CHEMOGENOMICS: AN EMERGING STRATEGY FOR RAPID TARGET AND DRUG DISCOVERY

*Markus Bredel**‡ *and Edgar Jacoby*§

Chemogenomics is an emerging discipline that combines the latest tools of genomics and chemistry and applies them to target and drug discovery. Its strength lies in eliminating the bottleneck that currently occurs in target identification by measuring the broad, conditional effects of chemical libraries on whole biological systems or by screening large chemical libraries quickly and efficiently against selected targets. The hope is that chemogenomics will concurrently identify and validate therapeutic targets and detect drug candidates to rapidly and effectively generate new treatments for many human diseases.

TRANSCRIPTIONAL PROFILING
The study of the transcriptome — the complete set of RNA transcripts that are produced by the genome at any one time — using high-throughput methods, such as microarray analysis.

**Division of Oncology, Stanford University School of Medicine, 269 Campus Drive, CCSR-1110, Stanford, California 94305-5151, USA.*
‡Department of General Neurosurgery — Neurocenter, University of Freiburg, Breisacher Str. 64, D-79106 Freiburg, Germany.
§Novartis Institutes for Biomedical Research, Discovery Technologies, Molecular and Library Informatics Program, CH-4002 Basel, Switzerland.
Correspondence to M.B. e-mail: mbredel@stanford.edu
doi:10.1038/nrg1317

Over the past five decades, pharmacological compounds have been identified that collectively target the products of ~400–500 genes in the human body; however, only ~120 of these genes have reached the market as the targets of drugs[1,2]. The Human Genome Project[3,4] has made available many potential new targets for drug intervention: several thousand of the approximately 30,000–40,000 estimated human genes[4] could be associated with disease and, similarly, several thousand human genes could be susceptible to drugs. Overlaying the domains of drug-susceptible and disease-associated genes will probably allow the identification of genes that could serve as new drug targets[1].

In this context, new approaches are needed to move rapidly from gene to drug. The integration of recent progress in high-throughput genomics (including TRANSCRIPTIONAL PROFILING) and several chemistry disciplines has markedly influenced current target- and drug-discovery practices and has given rise to a new research discipline, known as 'chemogenomics'. This discipline is best defined as the study of the genomic and/or proteomic response of an intact biological system — whether it be single cells or whole organisms — to chemical compounds, or the study of the ability of isolated molecular targets to interact with such compounds. This response is gauged by phenotypic readouts and high-throughput assay technologies. Note, however, that owing to the emergence of various subspecialties of chemogenomics (discussed in the next section) and the involvement of several disciplines, it is currently almost impossible to give a simple and common definition for this research discipline (BOX 1).

In chemogenomics-based drug discovery, large collections of chemical products are screened for the parallel identification of biological targets and biologically active compounds. The biologically active compounds that are discovered in this way are known as 'targeted therapeutics'[5] because they bind to and modulate specific molecular targets. Although the chemogenomics strategy seems to have much in common with large-scale serendipity, the basic idea of its application is to accelerate the pace of the drug-discovery process. In the past couple of years, many leading pharmaceutical drug development companies have established chemogenomics groups, and hundreds of novel targets and compounds that have been identified by chemogenomics are currently undergoing biological testing further along the drug-development pipeline.

Chemogenomics is a direct descendent of the conventional pharmaceutical approach to product discovery in that it also involves screening libraries of compounds for their effects on biological targets. However, it differs from the conventional pharmaceutical approach because it directly links such extensive library screens to

©2004 Nature Publishing Group

Box 1 | **Defining chemogenomics**

There is some confusion about the meaning of the term 'chemogenomics'[103]; this might be expected given the involvement of so many disciplines. In particular, there is considerable overlap among the related strategies described by the terms 'chemical genetics'[104] and 'chemical genomics'[105,106]. Although these three terms are sometimes used interchangeably, the primary goal of both of the last two strategies is the study of cellular function using small synthetic molecules as modulating ligands.

By contrast, the term 'chemogenomics' is often used to describe the focused exploration of target gene families, in which small molecule leads — identified by virtue of their interaction with a single member of a gene family — are used to study the biological role of other members of that family, the function of which is unknown. The additional members of a gene family are usually identified by sequence homology[107–109]. So, in this strategy, genetic sequences are annotated and translated into potential pharmacological targets. Such grouping of gene families according to genetic sequence homology has the potential to cover a broad range of therapeutic areas, because although genes in a family can encode structurally similar proteins, each protein of a gene family can exert a very different biological function.

The promotion of the target family-based approach to an industrial discovery technology was probably first emphasized by researchers at Glaxo Wellcome, who based it on the analysis of gene families that had been successfully explored up to that point[110]. The scientists at Glaxo Wellcome highlighted obvious advantages of system-based approaches, such as combining advances in gene cloning and expression, automation, COMBINATORIAL CHEMISTRY and bioinformatics. Since then, numerous pharmaceutical and biotechnology companies have built their business models on these principles. Most notably, the pioneering company Vertex Pharmaceuticals designed its entire drug-discovery process around target families in contrast to the traditional disease-area-orientated organization of pharmaceutical research[103,111].

Although chemogenomics historically refers to the gene-family-based approach to drug discovery, in this review, a broad definition of chemogenomics is emphasized, which includes every scientific approach that uses high-throughput technology to study the genomic effect of chemical compounds.

modern genomics methodologies. To accomplish this goal more effectively and systematically, it is necessary to have access to new synthetic COMPOUND LIBRARIES. Several chemical disciplines, such as combinatorial chemistry, SYNTHETIC CHEMISTRY and CHEMOINFORMATICS, have been used for this purpose. Traditional screening methods were limited to assessing gross responses to a limited panel of natural products; by integrating genomics, bioinformatics and the above-mentioned chemical disciplines, chemogenomics has replaced this approach with the rational development and rapid screening of target-specific chemical ligands.

Although chemogenomics has garnered support from virtually all areas of research, such as immune, inflammatory and hormone research, the application that stands to particularly gain from chemogenomics is cancer research, as the technique can help to identify treatment strategies that selectively target certain molecular alterations. Human cancers show complex and multiple pathogenic aberrations, for which targeted therapies are urgently needed to ameliorate the largely negative outcome of these diseases. Cancer research is therefore particularly poised to take advantage of the high-throughput nature of chemogenomics.

This article outlines the goals and strategies of chemogenomics and reviews the application of this method, particularly to cancer research and therapy. We also discuss the future prospects of this new discipline and the unique and mainly technical challenges that it faces. Such challenges include the need for a more refined integration of bioinformatics and chemoinformatics data, a more rational approach to selecting designed compounds from an almost infinite number of synthetic possibilities and the ability to build more FOCUSED LIBRARIES for screening.

Experimental chemogenomics

Chemogenomics integrates target and drug discovery by using active compounds (or ligands) as probes to characterize proteome functions. As in genetics, forward and reverse principles can be distinguished, depending on whether the investigative process moves from phenotype to target or from target to phenotype. The target is the gene product, whereas a ligand is any molecule that is able to bind to the target with a certain specificity. Physical compound libraries that consist of hundreds of thousands to millions of potential ligands have been developed. There are two different sorts of these compound libraries: natural product libraries and synthetic libraries. The advantage of the former is their inherently large-scale structural diversity and their intrinsic property as having been evolutionarily optimized within biological systems, whereas the latter are characterized by an almost infinite number of synthetic possibilities.

Reverse chemogenomics. In 'reverse chemogenomics', gene sequences of interest are first cloned and expressed as target proteins, which are then screened in a high throughput, 'target-based' manner by the compound library (FIG. 1b). Such a HIGH-THROUGHPUT SCREENING (HTS) method can involve many different bioassays[6–11], which monitor the effects of different compounds on specific targets (such as the ability to bind a protein), on specific cellular pathways (for example, the capacity to inhibit the mitogenic pathway of a tumour cell)[12] or on the phenotype of a whole cell or organism. The assays can be generally divided into cell-free, cell-based and organismal assays[13]. Cell-free, universal binding assays — in which several compounds are simultaneously tested for their binding affinity to a wide panel of specific targets — are usually simple, precise, highly automated and

COMBINATORIAL CHEMISTRY
A process for preparing large collections of compounds, or 'combinatorial libraries', by synthesizing all possible combinations of a set of smaller chemical structures or 'building blocks'.

COMPOUND LIBRARY
A structurally diverse collection of chemical molecules, typically containing several hundred thousand entities, that is used to identify new lead candidates.

SYNTHETIC CHEMISTRY
A branch of chemistry that focuses on the deliberate manufacture of pure compounds of defined structure and/or the development of new chemical reactions for this purpose.

CHEMOINFORMATICS
A generic term that encompasses the design, creation, organization, management, retrieval, analysis, dissemination, visualization and use of chemical information, with the intended purpose of guiding drug discovery and development.

FOCUSED LIBRARY
Compound libraries that are enriched for desired properties, such as target-binding affinity, by using computational library design methods.

HIGH-THROUGHPUT SCREENING
The large-scale, trial-and-error evaluation of compounds in a parallel target-based or cell-based assay.

©2004 Nature Publishing Group

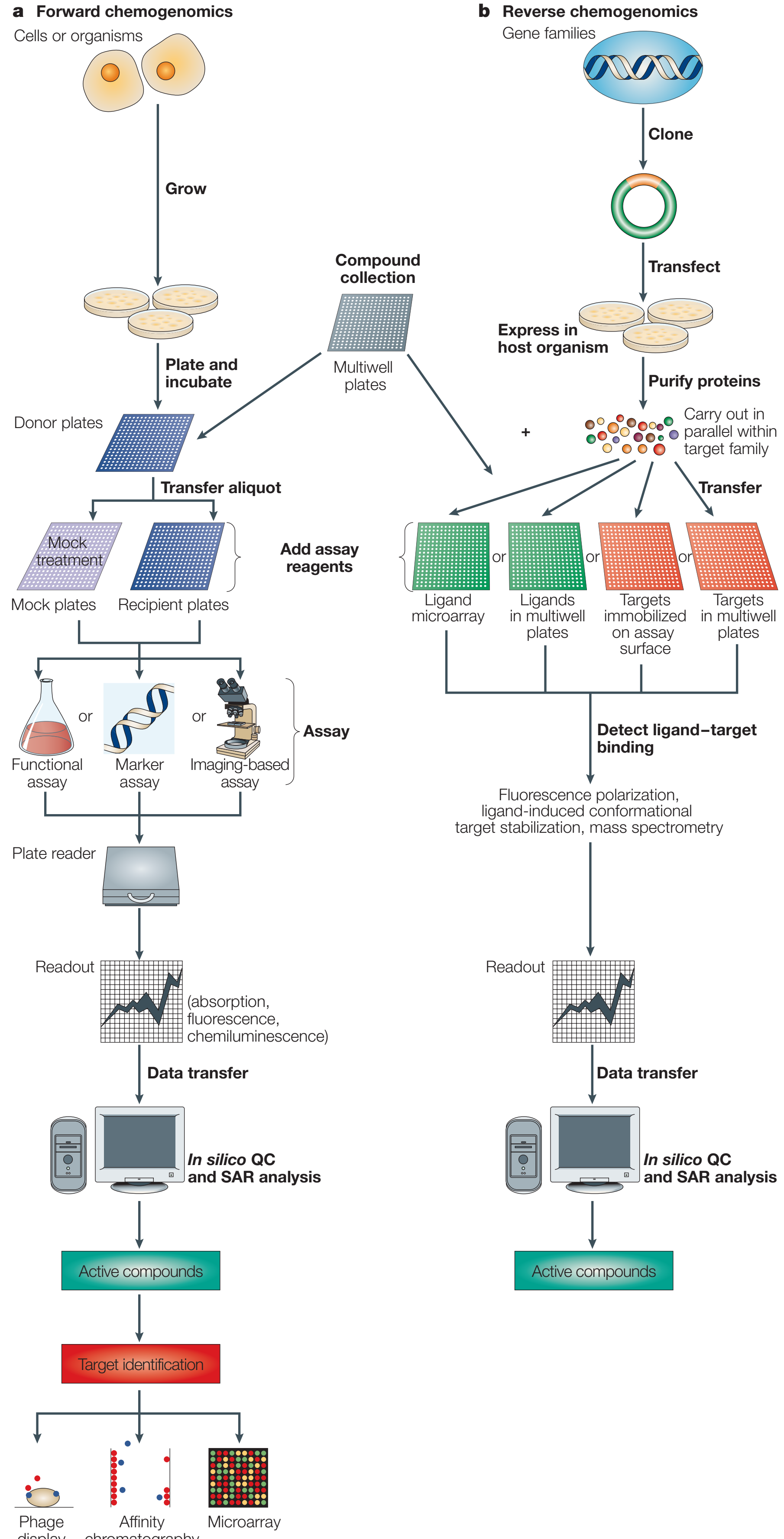


Figure 1 | **Forward- and reverse-experimental chemogenomic approaches.** The initial step in forward and reverse chemogenomics is to select a suitable collection of compounds and an appropriate model system in which to screen them. In both approaches, the sequential steps of the assay — the transfer of ligands, cells, assay reagents and plates — can be fully or semi-automated. IT integration is a key element in industrial setups. **a** | In forward chemogenomics, the cell or organismal model system is typically dispensed in multiwell microtitre or nanowell plates. Solutions of single ligands are added from the stock plates to different wells. After incubation, an aliquot is transferred from the donor plate to a new recipient plate, in which the ligand–target binding assay is carried out. The effects of a compound are assayed by one of several methods: functional assays directly measure cellular activities such as cell division; marker assays, such as reporter-gene assays[114] and whole-culture CYTOBLOTS[7], identify specific molecular events that act as surrogate transcriptional and post-transcriptional markers for phenotypic changes of interest; automated microscopy or imaging-based screening[115] are innovative approaches that attempt to capture further morphological changes. The end point of most cell-based high-throughput assays is a spectroscopic readout; readout data are automatically transferred to a microprocessor for final data calculation, including *in silico* quality control and structure–activity relationship (SAR) analysis. Active compounds that achieve the desired phenotypic change are then selected to identify their molecular targets. This can be done in several ways, of which AFFINITY MATRIX PURIFICATION, PHAGE DISPLAY or transcriptional or PROTEOMIC PROFILING are the most commonly used approaches. In profiling experiments, protein or RNA isolates of the treated model system are analysed in reference to mock treatment for global molecular drug signature assessment. **b** | In reverse chemogenomics, emphasis is especially placed on the parallel exploration of gene and protein families. Here, target gene sequences that show a certain degree of homology are expressed in a host cell: these family target proteins are purified, collected and subjected to assay design. Based on the SAR homology concept[116] that the degree of similarity of ligands determines similarities in target binding, candidate ligands that show a desirable similarity to a ligand that is known to interact with one member of a target family are selected.The ultimate goal is to identify new ligands that hit either the same target or analogous target-family members. Reverse chemogenomics is normally carried out using a cell-free binding system, in which either the target protein or the libraries are immobilized on assay plates or dispensed in multiwell plates, and the study compounds or target proteins, respectively, are added in solution. Various technologies are used to detect ligand–target binding. In fluorescence-based detection, an imaging camera automatically captures the fluorescence signal that corresponds to ligand–target binding. As in forward chemogenomics, readout data are transferred to a data analysis system, which uses sophisticated computational algorithms to mine the large amounts of — and, to some extent, noisy and error-prone — data.

©2004 Nature Publishing Group

CYTOBLOT
A cellular immunoassay that uses primary antibodies to gauge specific post-translational changes, such as the abundance, modification or conformational change of a protein, as a surrogate measure of a phenotypic change of interest.

AFFINITY MATRIX PURIFICATION
Purification of targets on the basis of their interactions with a ligand that is attached to an immobilized matrix, such as agarose beads, to form an affinity column.

PHAGE DISPLAY
A technique that fuses foreign peptides to capsid proteins on the phage surface. Immobilized libraries of phage-displayed peptides might be screened for binding to specific ligands; determination of the gene sequence of the selected phage identifies the peptide sequence.

PROTEOMIC PROFILING
The systematic analysis of protein expression of normal and diseased tissues that involves the separation, identification and characterization of proteins that are present in a biological sample.

PHYSICOCHEMICAL PROPERTIES
The characteristics of a compound that are relevant to pharmacokinetics studies, such as solubility or membrane permeability.

DRUG-LIKENESS
The concept that drugs share specific molecular properties that distinguish them from other natural or synthetic chemicals.

LEAD OPTIMIZATION
The concurrent optimization of many pharmacological design features, such as target-binding affinity and selectivity, by iterative design, synthesis and testing in biological model systems.

QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIP
(QSAR). An analysis that describes the association between the molecular structure of a compound and its ability to affect a biological target.

LIGAND-INDUCED CONFORMATIONAL STABILIZATION
A phenomenon in which substrates, inhibitors, cofactors and even other proteins provide enhanced stability to proteins on binding.

Box 2 | From chemical compound to drug

In the discovery and development of a new drug, different terms are assigned to the substances that are studied at different steps in the process. Only if a substance fulfills the requirements for a certain level is it subjected to testing on the next level.

The process starts with a 'compound' — a chemical substance of defined chemical structure and purity. Compounds that show specific binding properties and are active are referred to as 'hits' and are further validated; this includes verifying their chemical identity and purity, and determining their binding properties in one or more secondary assays.

Follow-up research on candidate compounds is then prioritized according to sophisticated chemoinformatics strategies that take into account practical factors, such as the PHYSICOCHEMICAL PROPERTIES of a compound, its 'DRUG-LIKENESS' and feasibility of synthesis. This so-called LEAD OPTIMIZATION process uses QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIP (QSAR) data, and, in particular, the synthesis of analogous synthetic compounds with refined properties. These refined properties include both an increased potency and target selectivity as well as drug-like characteristics that are compatible with testing in biological model systems. The synthetic analogues that are predicted to have improved properties are subjected to assay-based high-throughput screening (HTS) to compare their biological effects with those of the initial lead compounds. Repeated rounds of concomitant HTS and chemical analogue synthesis give rise to a progressive lead optimization.

Cellular or biological model systems are then used to validate the discovered hits *in vitro* and/or *in vivo* and to assess the potency and specificity of the effect of a compound.

Selected clinical candidates are tested in humans during phases 1–3 of the clinical studies according to predefined protocols that analyse their toxicity and efficiency profile for specific therapeutic indications. The end point of the drug-discovery and development process is a 'drug', a registered and approved chemical substance that is used in a dose-dependent manner for the treatment of a medical condition.

compatible with a very high throughput approach. Target–ligand interactions (so-called 'hits') are unambiguously identified in the absence of confounding variables. For example, fluorescent-based methods for detecting the LIGAND-INDUCED CONFORMATIONAL STABILIZATION of proteins[14] or MASS-SPECTROMETRY-based detection systems[15,16] have been described as a means to examine the effect of bound ligands. By contrast, cell-based and organismal assays — in which selected compounds are delivered directly to cells or organisms *in vitro* — identify hits within a relevant cellular context, but, because of the interaction with multiple targets, hits require additional mechanistic characterization. Whereas cell-free assays are primarily used in reverse chemogenomics, if target information is available, cell-based and organismal assays are predominantly used in forward chemogenomics (see below) to examine broad compound effects on intact biological systems.

The hits that are revealed in this way are used to generate lead compounds. These are then optimized by the careful selection of the most promising candidates and through the synthesis and testing of chemical ANALOGUES with similar and, it is to be hoped, improved properties (BOX 2). Reverse chemogenomics is therefore virtually identical to the target-based approaches that have been applied in drug discovery and molecular pharmacology over the past decade; however, these are now enhanced by parallel screening and by the ability to perform lead optimization on many targets that belong to one target family.

Forward chemogenomics. In 'forward chemogenomics' (FIG. 1a), the molecular basis of a desired phenotype is unknown. Here, a so-called 'phenotypic screen' is performed in single-cell organisms or cells from multicellular organisms using a panel of ligands. The biological systems can consist of prokaryotic and eukaryotic single-cell organisms (bacteria and fungi; for example, the yeast *Saccharomyces cerevisiae*[17–21]), physiological or pathological cells from complex multicellular vertebrate or mammalian organisms, or even whole higher organisms, such as fly, worm, zebrafish or mouse. Subsequently, high-throughput cell-based or organismal phenotypic assays are used to identify biologically active compounds[22]. In other words, in this approach, compounds are identified on the basis of their conditional phenotypic effect on a whole biological system rather than on the basis of their inhibition of a specific protein target. The phenotypic screen is designed to reveal a novel conditional phenotype (either a loss-of-function or a gain-of-function phenotype) and the affected protein or pathway. For example, a loss-of-function phenotype could be an arrest of tumour growth. Once biologically active compounds that lead to a target phenotype have been identified, efforts are directed towards the study of the mechanistic basis of the phenotype by identifying the gene and protein targets using various high-throughput methods[22,23] (FIG. 1a). The biological and structural information on the target can, in turn, be used in reverse chemogenomics to identify and develop, through HTS, new and more potent compounds that disrupt the function of the target. On the other hand, active ligands that are identified using reverse chemogenomics can be biologically validated by examining the phenotypic effect of altering the function of their protein targets in a forward chemogenomics setting. The main challenge of this chemogenomics strategy lies in designing phenotypic assays that lead immediately from screening to target identification.

Predictive chemogenomics. Whereas the principal goal of chemogenomics is to identify new therapeutic targets and drugs, 'predictive chemogenomics' strategies primarily attempt to holistically characterize treatment responses, coupled with the secondary aim of identifying novel therapeutic molecules. The central approach

©2004 Nature Publishing Group

MASS SPECTROMETRY
A technique that is used to determine the composition and abundance of the atoms in and the molecular mass of complex molecules, starting with a small amount of the sample.

[SYNTHETIC] ANALOGUE
Closely related, synthetically synthesized members of a chemotype — a family of molecules that demonstrate a unique core structure or scaffold — with minor chemical modifications that might show improved target-binding affinity and potency compared with the original natural lead compound.

PHARMACOGENOMICS
The study of how and which variations in the human genome affect the response to medications.

ULTRA HIGH-THROUGHPUT SCREENING
Screening activity that is accelerated to more than 100,000 tests per day.

PRIVILEGED STRUCTURE
A core or scaffolding structure that, independent of specific substituents attached to it, imparts a generic activity towards a protein family or a subset of such a family.

DIVERSITY SET [LIBRARY]
Diversity-orientated synthesis-based libraries augment the accessible structural diversity of the library by mimicking the structural complexity and diversity of natural products.

BOOTSTRAP [ANALYSIS]
A type of statistical analysis that is used to test the reliability of certain branches in the evolutionary tree. The bootstrap analysis proceeds by re-sampling the original data, with replacement, to create a series of bootstrap samples of the same size as the original data. The bootstrap value of a node is the percentage of times that a node is present in the set of trees that is constructed from the new data sets.

Table 1 | **Key *in silico* methods that are used to support chemogenomics approaches**

***In silico* method**	**Application setting and objective**	**References**
Similarity and pharmacophore* searching	Virtual screening or design of compound libraries to provide focused subsets, on the basis of the knowledge of a known active ligand.	80,118–120
	Homology-based searching aims to identify compounds that are active on a target for which there are no known active compounds but that are related by homology to one or more targets for which active compounds are known (using the 'structure–activity relationship homology concept')[38].	112
High-throughput docking‡ (HTD)	Virtual screening or design of compound libraries to provide focused subsets, on the basis of the knowledge of the 3D structure of the target.	80,118,121–126
	Identification or design of selective compounds: docking a focused library against a comprehensive panel of 3D structures of one protein family.	127
	Identification of potential targets and mechanisms of action: docking of selected compounds against the entire PDB§ database.	128
3D bioinformatics target-binding site comparison	Identification of potential targets and mechanisms of action of compounds on the basis of the structure-based comparison of the ligand-binding sites. Design of selective compounds within a target family with conserved molecular recognition.	129,130
Target and ligand annotation and ontologies	By linking the sequence information of targets to their ligands, ligand–target classification schemes allow ligands to be matched to targets on the basis of their sequence similarity and provide reference sets for homology-based similarity searching. The automatic build-up of compound annotations on the basis of Medline literature reports provides the knowledge basis for generating annotated compound libraries. These can then be used to guide experiments to determine the components of signalling pathways.	98,131,132

*Pharmacophore, the predicted combination of steric and electronic features and the spatial arrangement of chemical groups that determine the interaction of a set of compounds with a specific biological target. ‡Docking, the process of assessing how a ligand and a target fit together in three-dimensional space. High-throughput docking (HTD) is used to virtually dock drug-like compounds to their macromolecular targets to predict their potential activity against the targets. §PDB, Protein Data Bank, a worldwide biological repository for the processing and distribution of three-dimensional macromolecular structure data.

of predictive chemogenomics is to initially collect the genomic responses (for example, through microarray analysis) and the pharmacological responses (for example, through growth inhibition assay) of a cell type or tissue to treatment with various drugs. Each drug profile represents the drug's own signature at the transcriptional and molecular pharmacological level. Biostatistical integration of the genomic and pharmacological data then reveals predicted gene–drug relationships. This approach can be extended to look at families of drugs to extract the signatures that are common to a class of molecules (that is, the effect that is linked to a chemical structure) and those that are drug specific. This strategy will not necessarily reveal drug targets but might identify molecules that significantly influence the effect of a drug. Computational or *in silico* methods can complement experimental chemogenomics strategies in the search for such predictive molecules (TABLE 1).

Predictive chemogenomics has considerable overlap with 'PHARMACOGENOMICS'. In contrast to pharmacogenomics, however, predictive chemogenomics strategies generate gene–ligand response associations by concurrently considering the response profiles of thousands of drugs, rather than those of one molecule at a time.

Ligand and target selection

From a cost perspective, it is important that biologically active chemical candidates are identified quickly, efficiently and accurately. The original combinatorial chemistry approach to ligand and target identification involved the ULTRA HIGH-THROUGHPUT SCREENING (uHTS) of diverse-compound collections; however, it soon became clear that such massive screening is hardly applicable in the above experimental chemogenomics settings because of the necessary high level of financial investment and the substantial efforts needed for data handling (IT logistics and automation) and interpretation. In experimental chemogenomics, emphasis has therefore now been placed on the pre-selection of potential ligands to allow the study of smaller, focused, libraries of compounds. Several strategies have been applied to generate so-called 'targeted libraries', in which the design and selection of ligands is based on the information that is available on either the target itself (for example, through three-dimensional X-ray/nuclear magnetic resonance (NMR) structure) or on ligands that are known to interact with the target[24–27]. For targets with limited or no biostructural information, PRIVILEGED STRUCTURES are often considered in library design. Today, typical chemogenomics screening libraries contain several types of designed subset, including annotated known biologically active compounds, target-family focused libraries (for example, kinases, proteases), peptide mimetics (for example, β-strand, β-turn and α-helix structural mimetics), natural products and derivatives thereof, and DIVERSITY SETS of drug-like compounds[28–31].

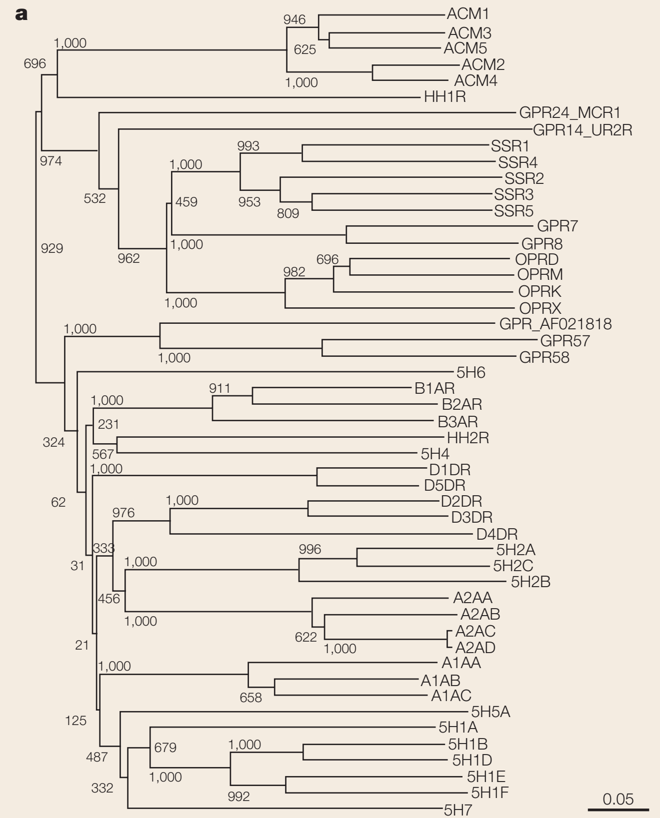
As well as using focused library design, cost-efficient chemogenomics requires genetic sequences of relevance to be dissected from those that do not contribute to the ligand–target interaction. For this purpose, sequence homology alignments and biostructural algorithms can be used to narrow down the number of targets to be tested[25].

©2004 Nature Publishing Group

Box 3 | Reverse-chemogenomics case study: designing a focused library of ligands for monoamine-related GPCRs

The successful design of family-focused target libraries depends on how similar the ligands and binding sites are among the members of a gene family. We recently proposed that the monoamine-related G-protein-coupled receptor (GPCR) subfamily, for which motif-based sequence searches identified 50 human GPCR members, recognizes its ligands through 3 binding sites. The sequence comparisons of the 50 identified GPCRs, based on the 7-transmembrane (7TM) domain, are shown in the dendrogram in panel **a** (the scale of sequence identity is indicated by the 5% distance bar and the numbers on the branches are BOOTSTRAP values out of 1,000 replicates). For the serotonin $5HT_{1A}$-receptor subtype, each of the three spatially distinct binding regions allows the receptor to bind a different ligand (panel **b**, top). These regions are located within the highly conserved 7TM domain of the receptor and overlap at the residue $D_{3.32}$ in TM3, which is responsible for the recognition of the basic amino group of the ligands.

This information motivated the design of the Novartis tertiary amine (TAM) combinatorial ligand library. The TAM structures, for which prototypes are shown in panel **c**, were designed to be similar in architecture and properties to known monoamine-related GPCR ligands, for which examples are shown in panel **b**. The successful search for antagonists for the $5HT_7$ GPCR, which has the $5HT_{1A}$ receptors as next neighbour in the sequence dendrogram, illustrates the use of the TAM library. By searching with $5HT_{1A}$ reference compounds in the TAM library (using the Similog method[112]), we were able to identify a 10% hit rate (pK_B <5 µM, where pK_B = the negative logarithm of the binding constant) when only a biological assay with limited capacity was available — that is, when only a limited number of componds could be screened. The hits corresponded to arylpiperazines (see panel **c**), which, in follow-up studies, were also active on other monoamine-related GPCRs[113]. Panel **a** reproduced with permission from REF. 113 © (2001) Wiley–VCH.



b

Serotonin
5-HT agonist

Propranalol
β antagonist

8-OH-DPAT
5-HT1A partial agonist

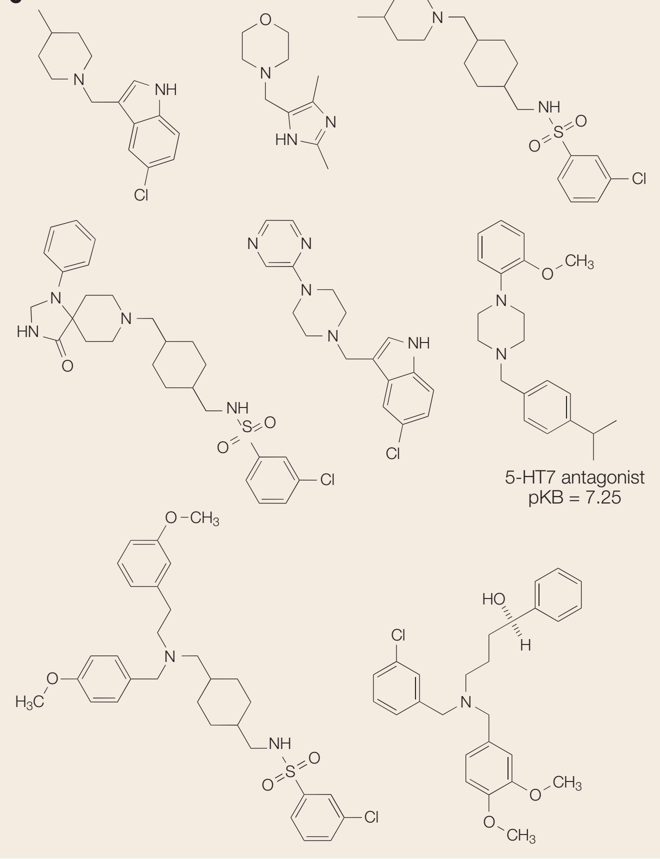
Ketanserin
5-HT2A antagonist

Janssen_1
D4 antagonist

WAY-100635
5-HT1A antagonist

RO-16814
β agonist

Kissei_1
D2 antagonist



©2004 Nature Publishing Group

Applications

Chemogenomics strategies are increasingly being harnessed by various fields of medical research – for example, those related to cancer or immune, inflammatory and hormone disease – in attempts to develop new targeted therapies as rapidly as possible (BOX 3). Several advances have been made by chemogenomics in understanding the molecular biology of various diseases and in identifying potential pharmacological therapies for them. TABLE 2 summarizes some recent results of chemogenomics research into human diseases.

In general, chemogenomics approaches can be used for three different purposes in disease research. First, chemogenomics can be used to identify new drug targets and might allow their biological functions to be understood. In this context, forward chemogenomics strategies are used to initially examine the phenotypic effect of a compound or a panel of compounds on a

Table 2 | **Chemogenomics applications***

Focus/approach	Ligands	Target	Disease	Model	Validation model	Refs
Target identification/F	FK228, Trichostatin A, Depudecin	HDAC	Cancer/ angiogenesis	*Hras*-Ras1, *vras*-NIH3T3 cells, mammalian cell lines	Proliferation/enzyme assay	37–39
Target identification/F	TNP-470	MetAP2	Cancer/angiogenesis	W303 yeast strain, endothelial cells	Yeast-deletion model	19,20
Target identification/F	Dihydroepo-nemycin	20S proteasome (LMP2, LMP7)	Cancer/ angiogenesis	EL4 murine thymoma cells	2D-gel electrophoresis, immunoblotting	133
Target & drug identification/F	PNRI-299	Ref1	Asthma	A549 lung epithelial cells	BALB/c mice	47
Target & drug identification/F	Radicicol-derivatives BR-1, BR-6, KT8529, KF25706	Hsp90, ACL	Cancer/ angiogenesis	NIH3T3, RAS-373, SRC-373 mouse fibroblasts, HeLa cells, normal rat and human tumour cells	BALB/c mice, *ex vivo* western blot, immuno-blotting, enzyme inhibition	32–34
Target & drug identification/F	FK506, cyclosporine A	Calcineurin, CPH1 & FPR1	Immune disease	Yeast strains	*In vitro* and mice	22,134, 135
Drug & target identification/F	K252a	CaMPKs	Neuroinflammation	BV-2 microglia cells	Western blot	136
Drug identification/F	Compound A	pRB	Cancer	HT29, HCT116, C33A cancer cells (cytoblot)	Western blot	137
Drug identification/F	Compounds 5a–5h	Ras-Raf	Cancer	MDCK epithelial cells, MDCK-F3 (HRas)	Phenotype reversal	12
Drug identification/R	CDK-731	MetAP-2	Cancer/angiogenesis	Endothelial cells	Phenotype reversal (proliferation inhibition)	36
Drug identification/R	Peptide 18 & analogues	CaMPKs/MLCK	Various	Enzyme inhibition assay	Kinetic analysis of target inhibition	138
Drug identification/R	Isatin oximes 30, 50, 51	UCH-L1	Cancer	High-throughput screening	Phenotype induction in H1299 lung cancer cells	139
Drug identification/ R and IS	Compounds 1a–c and 11l	HDAC	Cancer/ angiogenesis	*In vitro* enzyme assay, mouse A20 cells, virtual mutation	Phenotype reversal in murine erythro-leukemia cells	140,141
Drug identification/ R and IS	NSC-65828, R5A, H8A, N68A, des (121-123)	AGN	Cancer/ angiogenesis	Cell-free high-throughput screening, VHTS	Athymic mice (s.c. PC-3 prostate and HT-29 colon cancer cells), HPLC	122,142, 143
Drug identification/IS	1-850, D1-D4	TR	Hyper-thyroidism	VHTS	HeLa, GH4 rat pituitary cells	124
Drug identification/IS	Indoloquina-zolinones 3a, 4, 5	CK2	Cancer	VHTS	Rat liver kinase inhibition assay	125
Drug mechanism/F	Wortmannin	1,067 on/off targets	Immune/ inflammatory disease, etc.	6,025-strain homozygous and heterozygous yeast-deletion collection	Phenotype complementation by ORF reintroduction	21
Drug mechanism/F	Rapamycin	TOR1p, TOR2p, 106 gene mutations	Cancer, immune disease	2,216 homozygous and 50 heterozygous yeast-deletion strains	Transfection	18
Drug mechanism/F	Artesunate	54 genes	Cancer	55 human tumour cell lines	Transduction (ΔEGFR, GCS, CDC25A)	144
Response factor/P	>70,000 anti-cancer drugs	Many on/off targets	Cancer	60 human cancer cell lines (NCI60)		70,71,74, 76–78

*The list of chemogenomics studies is not exhaustive but constitutes a representative compilation of recent, selected examples in this field. ACL, ATP citrate lyase; AGN, angiogenin; CaMPK, calmodulin-regulated protein kinase; CDC25A, cell-division cycle 25A phosphatase; CK2, protein kinase CK2 (casein kinase II); EGFR, epidermal growth factor; F, forward chemogenomics; GCS, γ-glutamylcysteine synthetase; HDAC, histone deacetylase; HPLC, high-performance liquid chromatography; Hsp90, heat-shock protein 90 molecular chaperone; IS, *in silico* chemogenomics; MetAP2, methionine aminopeptidase type 2; MLCK, myosin light chain kinase; ORF, open reading frame; P, predictive chemogenomics; Ref1, redox effector factor-1; pRB, retinoblastoma gene protein; R, reverse chemogenomics; TOR, target of rapamycin; TR, thyroid hormone receptor; UCH-L1, neuronal ubiquitin C-terminal hydrolase; VHTS, virtual high-throughput screening.

©2004 Nature Publishing Group

biological system, followed by an investigation into how these compounds interact with the drug targets. As a result of these approaches, the number of newly identified targets (and compounds) is steadily increasing (TABLE 2). One of the earliest successes of this method was the discovery of the immunosuppressant FK-506 (which has entered clinical practice as tracrolimus) in 1987 and the subsequent identification of calcineurin as its molecular target. The application of forward chemogenomics to cancer and angiogenesis research, for example, has identified the heat-shock protein 90 (HSP90) molecular chaperone and ATP citrate lyase (ACL) as targets of radicicol (*Humicola fuscoatra*, an antifungal antibiotic that has anti-tumour and anti-neoplastic activity) and its analogues KT8529, KF25706, BR-1 and BR-6 (REFS 32–34). Second, chemogenomics approaches are applied to discover, in a high-throughput fashion, new chemical candidates for molecular targets and phenotypes of interest. Hundreds of novel drug-like ligands have been identified for various molecular targets in the past few years through reverse, *in silico* and forward chemogenomics (TABLE 2). These targets relate to different disease areas, such as cancer, asthma, neuroinflammation and hyperthyroidism. Finally, chemogenomics can be used to understand the mechanism of drug action, which also includes finding genetic markers of drug susceptibility (TABLE 2). For example, the complex molecular effects of rapamycin (an antibiotic derived from *Streptomyces hygroscopicus* that has anti-fungal, anti-inflammatory, anti-tumour and immunosuppressive properties) and wortmannin (a fungal metabolite isolated from *Penicillium wortmanni* that has anti-neoplastic and radiosensitizing effects) were recently examined in a chemogenomics fashion using homozygous and heterozygous collections of the yeast *S. cerevisiae*: defined and known regions of the yeast genome had been deleted and targets were identified on the grounds that a yeast strain bearing a deletion in a gene that encodes a protein target is more resistant to treatment[18,21]. Similarly, the heterozygous yeast-deletion model has been recently used to identify gene products that functionally interact with various compounds, including anticancer, antifungal and anticholesterol agents[35]. The published data show signs of promise for all three aspects of chemogenomics.

Combining forward and reverse chemogenomics. Several studies stress the power of combining forward and reverse chemogenomics in concurrent target and drug discovery. As such, targets identified for lead compounds by phenotypic screening have been subsequently used to develop more potent synthetic analogues by HTS[20,36]. To take a specific example, sequential forward and reverse chemogenomics have led to the initial identification of histone deacetylase (HDAC) and the subsequent development of HDAC inhibitors, such as depsipeptide, sodium phenylbutyrate, CI-994 and suberoylanilide hydroxamic acid (SAHA)[37–39]. Clinical testing of these compounds has already begun in patients with advanced solid cancers, peripheral and cutaneous T-cell lymphoma, acute myeloid leukaemia and MYELODYSPLASTIC SYNDROME[40–46]. In turn, phenotypic screening of compounds that have been recently discovered by reverse chemogenomics to inhibit a certain biological pathway can be used to identify the immediate target of compound action in this pathway. For example, the redox effector factor-1 (*Ref1*) gene has been recently identified as a therapeutic target for asthma, following HTS for small-molecule inhibitors of activator protein-1 (AP1) transcription[47].

The potential of chemogenomics to identify novel compounds for many targets has stimulated particular interest in its application to cancer research. Compared with other diseases, cancers normally demonstrate complex genetic changes. These changes involve multiple alterations at the genetic and gene-expression levels that lead to aberrant protein (target) abundance. The genetic and epigenetic profile is highly variable even among cancers that seem to be histologically similar. In addition, genetic instability can cause substantial intratumour genetic heterogeneity, which further increases the number of molecular aberrations. This presents two difficult challenges: how to identify the enormous quantity of pathogenic changes (and therefore potential targets) that are present in tumours and how to identify the many targeted therapeutics that are necessary to address as many genetic alterations as possible. Chemogenomics might successfully rise to both these challenges and it is in fact in the field of cancer research that currently most chemogenomics studies have been done[17,28].

Chemogenomics and cancer research

Cancer research has been affected by genomics more than any other research discipline. As in other areas, the ultimate goal of applying chemogenomics strategies to cancer research is to increase the odds of advancing the right compounds to the clinic with reduced attrition rates and reduced costs of drug discovery. Probably the most extensive HTS attempt for novel anticancer drug and target profiling is the *in vitro* cell-line screening project (IVCLSP) that has been undertaken by the Developmental Therapeutics Program (DTP) of the National Cancer Institute (NCI). Here, a panel of 60 untreated cancer cell lines — the so-called 'NCI60' — is used for the drug-discovery screening of 140,000 compounds from the NCI chemical library[13,48,49]. The aim is to 'fingerprint' and prioritize compounds for further evaluation. The biological response patterns of the different cell lines are used in pattern-recognition algorithms, which allow putative mechanisms of action to be assigned to each compound. Potential targets are assessed *en masse*, allowing the activity patterns and DESCRIPTORS of molecular structure to be linked to genomic profiles. Compounds that are deemed to be sufficiently interesting in the screen are subjected to further experimental testing[13,48,49].

Chemogenomics-based anticancer drug discovery faces special challenges. An essential precursor to the development of such drugs is a comprehensive linking of specific genes to individual cancers and an explanation of the roles of their proteins, a project that is still far from complete[50]. It is unlikely that any one drug will be able to

MYELODYSPLASTIC SYNDROME
One of a group of disorders of the bone marrow that is characterized by the abnormal development of one or more of the cell lines that are normally found in bone marrow, leading to anaemia, abnormally low white blood cell count, tendency to infection and bleeding problems.

DESCRIPTOR
A metric that is used to numerically describe a structure or other molecular attributes of a chemical compound.

©2004 Nature Publishing Group

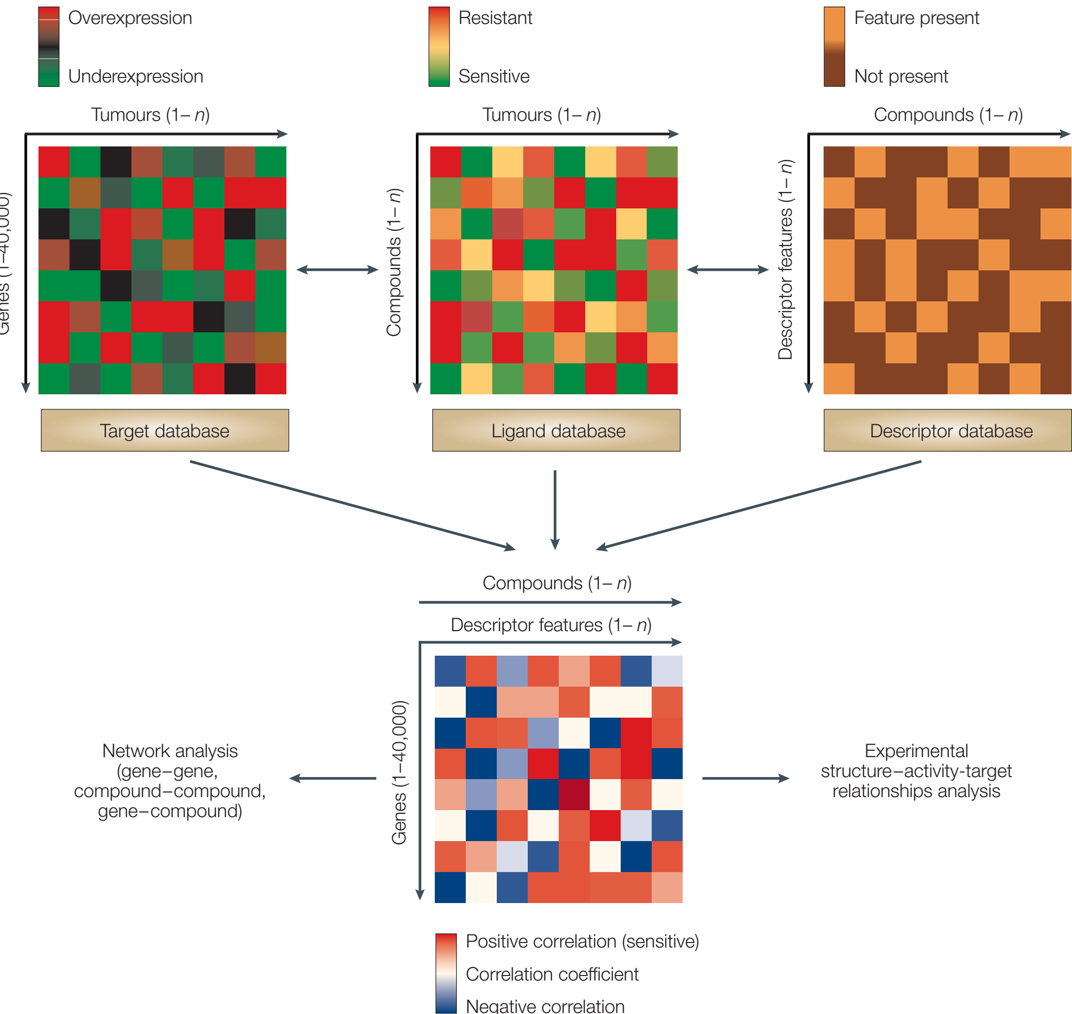


Figure 2 | **Predictive chemogenomics model.** The diagram shows an 'information-intensive approach' — originally proposed by Weinstein *et al.*[73] — to organize and inter-relate potential therapeutic targets, potential therapeutic drugs and molecular substructure classes. The principal aim of this approach is to identify relationships between the structural characteristics of compounds and the features of genes or gene products in cells (such as their patterns of expression) that might predict compound activities. Three databases (top row) are integrated to generate predictions about gene–compound relationships (bottom row): a target database of gene–disease correlations, a ligand database of compound activity–disease correlations and a descriptor database that includes compound–molecular structure correlations. Such predictive coupling of genomic and chemical information might allow genes that are selectively expressed in a tumour to be correlated not only with the compounds themselves but also with the subclasses and substructures of these compounds. This information can then be used to identify classes of compound for which detailed experimental structure–activity studies might be fruitful[78]. The Iconix Drugmatrix™ and the Cerep Discovery™ Bioprint matrix are two industrial implementations of the original National Cancer Institute (NCI) concept[117]: the Iconix Drugmatrix is a database that contains pharmacological and gene-expression profiles, as well as literature information of approximately 800 benchmark drugs, whereas the Cerep Bioprint matrix is a pharmacoinformatics system with a strong focus on G-protein-coupled receptor pharmacology and structure–activity data (see online links box).

modify cellular processes in such a way that it will cure most cancer diseases. Although cancers might initially arise as a result of a single-gene mutation, subclones form as the tumour progresses and additional genetic alterations accumulate[51]. Therefore, despite some published examples in which molecular correction of a single genetic alteration can bring about a therapeutic effect[52], the more likely situation is that a panel of drugs will be necessary to achieve a durable anticancer effect. This means that hundreds of small-molecule drugs with specific activity against the most prevalent genetic alterations will have to be developed. This, in turn, brings with it practical concerns related to cost implications.

Selective targeting of specific cancer alterations can meanwhile be a daunting task. For example, considerable work on kinase gene families[53] has led to the design of supposedly target-selective kinase inhibitors. However, all these inhibitors run the risk of interacting with additional targets. One way, albeit a costly one, to understand how these compounds interact with potentially many gene products would be to carry out chemogenomic profiling of candidate anticancer compounds against the

©2004 Nature Publishing Group

RELEVANCE NETWORK ANALYSIS
An analysis technique that is used to find functional genomic clusters by initially linking all genes in a data set by comprehensive pair-wise mutual information and then isolating clusters of genes by removing links that fall under a threshold.

primary therapeutic target, as well as against thousands of related targets. Such an approach should help to drive informed decisions on which compounds have the largest potential for therapeutic success.

Predictive chemogenomics in cancer research. Predictive chemogenomics strategies have been extensively used in cancer research to determine the relationship between genes, compounds and phenotypes. Using this approach, it has been possible to generate integrated databases of compound–gene interactions by experimental profiling[54–69], and libraries of thousands of anticancer compounds have been indirectly screened against the genomes of cancer cell lines. Such models have linked bio- and chemoinformatics by integrating databases that contain global gene-expression information and those that contain information on compound effectiveness, allowing compound–gene relationships to be inferred[70,71]. These models are grounded in the premise that similarities in cellular response patterns that are derived from genomic measurements and chemical screens represent associations between gene products and chemical activity. In other words, the relationship between patterns of gene expression and patterns of activity of compounds might provide incisive information on the mechanism of action of the compounds and on the molecular targets and modulators of activity in the cancer cells. In practical terms, initial experimental screening provides correlations between tumour–gene and tumour–compound pairs. *In silico* modelling then allows the correlative significance of individual gene-expression and compound-activity profiles to be ranked, and, finally, to identify candidate genes that are associated with compound susceptibility (FIG. 2).

A predictive chemogenomics model that is widely used in cancer research consists of the genomic profiles of the NCI60 (REF. 72) and the activity pattern of more than 70,000 potential anticancer compounds derived from the DTP database. This is an open repository of more than 600,000 compounds that consists of both synthetic and fully characterized, pure natural products; of these, sufficient quantities of approximately 140,000 compounds are available for the extramural discovery and development of new agents for the treatment of cancer, AIDS or opportunistic infections[73]. By identifying many gene–compound relationships[70,71], this model has shown that the integration of large genomic and molecular pharmacology databases might provide a valuable screen for predicting genes that underlie susceptibility to a particular compound. Additional sophisticated computational models have used this system to generate further regulatory genetic information[74–76]. For example, RELEVANCE NETWORK ANALYSIS, which allows individual genes to be directly or indirectly linked to other genes as well as to phenotypic measurements, has been used to deduce functional relationships between compound susceptibility and genomic profile in the NCI60 (REF. 76). Such relevance networks[75,76] offer a method to construct networks of similarity across disparate biological measures (FIG. 3). Here, so-called 'nodes' (for example, genes or compounds) with varying degrees of cross-connectivity are displayed, which represent features that are not only associated pair-wise, but also in aggregate.

Predictive chemogenomic approaches in the NCI60 have been recently extended to include structure-based descriptors[77]. The coupling of gene-expression levels, corresponding compound activity values and binary indices of the occurrence of thousands of compound substructures and chemical descriptor features has identified classes of compound for which detailed experimental studies might be fruitful[78].

Implicit in the goal of predictive chemogenomics is the notion that the activity of a potential anticancer compound can be predicted by genome-wide gene-expression patterns. The above studies lend support to the premise that connections can be created between genomic data sets and pharmacology data sets that allow gene–compound relationships to be easily explored and genomic data to be linked to the drug-discovery and development processes. However, it is widely appreciated that protein levels can vary significantly among genes that have similar mRNA-expression profiles and, in turn, that there can be a significant variance in the mRNA levels of proteins that are expressed with comparable abundance[79]; this means that the linking of genomic and compound activity–structure data sets

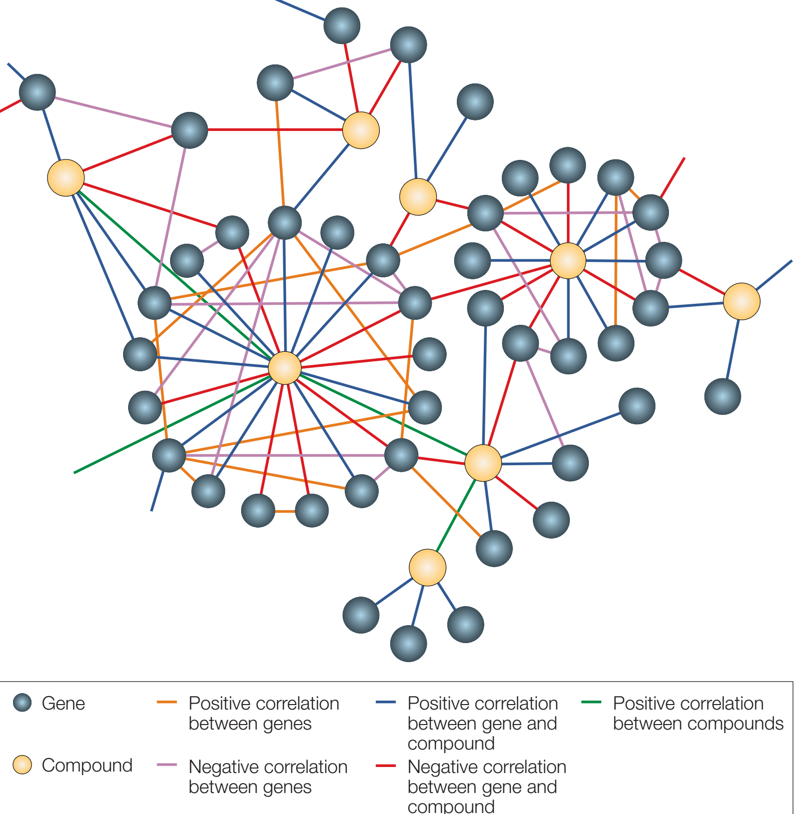


Figure 3 | **Relevance network between genome and drug effect.** A gene regulatory network of compound susceptibility — a chemical genetics network — can be constructed from the integrated databases in FIG. 2; that is, by incorporating baseline target expression and measures of drug activity in the same study system. Feature pairs include gene–compound, gene–gene and compound–compound relations.

©2004 Nature Publishing Group

might fail to extract meaningful information for drug discovery and for predicting the susceptibility of a cell or an organism to a compound.

Challenges and limitations

Chemogenomics is one example of the many innovative platforms that pharmaceutical and biotechnology research have generated to accelerate the pace of drug discovery. Such combined efforts have been motivated by the assumption that the judicious application of genomics can help to improve efficiencies throughout the drug-discovery process — for example, by identifying candidate failures early in the process — before moving into expensive later-phase trials.

However, the decreased output of commercialized drugs during the past few years indicates that genomics might currently be doing more to hinder drug-discovery programmes than to stimulate them. The sheer volume of data that are generated by chemogenomic analyses creates a gap between drug discovery and drug development. Considering that an important driving concept of chemogenomics is to reduce the time taken for drug development, it is under tremendous pressure to deliver some valid results quickly. Therefore, current chemogenomics programmes are anxiously seeking to accelerate their throughput at every stage of the drug-development process.

Hit selection and validation. HTS often generates an enormous number of hits. In addition, HTS data are noisy and error prone and include a substantial portion of false-positive and false-negative hits. This presents two main challenges: how to mine the large data sets to identify those hits with the greatest potential as leads and how to weed out false-positive hits. It is generally agreed that the bottleneck that is created by genomics-based drug discovery occurs not because the process fails to identify hits but because of the slow process of optimizing and validating them — that is, it is improving their ADMET PROPERTIES[37] and determining whether they can convincingly reverse or ameliorate a disease state[80]. Negligent hit validation often occurs at the expense of taking compounds through development, only to later fall short of success. Validation is not always straightforward, however, as sufficient biological information about many targets is often lacking and, in turn, for many new targets with true disease impact, leads cannot be identified or optimized because of the increasing sizes and concurrent stagnating diversities and complexities of chemical libraries.

The design, adaptation and refinement of sophisticated 'front-end' computational approaches that can assist in making an informed, quick decision as to which hits should be pursued will be crucial for the success of chemogenomics. Statistical and chemoinformatics approaches for assessing the quality control of HTS data and for mining their chemical and biological information have been developed, for example, by incorporating pattern-detection methods for the identification of pipetting artefacts or for the detection of chemical-class-related effects. The development of chemoinformatics methods and procedures, such as RECURSIVE PARTITIONING, PHYLOGENETIC-LIKE TREE ALGORITHMS or BINARY QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIPS (QSARs)[81–87], which support the automatic identification of hits that are frequently identified by HTS, false positives and negatives, as well as structure–activity relationship (SAR) information, is essential for generating knowledge from HTS data[83,88–90]. The myriad efforts that surround the design of appropriate analytic tools have to cope with the difficulty of integrating disparate types of information, especially PARSING and assimilating both chemical and genomics information data (tools such as Scitegic Pipeline pilot or Kensington Inforsense provide the required integration concepts (see online links box)).

Data integration. Chemogenomics-orientated drug-discovery programmes face many data-management challenges, which have partly been met by the recent development of innovative 'biochemoinformatics' (or pharmacoinformatics) platforms[91]. These INTEGRATION PLATFORMS aim to collect, store and disseminate diverse data sets — such as combinatorial library diversity data, small-molecule chemical structure and biomolecule protein structure data, annotation information, HTS bioassay data and imaging data — in as efficient and productive a manner as possible and to maximize the potential of these data sets.

In particular, integrating information about ligands and targets is complex. This process is related to the current bioinformatics project to create ontologies for proteomics, the aim of which is to generate a systematic definition of the structure and function of all proteins in a genome[92,93]. The key difficulty lies in integrating two ontologies — one for protein structure and the other for protein function — that have been developed separately and remain largely isolated[92]. The description of active sites and binding sites in protein structures is recognized here as one potential connection point that describes the protein function. Classifications that are based on molecular interactions[92,94], in which each protein is associated with a vector that consists of the probability of binding to various ligands — the central chemogenomics idea — might therefore become prominent in the future[94]. The emphasis on protein-structure similarity is recognized here as a guiding principle[95]. The establishment of standardized molecular informatics platforms and real drug-discovery ontologies at the genome level — that integrate the relevant chemical and biological knowledge — is therefore pursued by the academic and industrial drug-discovery organizations and by companies that are involved in informatics-based discovery[96–98].

Knowledge-based chemogenomics companies are currently developing comprehensive molecular information systems for several target classes, including G-protein-coupled receptors, kinases, ion channels and proteases. Their main contribution has been to integrate, in a comprehensive manner, data from patents and selected literature, including two-dimensional structures of the ligands, target sequence and classification, mechanisms of action, structure–activity data, assay results and bibliographic information, together with chemical

ADMET PROPERTIES [AND MECHANISMS]
The absorption, distribution, metabolism, excretion and toxicity (ADMET) are fundamental pharmacokinetic properties that determine the *in vivo* efficacy of a drug, together with its intrinsic biological activity on the target.

RECURSIVE PARTITIONING
A process for identifying complex structure–activity relationships in large sets by dividing compounds into a hierarchy of smaller and more homogeneous subgroups on the basis of the statistically most significant descriptor, such as structure fragments.

PHYLOGENETIC-LIKE TREE ALGORITHM
A method for analysing a data set of molecules that assists in identifying chemical classes of interest and sets of molecular features that correlate with a specified biological feature by combining elements of neural nets, genetic algorithms and substructure analysis.

BINARY QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIP
A method to assign probabilities of activity to compounds by establishing associations between structural features and molecular properties of these compounds and their biological activities.

PARSING
A process by which programming data is broken into smaller, more distinct chunks of information that can be more easily interpreted and acted on.

INTEGRATION PLATFORMS
A software system that connects different data domains and analysis applications under one graphical user interface.

©2004 Nature Publishing Group

and biological search engines. Further academic and commercial programmes are gathering information on other types of target, such as those involved in adverse reactions[99,100], in ADMET mechanisms and those that define metabolic and signalling pathways[101,102].

Conclusions

How effective chemogenomics will be at generating new treatments across various disease areas remains to be seen. The explosion in the data that has been made available by chemogenomics is yet to create a commensurate increase in the efficacy of the drug-discovery process. Refined chemogenomics strategies will require a better integration of data by new informatics tools and more thorough validation of hits. Beyond proper biological mechanistic target validation and progress in data mining, the challenge reverts to drug-discovery chemistry to create diversity-enriched compound libraries, specifically to provide ligands for the genetic sequences with which their biological and pharmacological impact can be systematically examined. With a recent trend in chemogenomics to focus on data quality rather than on the number of data points that can be generated, judicious selection of compounds for screening will become increasingly important; this will dictate the extent of screening activity and the value of the hits — both crucial variables of cost-efficient chemogenomics. Although powerful bio- and chemoinformatics tools might be developed in the future to provide the desired link between genomic and chemical space, the success of chemogenomics will not only depend very much on progress in overcoming current methodological obstacles but also on what can realistically be done within the limits of available budgets. Finally, supporters of chemogenomics as a means to revolutionize disease treatment will ultimately need to face the fact that — as with traditional drug-development approaches — any discovered targeted therapeutic candidate for clinical use must undergo rigorous testing in clinical trials, which can be prohibitively expensive.

1. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nature Rev. Drug Discov.* **1**, 727–730 (2002).
2. Drews, J. Genomic sciences and the medicine of tomorrow. *Nature Biotechnol.* **14**, 1516–1518 (1996).
3. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
4. Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
5. Kim, J. A. Targeted therapies for the treatment of cancer. *Am. J. Surg.* **186**, 264–268 (2003).
6. Wolcke, J. & Ullmann, D. Miniaturized HTS technologies — uHTS. *Drug Discov. Today* **6**, 637–646 (2001).
7. Stockwell, B. R., Haggarty, S. J. & Schreiber, S. L. High-throughput screening of small molecules in miniaturized mammalian cell-based assays involving post-translational modifications. *Chem. Biol.* **6**, 71–83 (1999).
8. Stockwell, B. R. Chemical genetics: ligand-based discovery of gene function. *Nature Rev. Genet.* **1**, 116–125 (2000).
9. Stockwell, B. R. Frontiers in chemical genetics. *Trends Biotechnol.* **18**, 449–455 (2000).
10. Walters, W. P. & Namchuk, M. Designing screens: how to make your hits a hit. *Nature Rev. Drug Discov.* **2**, 259–266 (2003).
11. Entzeroth, M. Emerging trends in high-throughput screening. *Curr. Opin. Pharmacol.* **3**, 522–529 (2003).
12. Muller, O. *et al.* Identification of potent Ras signaling inhibitors by pathway-selective phenotype-based screening. *Angew. Chem. Int. Ed. Engl.* **43**, 450–454 (2004).
13. Shoemaker, R. H. *et al.* Application of high-throughput, molecular-targeted screening to anticancer drug discovery. *Curr. Top. Med. Chem.* **2**, 229–246 (2002).
14. Pantoliano, M. W. *et al.* High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screen.* **6**, 429–440 (2001).
15. Shin, Y. G. & van Breemen, R. B. Analysis and screening of combinatorial libraries using mass spectrometry. *Biopharm. Drug Dispos.* **22**, 353–372 (2001).
16. Muckenschnabel, I., Falchetto, R., Mayr, L. M. & Filipuzzi, I. SpeedScreen: label-free liquid chromatography-mass spectrometry-based high-throughput screening for the discovery of orphan protein ligands. *Anal. Biochem.* **324**, 241–249 (2004).
17. Chan, T. F., Carvalho, J., Riles, L. & Zheng, X. F. A chemical genomics approach toward understanding the global functions of the target of rapamycin protein (TOR). *Proc. Natl Acad. Sci. USA* **97**, 13227–13232 (2000).
18. Griffith, E. C. *et al.* Methionine aminopeptidase (type 2) is the common target for angiogenesis inhibitors AGM-1470 and ovalicin. *Chem. Biol.* **4**, 461–471 (1997).
19. Sin, N. *et al.* The anti-angiogenic agent fumagillin covalently binds and inhibits the methionine aminopeptidase, MetAP-2. *Proc. Natl Acad. Sci. USA* **94**, 6099–6103 (1997).
20. Zewail, A. *et al.* Novel functions of the phosphatidylinositol metabolic pathway discovered by a chemical genomics screen with wortmannin. *Proc. Natl Acad. Sci. USA* **100**, 3345–3350 (2003).
21. Marton, M. J. *et al.* Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nature Med.* **4**, 1293–1301 (1998).
Describes an approach based on genome-wide gene-expression patterns to identify the immediate pathways that are altered by a drug and to detect drug effects that are mediated by unintended targets.
22. Dolma, S., Lessnick, S. L., Hahn, W. C. & Stockwell, B. R. Identification of genotype-selective antitumor agents using synthetic lethal chemical screening in engineered human tumor cells. *Cancer Cell* **3**, 285–296 (2003).
23. Kwon, H. J. Chemical genomics-based target identification and validation of anti-angiogenic agents. *Curr. Med. Chem.* **10**, 717–736 (2003).
24. Carr, R. & Jhoti, H. Structure-based screening of low-affinity compounds. *Drug Discov. Today* **7**, 522–527 (2002).
25. Bleicher, K. H. Chemogenomics: bridging a drug discovery gap. *Curr. Med. Chem.* **9**, 2077–2084 (2002).
26. Jung, M., Kim, H. & Kim, M. Chemical genomics strategy for the discovery of new anticancer agents. *Curr. Med. Chem.* **10**, 757–762 (2003).
27. Goodnow, R. A. Jr, Guba, W. & Haap, W. Library design practices for success in lead generation with small molecule libraries. *Comb. Chem. High Throughput Screen.* **6**, 649–660 (2003).
Summary of current practices in the design of compound libraries for use in drug discovery, focusing on the generation of novel structures that are amenable to rapid and efficient lead optimization.
28. Root, D. E., Flaherty, S. P., Kelley, B. P. & Stockwell, B. R. Biological mechanism profiling using an annotated compound library. *Chem. Biol.* **10**, 881–892 (2003).
Annotated compound libraries with known biological activity are introduced to guide experiments for pathway elucidation. Algorithms for the build-up of annotations from Medline reports and for scoring statistically enriched mechanisms are also described.
29. Eguchi, M. *et al.* Chemogenomics with peptide secondary structure mimetics. *Comb. Chem. High Throughput Screen.* **6**, 611–621 (2003).
30. Abel, U., Koch, C., Speitling, M. & Hansske, F. G. Modern methods to produce natural-product libraries. *Curr. Opin. Chem. Biol.* **6**, 453–458 (2002).
31. Burke, M. D. & Schreiber, S. L. A planning strategy for diversity-oriented synthesis. *Angew. Chem. Int. Ed. Engl.* **43**, 46–58 (2004).
32. Ki, S. W. *et al.* Radicicol binds and inhibits mammalian ATP citrate lyase. *J. Biol. Chem.* **275**, 39231–39236 (2000).
33. Sharma, S. V., Agatsuma, T. & Nakano, H. Targeting of the protein chaperone, HSP90, by the transformation suppressing agent, radicicol. *Oncogene* **16**, 2639–2645 (1998).
34. Soga, S. *et al.* KF25706, a novel oxime derivative of radicicol, exhibits *in vivo* antitumor activity via selective depletion of Hsp90 binding signaling molecules. *Cancer Res.* **59**, 2931–2938 (1999).
35. Giaever, G. *et al.* Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc. Natl Acad. Sci. USA* **101**, 793–798 (2004).
Demonstrates the efficacy of a genome-wide protocol in yeast that allows the identification of gene products that functionally interact as on- or off-targets with small molecules, thereby allowing an understanding of the *in vivo* response to small-molecule perturbants.
36. Han, C. K. *et al.* Design and synthesis of highly potent fumagillin analogues from homology modeling for a human MetAP-2. *Bioorg. Med. Chem. Lett.* **10**, 39–43 (2000).
37. Ueda, H., Nakajima, H., Hori, Y., Goto, T. & Okuhara, M. Action of FR901228, a novel antitumor bicyclic depsipeptide produced by *Chromobacterium violaceum* no. 968, on Ha-ras transformed NIH3T3 cells. *Biosci. Biotechnol. Biochem.* **58**, 1579–1583 (1994).
38. Yoshida, M., Kijima, M., Akita, M. & Beppu, T. Potent and specific inhibition of mammalian histone deacetylase both *in vivo* and *in vitro* by trichostatin A. *J. Biol. Chem.* **265**, 17174–17179 (1990).
39. Kwon, H. J., Owa, T., Hassig, C. A., Shimada, J. & Schreiber, S. L. Depudecin induces morphological reversion of transformed fibroblasts via the inhibition of histone deacetylase. *Proc. Natl Acad. Sci. USA* **95**, 3356–3361 (1998).
40. Sandor, V. *et al.* Phase I trial of the histone deacetylase inhibitor, depsipeptide (FR901228, NSC 630176), in patients with refractory neoplasms. *Clin. Cancer Res.* **8**, 718–728 (2002).
41. Piekarz, R. L. *et al.* Inhibitor of histone deacetylation, depsipeptide (FR901228), in the treatment of peripheral and cutaneous T-cell lymphoma: a case report. *Blood* **98**, 2865–2868 (2001).
42. Marshall, J. L. *et al.* A phase I trial of depsipeptide (FR901228) in patients with advanced cancer. *J. Exp. Ther. Oncol.* **2**, 325–332 (2002).
43. Carducci, M. A. *et al.* A Phase I clinical and pharmacological evaluation of sodium phenylbutyrate on an 120-h infusion schedule. *Clin. Cancer Res.* **7**, 3047–3055 (2001).
44. Gore, S. D. *et al.* Impact of prolonged infusions of the putative differentiating agent sodium phenylbutyrate on myelodysplastic syndromes and acute myeloid leukemia. *Clin. Cancer Res.* **8**, 963–970 (2002).
45. Nemunaitis, J. J. *et al.* Phase I study of oral CI-994 in combination with gemcitabine in treatment of patients with advanced cancer. *Cancer J.* **9**, 58–66 (2003).
46. Kelly, W. K. *et al.* Phase I clinical trial of histone deacetylase inhibitor: suberoylanilide hydroxamic acid administered intravenously. *Clin. Cancer Res.* **9**, 3578–3588 (2003).
47. Nguyen, C. *et al.* Chemogenomic identification of Ref-1/AP-1 as a therapeutic target for asthma. *Proc. Natl Acad. Sci. USA* **100**, 1169–1173 (2003).
48. Weinstein, J. N. & Buolamwini, J. K. Molecular targets in cancer drug discovery: cell-based profiling. *Curr. Pharm. Des.* **6**, 473–483 (2000).
49. Monks, A. *et al.* Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines. *J. Natl Cancer Inst.* **83**, 757–766 (1991).

©2004 Nature Publishing Group

50. Zheng, X. F. & Chan, T. F. Chemical genomics in the global study of protein functions. *Drug Discov. Today* **7**, 197–205 (2002).
51. Vogelstein, B. & Kinzler, K. W. The multistep nature of cancer. *Trends Genet.* **9**, 138–141 (1993).
52. Weinstein, I. B. Cancer. Addiction to oncogenes — the Achilles heal of cancer. *Science* **297**, 63–64 (2002).
53. Robinson, D. R., Wu, Y. M. & Lin, S. F. The protein tyrosine kinase family of the human genome. *Oncogene* **19**, 5548–5557 (2000).
54. Dan, S. *et al.* An integrated database of chemosensitivity to 55 anticancer drugs and gene expression profiles of 39 human cancer cell lines. *Cancer Res.* **62**, 1139–1147 (2002).
55. Zembutsu, H. *et al.* Genome-wide cDNA microarray screening to correlate gene expression profiles with sensitivity of 85 human cancer xenografts to anticancer drugs. *Cancer Res.* **62**, 518–527 (2002).
56. Sotiriou, C. *et al.* Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer. *Breast Cancer Res.* **4**, R3 (2002).
57. Wittig, R. *et al.* Candidate genes for cross-resistance against DNA-damaging drugs. *Cancer Res.* **62**, 6698–6705 (2002).
58. Hoshida, Y. *et al.* Identification of genes associated with sensitivity to 5-fluorouracil and cisplatin in hepatoma cells. *J. Gastroenterol.* **37** (Suppl. 14), 92–95 (2002).
59. Kihara, C. *et al.* Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. *Cancer Res.* **61**, 6474–6479 (2001).
60. Komatani, H. *et al.* Identification of breast cancer resistant protein/mitoxantrone resistance/placenta-specific, ATP-binding cassette transporter as a transporter of NB-506 and J-107088, topoisomerase I inhibitors with an indolocarbazole structure. *Cancer Res.* **61**, 2827–2832 (2001).
61. Levenson, V. V., Davidovich, I. A. & Roninson, I. B. Pleiotropic resistance to DNA-interactive drugs is associated with increased expression of genes involved in DNA replication, repair, and stress response. *Cancer Res.* **60**, 5027–5030 (2000).
62. Sakamoto, M. *et al.* Analysis of gene expression profiles associated with cisplatin resistance in human ovarian cancer cell lines and tissues using cDNA microarray. *Hum. Cell* **14**, 305–315 (2001).
63. Turton, N. J. *et al.* Gene expression and amplification in breast carcinoma cells with intrinsic and acquired doxorubicin resistance. *Oncogene* **20**, 1300–1306 (2001).
64. Vikhanskaya, F., Marchini, S., Marabese, M., Galliera, E. & Broggini, M. p73a overexpression is associated with resistance to treatment with DNA-damaging agents in a human ovarian cancer cell line. *Cancer Res.* **61**, 935–938 (2001).
65. Weldon, C. B. *et al.* Identification of mitogen-activated protein kinase kinase as a chemoresistant pathway in MCF-7 cells by using gene expression microarray. *Surgery* **132**, 293–301 (2002).
66. Watts, G. S. *et al.* cDNA microarray analysis of multidrug resistance: doxorubicin selection produces multiple defects in apoptosis signaling pathways. *J. Pharmacol. Exp. Ther.* **299**, 434–441 (2001).
67. Duan, Z., Feller, A. J., Penson, R. T., Chabner, B. A. & Seiden, M. V. Discovery of differentially expressed genes associated with paclitaxel resistance using cDNA array technology: analysis of interleukin (IL) 6, IL-8, and monocyte chemotactic protein 1 in the paclitaxel-resistant phenotype. *Clin. Cancer Res.* **5**, 3445–3453 (1999).
68. Kudoh, K. *et al.* Monitoring the expression profiles of doxorubicin-induced and doxorubicin-resistant cancer cells by cDNA microarray. *Cancer Res.* **60**, 4161–4166 (2000).
69. Maxwell, P. J. *et al.* Identification of 5-fluorouracil-inducible target genes using cDNA microarray profiling. *Cancer Res.* **63**, 4602–4606 (2003).
70. Scherf, U. *et al.* A gene expression database for the molecular pharmacology of cancer. *Nature Genet.* **24**, 236–244 (2000).
This paper has pioneered the large-scale analysis of the relationship between patterns of gene expression and patterns of anticancer drug activity, thereby directly linking bioinformatics and chemoinformatics.
71. Staunton, J. E. *et al.* Chemosensitivity prediction by transcriptional profiling. *Proc. Natl Acad. Sci. USA* **98**, 10787–10792 (2001).
72. Ross, D. T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genet.* **24**, 227–235 (2000).
73. Weinstein, J. N. *et al.* An information-intensive approach to the molecular pharmacology of cancer. *Science* **275**, 343–349 (1997).
74. Bao, L., Guo, T. & Sun, Z. Mining functional relationships in feature subspaces from gene expression profiles and drug activity profiles. *FEBS Lett.* **516**, 113–118 (2002).
75. Moriyama, M. *et al.* Relevance network between chemosensitivity and transcriptome in human hepatoma cells. *Mol. Cancer Ther.* **2**, 199–205 (2003).
76. Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R. & Kohane, I. S. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc. Natl Acad. Sci. USA* **97**, 12182–12186 (2000).
Describes a method that allows a gene to be linked to several other genes as well as to phenotypic measures of drug susceptibility through the construction of a chemical genetic network.
77. Wallqvist, A., Rabow, A. A., Shoemaker, R. H., Sausville, E. A. & Covell, D. G. Establishing connections between microarray expression data and chemotherapeutic cancer pharmacology. *Mol. Cancer Ther.* **1**, 311–320 (2002).
78. Blower, P. E. *et al.* Pharmacogenomic analysis: correlating molecular substructure classes with microarray gene expression data. *Pharmacogenomics J.* **2**, 259–271 (2002).
The authors present a model that allows systematic and fluent exploration of the relationships between the structural features of a compound and global gene expression.
79. Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* **19**, 1720–1730 (1999).
80. Bleicher, K. H., Bohm, H. J., Muller, K. & Alanine, A. I. Hit and lead generation: beyond high-throughput screening. *Nature Rev. Drug Discov.* **2**, 369–378 (2003).
Describes the process from hit to lead generation in a modern industrial setup and focuses on the role of quality versus quantity to improve the attrition rates at later stages of drug development.
81. Cramer, R. D., Patterson, D. E. & Bunce, J. D. Recent advances in comparative molecular field analysis (CoMFA). *Prog. Clin. Biol. Res.* **291**, 161–165 (1989).
82. Rusinko, A., Farmen, M. W., Lambert, C. G., Brown, P. L. & Young, S. S. Analysis of a large structure/biological activity data set using recursive partitioning. *J. Chem. Inf. Comput. Sci.* **39**, 1017–1026 (1999).
83. Nicolaou, C. A., Tamura, S. Y., Kelley, B. P., Bassett, S. I. & Nutt, R. F. Analysis of large screening data sets via adaptively grown phylogenetic-like trees. *J. Chem. Inf. Comput. Sci.* **42**, 1069–1079 (2002).
84. Labute, P. Binary QSAR: a new method for the determination of quantitative structure activity relationships. *Pac. Symp. Biocomput*. 444–455 (1999).
85. Stanton, D. T., Morris, T. W., Roychoudhury, S. & Parker, C. N. Application of nearest-neighbor and cluster analyses in pharmaceutical lead discovery. *J. Chem. Inf. Comput. Sci.* **39**, 21–27 (1999).
86. Hopfinger, A. J. & Duca, J. S. Extraction of pharmacophore information from high-throughput screens. *Curr. Opin. Biotechnol.* **11**, 97–103 (2000).
87. Gedeck, P. & Willett, P. Visual and computational analysis of structure–activity relationships in high-throughput screening data. *Curr. Opin. Chem. Biol.* **5**, 389–395 (2001).
88. Engels, M. F. Creating knowledge from high-throughput screening data. *Ernst Schering Res. Found. Workshop*, 87–101 (2003).
89. Raymond, J. W. & Willett, P. Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *J. Comput. Aided Mol. Des.* **16**, 521–533 (2002).
90. Roberts, G., Myatt, G. J., Johnson, W. P., Cross, K. P. & Blower, P. E. Jr. LeadScope: software for exploring large sets of screening data. *J. Chem. Inf. Comput. Sci.* **40**, 1302–1314 (2000).
91. Agrafiotis, D. K., Lobanov, V. S. & Salemme, F. R. Combinatorial informatics in the post-genomics ERA. *Nature Rev. Drug Discov.* **1**, 337–346 (2002).
92. Lan, N., Montelione, G. T. & Gerstein, M. Ontologies for proteomics: towards a systematic definition of structure and function that scales to the genome level. *Curr. Opin. Chem. Biol.* **7**, 44–54 (2003).
93. Stevens, R., Goble, C. A. & Bechhofer, S. Ontology-based knowledge representation for bioinformatics. *Brief. Bioinform.* **1**, 398–414 (2000).
94. Karp, P. D. An ontology for biological function based on molecular interactions. *Bioinformatics* **16**, 269–285 (2000).
95. Koch, M. A., Breinbauer, R. & Waldmann, H. Protein structure similarity as guiding principle for combinatorial library design. *Biol. Chem.* **384**, 1265–1272 (2003).
96. Glen, R. Developing tools and standards in molecular informatics. Interview by Susan Aldridge. *Chem. Commun. (Camb.)* 2745–2747 (2002).
97. Strausberg, R. L. & Schreiber, S. L. From knowing to controlling: a path from genomics to drugs using small molecule probes. *Science* **300**, 294–295 (2003).
98. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
99. Ji, Z. L. *et al.* Drug Adverse Reaction Target Database (DART) : proteins related to adverse drug reactions. *Drug Saf.* **26**, 685–690 (2003).
100. Sun, L. Z., Ji, Z. L., Chen, X., Wang, J. F. & Chen, Y. Z. ADME-AP: a database of ADME associated proteins. *Bioinformatics* **18**, 1699–1700 (2002).
101. Ji, Z. L. *et al.* Internet resources for proteins associated with drug therapeutic effects, adverse reactions and ADME. *Drug Discov. Today* **8**, 526–529 (2003).
102. Chen, X., Ji, Z. L. & Chen, Y. Z. TTD: Therapeutic Target Database. *Nucleic Acids Res.* **30**, 412–415 (2002).
103. Caron, P. R. *et al.* Chemogenomic approaches to drug discovery. *Curr. Opin. Chem. Biol.* **5**, 464–470 (2001).
104. Mitchison, T. J. Towards a pharmacological genetics. *Chem. Biol.* **1**, 3–6 (1994).
105. Zheng, X. F. & Chan, T. F. Chemical genomics: a systematic approach in biological research and drug discovery. *Curr. Issues Mol. Biol.* **4**, 33–43 (2002).
106. Sehgal, A. Drug discovery and development using chemical genomics. *Curr. Opin. Drug Discov. Devel.* **5**, 526–531 (2002).
107. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85**, 2444–24448 (1988).
108. Duckworth, D. M. & Sanseau, P. *In silico* identification of novel therapeutic targets. *Drug Discov. Today* **7**, S64–S69 (2002).
109. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
110. Lehmann, J. *et al.* Redesigning drug discovery. *Nature* **384** (Suppl.), 1–5 (1996)
111. Murcko, M. & Caron, P. Transforming the genome to drug discovery. *Drug Discov. Today* **7**, 583–584 (2002).
112. Schuffenhauer, A., Floersheim, P., Acklin, P. & Jacoby, E. Similarity metrics for ligands reflecting the similarity of the target proteins. *J. Chem. Inf. Comput. Sci.* **43**, 391–405 (2003).
113. Jacoby, E. A novel chemogenomics knowledge-based ligand design strategy — application to G protein-coupled receptors. *Quantitative Structure–Activity Relationships* **20**, 115–123 (2001).
114. New, D. C., Miller-Martini, D. M. & Wong, Y. H. Reporter gene assays and their applications to bioassays of natural products. *Phytother. Res.* **17**, 439–448 (2003).
115. Boland, M. V. & Murphy, R. F. Automated analysis of patterns in fluorescence-microscope images. *Trends Cell Biol.* **9**, 201–202 (1999).
116. Frye, S. V. Structure–activity relationship homology (SARAH): a conceptual framework for drug discovery in the genomic era. *Chem. Biol.* **6**, R3–R7 (1999).
This fundamental article introduces the 'structure–activity relationship homology' concept, which is the baseline for carrying out target-family reverse chemogenomics.
117. Krejsa, C. M. *et al.* Predicting ADME properties and side effects: the BioPrint approach. *Curr. Opin. Drug Discov. Devel.* **6**, 470–480 (2003).
118. Bajorath, J. Integration of virtual and high-throughput screening. *Nature Rev. Drug Discov.* **1**, 882–894 (2002).
This review article covers the current concepts that are involved in integrating both virtual and high-throughput screening.
119. Langer, T. & Krovat, E. M. Chemical feature-based pharmacophores and virtual library screening for discovery of new leads. *Curr. Opin. Drug Discov. Devel.* **6**, 370–376 (2003).
120. Xue, L. & Bajorath, J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. *Comb. Chem. High Throughput Screen.* **3**, 363–372 (2000).
121. Glen, R. C. & Allen, S. C. Ligand-protein docking: cancer research at the interface between biology and chemistry. *Curr. Med. Chem.* **10**, 763–767 (2003).
122. Abagyan, R. & Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **5**, 375–382 (2001).
123. Jenkins, J. L. & Shapiro, R. Identification of small-molecule inhibitors of human angiogenin and characterization of their binding interactions guided by computational docking. *Biochemistry* **42**, 6674–6687 (2003).
124. Schapira, M. *et al.* Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking. *Proc. Natl Acad. Sci. USA* **100**, 7354–7359 (2003).

©2004 Nature Publishing Group

125. Vangrevelinghe, E. *et al.* Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. *J. Med. Chem.* **46**, 2656–2662 (2003).
126. Schneider, G. & Bohm, H. J. Virtual screening and fast automated docking methods. *Drug Discov. Today* **7**, 64–70 (2002).
127. Lamb, M. L. *et al.* Design, docking, and evaluation of multiple libraries against multiple targets. *Proteins* **42**, 296–318 (2001).
128. Chen, X., Ung, C. Y. & Chen, Y. Can an *in silico* drug-target search method be used to probe potential mechanisms of medicinal plant ingredients? *Nat. Prod. Rep.* **20**, 432–44 (2003).
Describes the application of high-throughput docking for identifying drug targets in an automated fashion. The entire PDB protein structure repertoire is docked against selected natural products. Predictions for their mechanism of action are successfully made.
129. Gunther, J., Bergner, A., Hendlich, M. & Klebe, G. Utilising structural knowledge in drug design strategies: applications using Relibase. *J. Mol. Biol.* **326**, 621–636 (2003).
130. Michalovich, D., Overington, J. & Fagan, R. Protein sequence analysis *in silico*: application of structure-based bioinformatics to genomic initiatives. *Curr. Opin. Pharmacol.* **2**, 574–580 (2002).
131. Jacoby, E., Schuffenhauer, A. & Floersheim, P. Chemogenomics knowledge-based strategies in drug discovery. *Drug News Perspect.* **16**, 93–102 (2003).
132. Schuffenhauer, A. *et al.* An ontology for pharmaceutical ligands and its application for *in silico* screening and library design. *J. Chem. Inf. Comput. Sci.* **42**, 947–955 (2002).
Provides a foundation for linking the fields of chemoinformatics and bioinformatics by establishing ligand-discovery ontologies. The application for similarity searching and focused library design are highlighted.
133. Meng, L., Kwok, B. H., Sin, N. & Crews, C. M. Eponemycin exerts its antitumor effect through the inhibition of proteasome function. *Cancer Res.* **59**, 2798–2801 (1999).
134. Kino, T. *et al.* FK-506, a novel immunosuppressant isolated from a *Streptomyces*. II. Immunosuppressive effect of FK-506 *in vitro*. *J. Antibiot. (Tokyo)* **40**, 1256–1265 (1987).
135. Kino, T. *et al.* FK-506, a novel immunosuppressant isolated from a *Streptomyces*. I. Fermentation, isolation, and physico-chemical and biological characteristics. *J. Antibiot. (Tokyo)* **40**, 1249–1255 (1987).
136. Mirzoeva, S. *et al.* Screening in a cell-based assay for inhibitors of microglial nitric oxide production reveals calmodulin-regulated protein kinases as potential drug discovery targets. *Brain Res.* **844**, 126–134 (1999).
137. Barrie, S. E. *et al.* High-throughput screening for the identification of small-molecule inhibitors of retinoblastoma protein phosphorylation in cells. *Anal. Biochem.* **320**, 66–74 (2003).
138. Lukas, T. J., Mirzoeva, S., Slomczynska, U. & Watterson, D. M. Identification of novel classes of protein kinase inhibitors using combinatorial peptide chemistry based on functional genomics knowledge. *J. Med. Chem.* **42**, 910–919 (1999).
139. Liu, Y. *et al.* Discovery of inhibitors that elucidate the role of UCH-L1 activity in the H1299 lung cancer cell line. *Chem. Biol.* **10**, 837–846 (2003).
140. Wittich, S. *et al.* Structure–activity relationships on phenylalanine-containing inhibitors of histone deacetylase: *in vitro* enzyme inhibition, induction of differentiation, and inhibition of proliferation in Friend leukemic cells. *J. Med. Chem.* **45**, 3296–3309 (2002).
141. Mai, A. *et al.* Binding mode analysis of 3-(4-benzoyl-1-methyl-1H-2-pyrrolyl)-*N*-hydroxy-2-propenamide: a new synthetic histone deacetylase inhibitor inducing histone hyperacetylation, growth inhibition, and terminal cell differentiation. *J. Med. Chem.* **45**, 1778–1784 (2002).
142. Kao, R. Y. *et al.* A small-molecule inhibitor of the ribonucleolytic activity of human angiogenin that possesses antitumor activity. *Proc. Natl Acad. Sci. USA* **99**, 10066–10071 (2002).
143. Jenkins, J. L., Kao, R. Y. & Shapiro, R. Virtual screening to enrich hit lists from high-throughput screening: a case study on small-molecule inhibitors of angiogenin. *Proteins* **50**, 81–93 (2003).
144. Efferth, T. *et al.* Molecular modes of action of artesunate in tumor cell lines. *Mol. Pharmacol.* **64**, 382–394 (2003).

Acknowledgements

P. Acklin, H.-J. Roth, P. Schoeffter, A. Schuffenhauer and J. Zimmermann (Novartis) are acknowledged for various support and discussions. M.B. is supported by the Emmy Noether Programme of the Deutsche Forschungsgemeinschaft.

Competing interests statement

The authors declare that they have no competing financial interests.

Online links

DATABASES
The following terms in this article are linked online to:
NCBI: http://www.ncbi.nlm.nih.gov
AP1 | ATP citrate lyase | HSP90

FURTHER INFORMATION
Cerep Discovery™:
www.cerep.fr/Cerep/Utilisateur/CerepDiscovery/frs_Tech.asp
Developmental Therapeutics Program, National Cancer Institute: http://dtp.nci.nih.gov
Eidogen: www.eidogen.com
Iconix Pharmaceuticals DrugMatrix™:
www.iconixpharm.com/products/products_drugmatrix.html
Inpharmatica: www.inpharmatica.co.uk
Kensington InforSense: www.inforsense.com
Protein Data Bank: http://www.rcsb.org/pdb
Scitegic Pipeline pilot: www.scitegic.com
Access to this interactive links box is free online.

©2004 Nature Publishing Group