiling of lymphoid malignancies. *Annu. Rev. Med.* 53:303–18
33. Quackenbush J. 2006. Microarray analysis and tumor classification. *N. Engl. J. Med.* 354:2463–72
34. Rosenblatt KP, Bryant-Greenwood P, Killian JK, et al. 2004. Serum proteomics in cancer diagnosis and management. *Annu. Rev. Med.* 55:97–112
35. Deleted in proof

36. Aebersold R, Anderson L, Caprioli R, et al. 2005. Perspective: a program to improve protein biomarker discovery for cancer. *J. Proteome Res.* 4:1104–9
37. Conrads TP, Zhou M, Petricoin EF III, et al. 2003. Cancer diagnosis using proteomic patterns. *Expert Rev. Mol. Diagn.* 3:411–20
38. Institute of Medicine of the National Academies of Science Forum on Drug Discovery, Development, and Translation. 2006. **http://www.iom.edu/CMS/3740/24155/24422.aspx**
39. Dollery CT. 1995. Frontiers in the development of new medicines in relation to clinical pharmacology. *Pharmacol. Toxicol.* 76:334–38
40. Nimjee SM, Rusconi CP, Sullenger BA. 2005. Aptamers: an emerging class of therapeutics. *Annu. Rev. Med.* 56:555–83
41. Jain KK. 2005. Applications of Amplichip CYP450. *Mol. Diagn.* 9:119–27
42. Suntharalingam G, Perry MR, Ward S, et al. 2006. Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. *N. Engl. J. Med.* 355:1018–28
43. Sussman HE, Smit MA, ed. 2006. *Genomes*. New York: Cold Spring Harbor Lab. Press
44. Goodyear M. 2006. Learning from the TGN1412 trial. *BMJ* 332:677–78
45. Solberg LA Jr. 2005. Biologic aspects of thrombopoietins and the development of therapeutic agents. *Curr. Hematol. Rep.* 4:423–28
46. Rossert J, Pure Red Cell Aplasia Global Scientific Advisory Board (GSAB). 2005. Erythropoietin-induced, antibody-mediated pure red cell aplasia. *Eur. J. Clin. Invest.* 35(Suppl. 3):95–99
47. Solberg LA Jr. 2005. Biologic aspects of thrombopoietins and the development of therapeutic agents. *Curr. Hematol. Rep.* 4:423–28
48. Koyama R, Nakanishi K, Kato M, et al. 2005. Hypoglycemia and hyperglycemia due to insulin antibodies against therapeutic human insulin: treatment with double filtration plasmapheresis and prednisolone. *Am. J. Med. Sci.* 329:259–64
49. Deleted in proof
50. Condra JH, Miller MD, Hazuda DJ, et al. 2002. Potential new therapies for the treatment of HIV-1 infection. *Annu. Rev. Med.* 53:541–55
51. Druker BJ. 2004. Imatinib as a paradigm of targeted therapies. *Adv. Cancer Res.* 91:1–30
52. Druker BJ. 2006. Circumventing resistance to kinase inhibitor therapy. *N. Engl. Med. Med.* 354:2594–96
53. Cobleigh MA, Vogel CL, Tripathy D, et al. 1999. Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. *J. Clin. Oncol.* 17:2639–48
54. 2006. Common ground on the critical path. *Nat. Rev. Drug Discov.* 5:267
55. Czernin J, Phelps ME. 2002. Positron emission tomography scanning: current and future applications. *Annu. Rev. Med.* 53:89–112
56. Zhang J, Hutton G. 2005. Role of magnetic resonance imaging and immunotherapy in treating multiple sclerosis. *Annu. Rev. Med.* 56:273–302
57. Meyer MA. 2006. Natalizumab for relapsing multiple sclerosis. *N. Engl. J. Med.* 354:2387–89
58. Woods NB, Bottero V, Schmidt M, et al. 2006. Gene therapy: therapeutic gene causing lymphoma. *Nature* 440:1123
59. Shaw AR. 2006. The rotavirus vaccine saga. *Annu. Rev. Med.* 57:167–80
60. Blanck HM, Khan LK, Serdula MK. 2004. Prescription weight loss pill use among Americans: patterns of pill use and lessons learned from the Fen-Phen market withdrawal. *Prev. Med.* 39:1243–48

61. Langreth R, Herper M. 2006. Pill pushers: How the drug industry abandoned science for salesmanship. *Forbes* May 8, pp. 95–100
62. Kondro W. 2004. Lawsuits mount in wake of rofecoxib (Vioxx) withdrawal. *Can. Med. Assoc. J.* 171:1335
63. Couzin J. 2004. Drug safety. Withdrawal of Vioxx casts a shadow over COX-2 inhibitors. *Science* 306:384–85
64. Pollitt RJ. 2006. International perspectives on newborn screening. *J. Inherit. Metab. Dis.* 29:390–96
65. Knoppers BM. 2005. Biobanking: international norms. *J. Law. Med. Ethics* 33:7–14

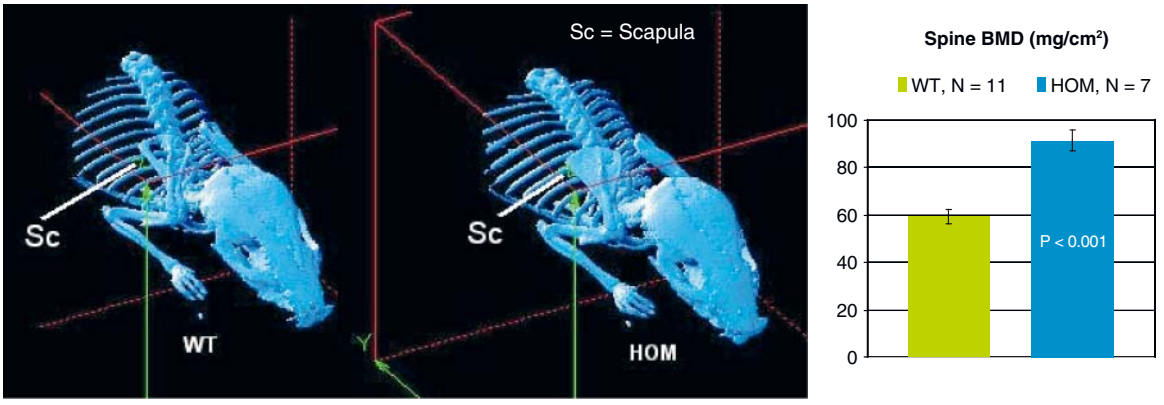

Figure 2

Knockout of secreted protein shows a general increase in bone density by CAT scan. Image courtesy of Lexicon Genetics, Inc.

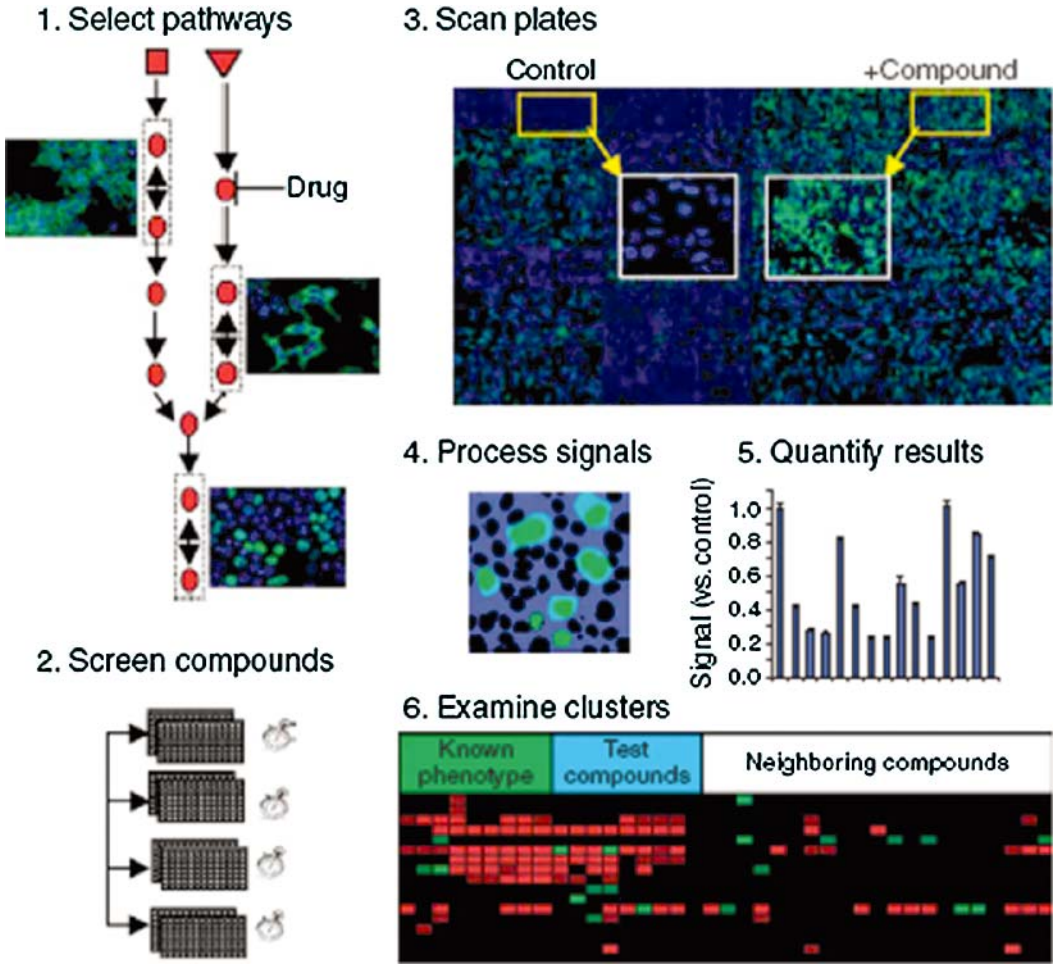

Figure 3

Strategy for pharmacologic profiling of compounds with high-content PCAs. (1) Pathways of interest (*red spheres connected by arrows*) are selected and high-content PCAs are created. Assays measure dynamics of specific pathway activation or inhibition by quantifying changes in abundance or location of protein complexes coupled to that pathway that are elicited in response to activator (*red square* and *triangle*) or inhibitor (⊥) drugs. *Inset*: Images of three such assays that report on dynamic complexes coupled to the individual pathways (*dotted-line boxes*) localized to membrane, cytosol, and nucleus. PCA signal is in green; nuclear (Hoechst) staining is in blue. (2) Cells expressing PCAs arrayed in 96-well plates are treated with compounds or vehicle controls, fixed after specified times, and treated with cell-compartment-specific counterstains. (3) Multiple images are captured from control and compound-treated wells. Pixel intensities from PCA signals are extracted from one or several cell compartments on the basis of colocalization with counterstain (4) and tabulated for individual compound treatments (5). Data for each compound versus PCA response at different times are represented as an array. Changes in signal intensity or location for compound versus vehicle control are represented by a color code, where green represents an increase and red a decrease in PCA signal versus control in units of coefficient of variation of each assay. Data are clustered by compounds and assays to identify on-pathway or off-pathway effects of compounds on specific pathways. The matrix also allows identification of test compounds that cluster with drugs of a known phenotype that they are expected to share. From Reference 31 with permission.