

Chemogenomics and biotechnology

Arthur Wuster and M. Madan Babu

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 0QH, UK

A robust knowledge of the interactions between small molecules and specific proteins aids the development of new biotechnological tools and the identification of new drug targets, and can lead to specific biological insights. Such knowledge can be obtained through chemogenomic screens. In these screens, each small molecule from a chemical library is applied to each cell type from a library of cells, and the resulting phenotypes are recorded. Chemogenomic screens have recently become very common and will continue to generate large amounts of data. The interpretation of this data will occupy biologists and chemists alike for some time to come. This review discusses methods for the acquisition and interpretation of chemogenomic data, in addition to possible applications of chemogenomics in biotechnology.

Introduction

The effects of small molecules on cells were central to the research of Paul Ehrlich (1854–1915). For much of his career, he strived to identify ‘magic bullets’, small molecules (see Glossary) that would enable him to target specific tissues or microbes while sparing others [1]. To find a cure for syphilis, he systematically screened a library of hundreds of potential drugs for their effect against *Treponema pallidum*, the causative agent of the disease. After testing 605 different compounds, he eventually identified arsphenamine, which he later marketed as Salvarsan 606 [2]. With this strategy, he initiated a whole new approach to drug discovery that has persisted until today [3]. The aim of Ehrlich’s chemical screen was simply to find a small molecule that was active against a single pathogen. The data structure created by his screen was one-dimensional, with the only dimension being a vector of the 606 compounds tested. Nowadays, chemical screens are more complex and wider in scope.

Here, we review two-dimensional chemogenomic screens (see Glossary). In these screens, the first dimension is a chemical library, just as it was in Ehrlich’s screen. The second dimension is a library of different cell types. These cell types can be well-defined mutants, such as in a library of yeast (*Saccharomyces cerevisiae*) deletion strains, where in each strain a different gene has been deleted; alternatively, the cell types can be defined in other ways, such as in a library of cancer cell lines or a library of meiotic recombinants [4,5]. The resulting data structure is a two-dimensional matrix in which each data point has two coordinates and one specific associated value (Figure 1a). The chemical coordinate specifies the small molecule that was applied, whereas the genetic coordinate specifies the cell type. The value of each data point is a measurement of the phenotype

of interest, such as viability, growth rate, or cell size and shape. Please refer to Box 1 for studies that measure different phenotypes.

The results of several chemogenomic screens with different designs have recently been published (Figure 2). They all have in common the data structure described above, but the experimental designs and aims, such as the identification of cellular targets of small molecules [6] or the characterisation of cellular pathways [7], vary between them. The methodologies described below have mainly been applied to *S. cerevisiae*. Nevertheless, the methodologies and their findings can also be applied to other systems [8] such as the human genome or pathogen and agricultural genomes.

Experimental setup

There are several different methodologies to carry out chemogenomic screens. The resulting data structure is similar among them all, but the interpretation of this data structure depends on the design of the experiment. For yeast, at least three different types of mutant libraries can be generated, such as heterozygous deletions, homozygous deletions, and overexpression libraries (Figure 2). In the following sections, we discuss how each of these library types can be used to generate chemogenomic data. For each mutant library, we assume that the measured phenotype refers to changes in growth rate or viability. The measured phenotypes can be caused by a variety of molecular interactions between proteins and small molecules (Figure 3), the interpretation of which depends on the choice of experimental setup.

Heterozygous deletions

This library type consists of diploid yeast cells. Almost every single one of the 6500 yeast genes can be deleted

Glossary

Chemical genetics: modulation of protein function using small molecules.

Chemical space: entirety of theoretically possible arrangements of atoms that result in small molecules. Chemical space can be explored with approaches such as diversity-oriented synthesis (DOS).

Chemogenomics: interactions between the genome (i.e. the sum of the genes in a cell) and small molecules.

DNA microarray: hybridisation-based technology that enables the semi-quantitative measurement of the abundance of individual nucleic acid chains within a cell.

Green biotechnology: branch of biotechnology concerned with the modification of crops and other plants.

log P value: the partition value P, often expressed as log P, is a measure of the ability of a small molecule to dissolve in lipids (lipophilicity).

Reverse genetics: approach to determine the function of a gene with known nucleotide sequence.

RNA interference (RNAi): inhibition of gene expression by double-stranded RNA fragments complementary to the target gene sequence using specific cellular pathways.

Small molecules: chemical molecules of small molecular weight compared to macromolecules such as DNA, RNA or proteins.

Corresponding authors: Wuster, A. (awuster@mrc-lmb.cam.ac.uk); Madan Babu, M. (madanm@mrc-lmb.cam.ac.uk).

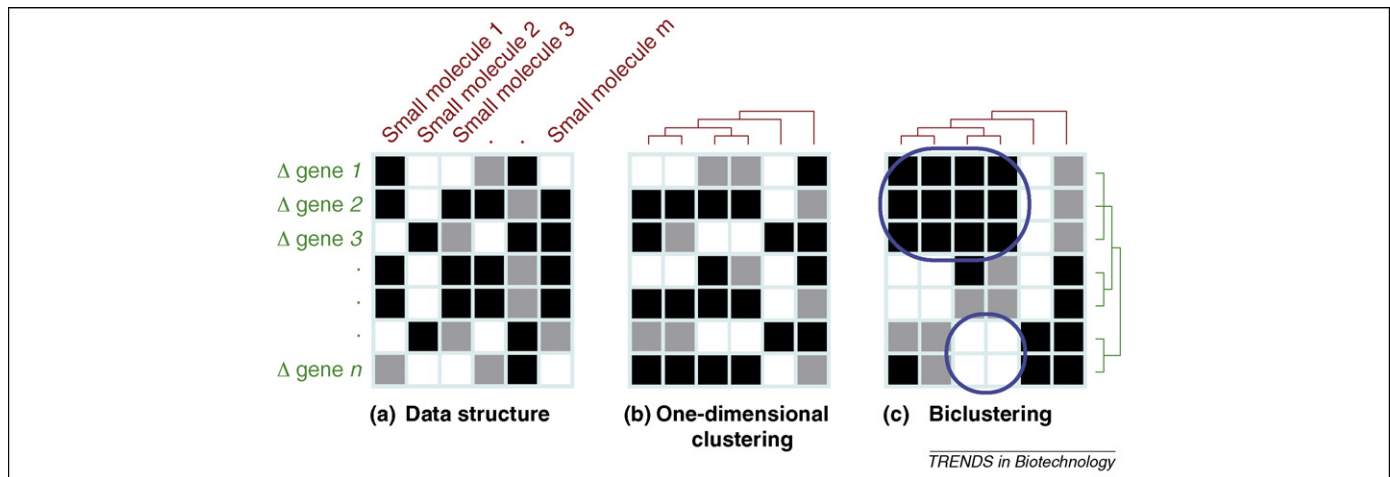


Figure 1. The chemogenomic data structure and clustering methods. (a) A hypothetical data structure obtained from a chemogenomic screen. One dimension (columns) refers to the small molecules added, and the other dimension (rows) refers to cell type, in this case defined deletion (Δ) mutants. The values, here represented as white, grey and black boxes, refer to the growth or viability of the specific mutant when treated with a specific small molecule. (b) The columns of the matrix (small molecules) have been clustered according to the similarity of their effect on different cell types. (c) Rows and columns of the matrix have been biclustered (see text). Two resulting biclusters are highlighted in blue.

independently, as long as the deletion is only on one of the two chromosomes of a chromosome pair (i.e. heterozygous) [9]. In such a library, even cells with a deletion in an essential gene will normally survive, because they still have another copy of the gene on the second chromosome. Data from a heterozygous deletion library and a homozygous

deletion library, which is described in the next section, need to be interpreted in different ways [10–12]. The rationale behind screening against a heterozygous deletion library is that a cell that has only one copy of a particular gene produces less of the gene product encoded by that gene. Such a cell is rendered more sensitive to a compound that is

Screen design →	Homozygous deletants	Heterozygous deletants	Overexpression
↓ Detection			
Non-competitive arrays Well plate	Ref. [32] 21 conditions 4710 mutants Aim: determine gene function	Ref. [11] 1 small molecule > 5000 mutants Aim: determine mode of action for drugs	Ref. [14] 1000 small molecules 7296 mutants Aim: identify protein targets for small molecules
Competitive barcoded pools 	-	Ref. [12] 82 small molecules 4111 mutants Aim: determine mode of action for drugs Ref. [10] 10 small molecules 6000 mutants Aim: determine functional interactions of small molecules	Ref. [6] 2 small molecules 3929 mutants Aim: identify protein targets for small molecules

Figure 2. A selection of recent chemogenomic screens using some of the different methodologies and cell libraries available in yeast. This table shows selected chemogenomic screens carried out in the recent past. The columns of the table correspond to three different types of mutant libraries (screen design) with differential gene expression: expression of a gene product can be avoided altogether by deleting the gene from both chromosomes (homozygous deletion; empty red arrows); alternatively, expression can be reduced by deleting the gene on only one chromosome (heterozygous deletion; filled and empty red arrows); finally, expression can be increased by introducing extra copies of the gene into the cell (overexpression; multiple red arrows). The rows of the table correspond to two different detection methods that can be used to observe the phenotype of the mutants when treated with small molecules. The small molecule can be added to a well plate in which each mutant occupies a separate well. The effect is then observed directly (non-competitive arrays). Alternatively, the bar-coded mutants can all be placed into a flask containing a small molecule. The effect on the growth of each individual mutant is then observed by measuring the abundance of the different mutants using microarrays with probes that are complementary to the bar codes of the mutants. The intensities of specific probes are proportional to the abundance of the mutant strains in the pool (competitive bar-coded pools).

Box 1. Phenotypes for large-scale screening

Besides viability, several parameters can be assessed with chemogenomic screens:

- Fitness: growth rate in laboratory conditions
- Fitness: growth rate in competition with other strains
- Cell morphology (e.g. shape, size, volume) [60]
- Gene expression profiles (e.g. transcript fingerprint) [61]
- Behaviour (e.g. clustering, movement) [62]
- Cell cycle abnormalities [63]
- Cytoblots [64]

capable of disrupting the function of this gene product than a wild type cell would be. This effect is also referred to as haploinsufficiency, and the method has also been described as haploinsufficiency profiling [10]. Data points obtained from such a screen that show decreased viability might therefore directly point to an interaction between a chemical compound and a gene product.

Homozygous deletions

This type of library consists of yeast cells in which each non-essential gene is deleted from both chromosomes of a diploid chromosome pair. For obvious reasons, in such a homozygous deletion library only non-essential genes can be deleted, because the deletion of any of the ~1000 essential yeast genes will by definition be lethal [9]. The

interpretation of such a chemogenomic screen is more complex for a homozygous deletion library than for a heterozygous deletion library. Clearly, a small molecule cannot act on a gene product after the respective gene has been knocked out. Therefore, if a homozygous deletion strain is sensitive to a particular small molecule, this can be interpreted in different ways. One possibility is that the deleted gene is different from the gene targeted by the small molecule but that it has a similar function. In the absence of small molecules, disruption of one of the two genes is expected to have little effect, because the second gene provides a back-up function. However, the disruption of the function of both genes – the one by the small molecule and the other by deletion – might result in decreased viability. This has been verified in a study that compared the effect of knocking out one gene and targeting the other with a small molecule with the effect of knocking out both genes, a method that is also referred to as a synthetic genetic array (SGA) [7]. This study considered the ergosterol biosynthesis gene *ERG11*, a target of the small molecule fluconazole, as an example. There are 27 genes that cause synthetic lethality when mutated together with *ERG11*. Thirteen of these 27 mutants also cause lethality when *ERG11* is not mutated but when fluconazole is added. Comparing the lethality profile of other small molecules to the synthetic lethal profiles of their targets, it was found that SGA profiles generally tend to have a significant overlap with the chemogenomic profiles [7]. This shows that the effect of deleting *ERG11* is similar to the effect of disrupting its function with fluconazole.

In an approach that is complementary to the homozygous deletion library approach described above, it would be possible to screen for deletions that confer resistance to a compound whose addition would be deleterious to the wild type strain. An example of this would be if a compound that is not normally toxic to the cell becomes toxic as a result of the action of a cellular enzyme modifying it. In this case, deletion of the genes encoding the modifying enzyme would confer resistance to the cell. For example, azidothymine becomes active against DNA polymerases and reverse transcriptases only after it has undergone several cycles of phosphorylation [13]. Mutating the genes responsible for this phosphorylation could potentially decrease the sensitivity of the cell towards azidothymine. In this way, genes that are responsible for causing side effects to certain drugs could be identified.

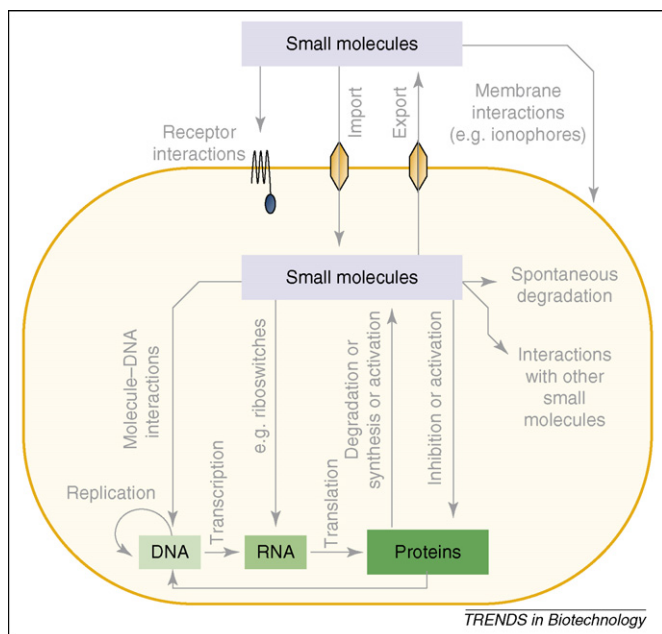


Figure 3. Possible interactions between small molecules and the components of a cell. This figure shows ways in which small molecules can affect the molecular phenotype of a cell by interacting with proteins and other biomolecules. Small molecules can enter the cell via an import channel or by diffusion through the membrane, or they can interact with a membrane receptor and trigger a cellular response. Once inside the cell, they can be degraded, modified, exported out of the cell again, or they can interact (inhibition or activation) with cytoplasmic proteins and other cellular components. Deletion or mutation of the proteins involved in the above-mentioned processes can therefore affect the molecular phenotype in the presence of the small molecules. For example, if a protein that degrades a specific small molecule were deleted, the small molecule would be present for a prolonged period of time and would therefore have a stronger phenotype. The small molecules can also interact with the informational macromolecules RNA (e.g. riboswitches) and DNA by interfering with DNA replication, transcription or translation. Thus, they can have an effect on the central dogma, which essentially states that information flows from DNA to RNA to protein [21,65]. This therefore enables studies that probe the function of specific information-carriers in the cell.

Overexpression

A third approach investigates the effect of small molecules under conditions of increased gene dosage that can be amended by increasing gene copy number [6,8]. One such study [14] used an overexpression library to screen for genes that confer increased resistance to small molecules when multiple copies of a gene, and therefore presumably higher levels of the gene product, are present. In this way it was possible to show, among other things, that the breast cancer therapeutic tamoxifen disrupts calcium homeostasis [14].

Beyond deletion and overexpression libraries

Alternative approaches for generating libraries of different cell types include decreased abundance by mRNA

perturbation (DAmP). DAmP enables fine-tuning of protein concentration by disrupting the 3' untranslated region of a gene by insertion of an antibiotic resistance marker, which greatly destabilizes the corresponding mRNA [15]. DAmP leads to proteins that are synthesized at substantially lower levels than in the wild type, but under their natural transcriptional regulation. DAmP can therefore be used to gain insights into the deleterious and alleviating effects caused by small molecules at differential protein levels [16].

Chemical libraries

In addition to various types of libraries of cell types, different small molecule libraries can also be used for chemogenomic screens [17–20]. There are two fundamentally different approaches to the design of small molecule libraries. One approach uses small molecule libraries that show as much chemical diversity as possible. The other approach draws its small molecule library from only a small fraction of defined chemical space (see Glossary) [21,22]. The range of small molecules in such a chemical library can be limited by choosing those compounds that are likely to have some biological activity. By analysing the chemical descriptors of approved drugs, Lipinski's 'rule of five' was developed [23]. This states that most small molecules that are also approved drugs have less than five hydrogen-bond donors, less than ten hydrogen-bond acceptors, and a log P value (see Glossary), a measure of lipophilicity, of less than five. Moreover, small molecules with potential biological activity need to be able to enter the cell, and this limits their size to ~ 1500 Da, because larger molecules are unlikely to cross the cell membrane unaided [24]. In addition, there is evidence that naturally occurring small molecules are more likely to be able to modulate protein function [25,26]. The development of carefully designed chemical libraries remains a major focus of research [18].

Detection methods

Chemogenomic screens can be carried out in two fundamentally different designs: non-competitive arrays and competitive mutant pools (Figure 2).

Non-competitive arrays

This design consists of a set of arrays that contain each of the deletion or overexpression strains separate from the others so that they can be observed individually. After each compound from the chemical library is added to a different set of arrays, changes in growth can be measured independently for each strain and each small molecule. Given that each mutant strain is physically isolated, strains do not interact or compete for resources in this design.

Competitive mutant pools

In this design, each mutant has a unique DNA sequence also referred to as a bar code. All bar-coded mutant strains are initially grown together in the same flask. This mutant pool is then divided into multiple flasks and each compound from the chemical library is added to a different flask. In each flask, the different strains compete for resources, and fitter strains increase in concentration in

the presence of the small molecule. For analysis, the genetic material of the contents of the flasks is hybridised to microarrays (see Glossary), which contain probes that are complementary to the bar codes of the individual deletion strains. The observed differential intensities of the probes correspond to the differential concentration and therefore the differential viability of the respective strains in the presence of the small molecule. However, in the design of this screen, the intrinsically different growth rates of mutants compared with wild type cells need to be taken into account when interpreting the obtained data. It is therefore important that the data are normalised by the growth rate of each mutant without any small molecule treatment.

It has been observed that the results of the array-based design and the pool-based design overlap but differ [27]. A possible explanation for this is interactions between strains in mutant pools. For example, certain deletion strains could accumulate intermediates in a metabolic pathway from which an enzyme has been deleted, and these intermediates, when secreted, might interfere with the viability of other strains in the same flask. Alternatively, yeast strains grown in flasks might behave differently from yeast strains grown in arrays, owing to the pressures of inter-strain competition for resources, which is not a limitation in the array-based detection method.

Data interpretation and analysis

To Understand and interpret chemogenomic data is not a trivial task, and the observed phenotypic effects can have many causes and can also be indirect (Figure 3). Here, we discuss the most common methods for interpreting chemogenomic data.

Clustering

The application of clustering algorithms is a standard way of analysing multidimensional data. In chemogenomic screens, data clustering can be applied to both the chemical and the genetic dimension. A cluster in the chemical dimension will result in groups of small molecules that have similar effects on the different cell types (Figure 1b). It has been shown that clusters in the chemical dimension often contain small molecules that are chemically similar [10]. Therefore, the action of untested small molecules could be predicted based on similarity to tested small molecules from previous screens. Accordingly, we could expect that a cell type cluster would be enriched for mutated genes (deletion or overexpression) that have a certain function in common.

To identify such clusters, several supervised or unsupervised algorithms that are also applied to microarray data have been described. They include hierarchical clustering, *k*-means, and self-organising maps [28,29]. Alternatively, biclustering, an approach that analyses both dimensions at the same time, can be employed [30,31]. The biclusters obtained in this way contain a subset of mutated genes that interact similarly with a subset of small molecules, and vice versa. The aim of biclustering is to sort both dimensions simultaneously in such a way that functionally significant blocks can be identified among the overall pattern (Figure 1c). Another important

advantage of biclustering is that it can identify mutated genes or small molecules that are associated with more than one cluster. This is a significant improvement over simpler clustering procedures because many genes might have more than one function. Biclustering approaches have successfully been applied to chemogenomic data [12,32]. Probabilistic sparse matrix factorisation (PSMF) is an alternative algorithm to biclustering, and in some cases it is more sensitive than hierarchical clustering. For example, the small molecules verrucarin and neomycin sulphate are clustered together by PSMF but not by hierarchical clustering [12]. For a detailed review of the various clustering methods and how they are implemented, please see references [33,34].

Matrix operations

Although clustering is a helpful tool for the interpretation of chemogenomic data, it is not the only method available. Weinstein and coworkers [35] have taken an interesting approach by integrating two different two-dimensional matrices obtained from a chemogenomic screen, and this strategy could also prove valuable for the interpretation of other datasets. Their first matrix had small molecules as its first dimension and cell lines as its second dimension, and the associated values referred to differential growth. In a second matrix, genes were represented in the first dimension and cell lines in the second dimension, and the associated values referred to transcript levels as measured by microarrays. The common dimension in both matrices was the cell line. The authors then correlated the expression of each gene in each cell line with its response to each small molecule, and thereby obtained a third two-dimensional matrix, which had small molecules as one dimension and genes as the other dimension. From this matrix, the authors identified the genes that are expressed at a differential level in cell lines that have differential sensitivity to a specific small molecule. In this way, they were able to show how the transcript levels of particular genes relate to sensitivity to specific small molecules.

Matrix data can in many cases also be visualized and interpreted in networks. Small molecule–gene networks have also been used to visualize and analyse chemogenomic-like datasets [36–39]. One conclusion that has been possible from the analysis of such drug-target networks is that there has been a trend for more rational drug design over the past decade [37,40].

Applications

The immediate purpose of a chemogenomic screen is to characterize the effect that a set of small molecules has at the gene or protein level. From a biotechnological point of view, such chemogenomic data can allow for the identification of proteins as novel drug targets [41]. In a screen of small molecules against a library of heterozygous yeast deletions, the gene products that interact with each of the small molecules can be identified. In a chemogenomic screen of mammalian cell lines, the transcription factor STAT3 was identified as a potential target for apratoxin A, a small molecule that is cytotoxic against tumour cells [42]. Many yeast genes have orthologues in humans [43]. This raises the possibility that interactions with small

molecules are conserved and that orthologues could be potential drug targets or might help us gain an understanding of the side effects of drugs at the molecular level [8]. Rapamycin is an example of a small molecule with a molecular mechanism of action that is conserved between yeast and humans [6]. This is of particular interest because the majority of medicines are small molecules [3], and many of them, including rapamycin, are used as drugs. Furthermore, any compound that decreases the viability of *S. cerevisiae* might also be active against related pathogenic fungi and is therefore a candidate for the development of anti-fungal agents.

Small molecules that are found to be active against specific proteins could also serve as alternatives to genetic methods. In reverse genetics (see Glossary), to disrupt the function of a protein, the DNA sequence that encodes the protein is mutated. Although this approach has generated a large variety of useful data, it has some drawbacks. First of all, mutagenesis can be difficult to implement in higher animals such as mammals. Furthermore, mutagenesis strategies do not exhibit a large degree of spatial and temporal flexibility, because it is difficult to mutate a gene for only a transient period of time or in a limited and clearly defined set of cells. An additional problem is that in multifunctional proteins it can be difficult to disrupt only one protein function while leaving others intact. Chemical genetics (see Glossary) has the potential to overcome these problems [25,44–46]. The ultimate goal of chemical genetics is to identify small molecules that are able to modulate the function of as many proteins in a cell as possible, either by simulating a gain-of-function or a loss-of-function mutation [47]. It has been estimated that between 8 and 14 percent of proteins in eukaryotic genomes are ‘druggable’ in this way [48]. Chemical genetics can be more easily employed in mammalian cells than mutagenesis strategies and in a manner that is specific to a set of clearly defined cells, or during a clearly defined time window by adding and removing the compounds used in the screen. Furthermore, in a multifunctional protein, potentially only one specific function can be modulated without affecting the others.

The potential of chemical genetics in drug development is well recognized. Recent applications include the targeting of cancer cell lines that carry mutations in proteins involved in cancer pathways with specific small molecules [49,50]. However, other areas of biotechnology, including ‘green biotechnology’ (see Glossary), might also benefit from the application of chemical genetics. Crop plants could be improved with the use of small molecules that target specific proteins; for example, those involved in drought resistance pathways [51]. Given that such small molecules can be applied in a manner similar to pesticides, we could avoid the introduction of genetically modified crops, which is often associated with environmental, legal or technical difficulties.

Besides these practical applications, chemogenomic data can also offer fundamental biological insights. Grouping genes according to their chemogenomic profiles can lead to the identification of clusters that are enriched for functional categories as defined by the Gene Ontology (GO) database [28]. If a gene of previously unknown function is

found in such a cluster, it is likely that it has a function similar to the other genes in the same cluster. Chemogenomics might therefore enable us to determine the previously unknown function of certain genes [32]. Moreover, it might even be possible to define novel functional classes of genes. More specifically, if a small molecule causes decreased viability when a set of genes is deleted, this set of genes can be interpreted as being necessary for conferring resistance to the small molecule. Therefore, these genes might have a previously unknown function in common. For example, deletion of members of the protein complexes SAGA, Swi-Snf and Ino80 caused sensitivity to cadmium, cycloheximide, hydroxyurea and glycerol [32]. Therefore, it has been assumed that these complexes share a function, which is to regulate genes that are required in the presence of these small molecules.

Discussion

It should be noted that small molecules interact not only with proteins but also with DNA, RNA and other cellular biomolecules such as membrane lipids and polysaccharides (Figure 3). These interactions are not necessarily detected in the types of genetic screens discussed above. Different methods can be used to detect such interactions; however, these methods are often less specific. Microarrays which measure the effect of small molecules of a chemical library on RNA levels, can be used to generate two-dimensional chemogenomic data. For this purpose, the small molecules in the chemical library are added to a cell, and the differential RNA levels are compared against those in the same cell without the small molecule. While mutant libraries often result in rather sparse matrices with only a few mutants showing differential growth under the influence of a small molecule, the number of transcripts which show differential concentrations as measured by microarrays tends to be much higher. This can be explained by the fact that in most cases the small molecule only indirectly causes changes in transcript levels by interacting with transcription factors, and that the differentially expressed genes will not be specific. For example, cellular responses induced by different stresses have been shown to share a common set of differentially expressed genes [52].

Conclusions

A major challenge for the future of chemogenomics will be to extract meaningful information from the data various screens produce. Although the amount of chemogenomic data continues to rise, the application of novel methods to specific biotechnological and biomedical applications (translational research) have not been fully exploited. In the future, new analysis tools will need to be required to deal with the large amount of data generated. Efforts have already been made to organize data that is relevant for chemogenomics in public databases, such as the Broad Institute's ChemBank database (<http://chembank.broad.harvard.edu/>), the National Center for Biotechnology Information's PubChem database (<http://pubchem.ncbi.nlm.nih.gov/>), or the European Bioinformatics Institute's ChEBI database (<http://www.ebi.ac.uk/chebi/>). Furthermore, the development of standardized representations for small molecules like InChi (<http://www.iupac.org/inchi/>) and SMILES (<http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>) facilitates data integration and comparison in conjunction with publicly available chemical structure lookup services such as CCLS (<http://cactus.nci.nih.gov/cgi-bin/lookup/search>).

A large number of chemogenomic studies to date have been carried out on the yeast *S. cerevisiae*. The primary reason for this is the availability of genetic libraries of strains in which almost every yeast gene is deleted. Although yeast has been the focus of most studies, there is no reason to limit the chemogenomic approaches described here to this organism. For example, a knockout library of the Gram-negative bacterium *Escherichia coli* has recently become available [53]. Furthermore, a library of RNA interference (RNAi, see Glossary) knockouts is now available for a variety of cells [54–56] and has been used for chemogenomic screens [57]. Moreover, cell types do not always have to be as clearly defined as they are in mutant libraries. For example, in a screen measuring the viability of 60 cancer cell lines in response to various drugs, the precise mutation that defined each cancer cell line was not known [35]. Taken together, this demonstrates that another major focus of chemogenomics will be to translate the methodologies and findings from yeast to other model systems.

Although the approaches discussed here do not measure the interactions between small molecules and proteins directly, they are able to determine small molecule targets indirectly. By contrast, affinity-based methods most often immobilise the small molecules and aim to identify the proteins that directly bind them [58,59]. Affinity-based methods directly determine the interactions between small molecules and proteins and have the advantage that they provide more easily interpretable data. Unlike chemogenomic approaches, they do not provide any information about changes in cellular phenotype associated with the small molecule–protein interaction. Affinity-based methods and chemogenomic methods can therefore be considered to have complementary aims.

Although the approaches discussed here do not measure the interactions between small molecules and proteins directly, they are able to determine small molecule targets indirectly. By contrast, affinity-based methods most often immobilise the small molecules and aim to identify the proteins that directly bind them [58,59]. Affinity-based methods directly determine the interactions between small molecules and proteins and have the advantage that they provide more easily interpretable data. Unlike chemogenomic approaches, they do not provide any information about changes in cellular phenotype associated with the small molecule–protein interaction. Affinity-based methods and chemogenomic methods can therefore be considered to have complementary aims.

Acknowledgements

We thank Varodrom Charoensawan, Winsome Cheung and Michael Bremang for critically reading the manuscript. We also thank the anonymous referees for helpful feedback. M.M.B. acknowledges Darwin College and Schlumberger for generous support.

References

- 1 Ehrlich, P. (1911) Aus Theorie und Praxis der Chemotherapie. *Folia Serologica* VII, 697–714
- 2 Gensini, G.F. *et al.* (2007) The contributions of Paul Ehrlich to infectious disease. *J. Infect.* 54, 221–224
- 3 Lipinski, C. and Hopkins, A. (2004) Navigating chemical space for biology and medicine. *Nature* 432, 855–861
- 4 Perlstein, E.O. *et al.* (2007) Quantifying fitness distributions and phenotypic relationships in recombinant yeast populations. *Proc. Natl. Acad. Sci. U. S. A.* 104, 10553–10558
- 5 Perlstein, E.O. *et al.* (2007) Genetic basis of individual differences in the response to small-molecule drugs in yeast. *Nat. Genet.* 39, 496–502
- 6 Butcher, R.A. *et al.* (2006) Microarray-based method for monitoring yeast overexpression strains reveals small-molecule targets in TOR pathway. *Nat. Chem. Biol.* 2, 103–109
- 7 Parsons, A.B. *et al.* (2004) Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat. Biotechnol.* 22, 62–69

- 8 Luesch, H. (2006) Towards high-throughput characterization of small molecule mechanisms of action. *Mol. Biosyst.* 2, 609–620
- 9 Deutschbauer, A.M. *et al.* (2005) Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* 169, 1915–1925
- 10 Giaever, G. *et al.* (2004) Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 101, 793–798
- 11 Baetz, K. *et al.* (2004) Yeast genome-wide drug-induced haploinsufficiency screen to determine drug mode of action. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4525–4530
- 12 Parsons, A.B. *et al.* (2006) Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell* 126, 611–625
- 13 Brenner, C. (2004) Chemical genomics in yeast. *Genome Biol.* 5, 240
- 14 Luesch, H. *et al.* (2005) A genome-wide overexpression screen in yeast for small-molecule target identification. *Chem. Biol.* 12, 55–63
- 15 Muhlrud, D. and Parker, R. (1999) Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. *RNA* 5, 1299–1307
- 16 Schuldiner, M. *et al.* (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 123, 507–519
- 17 Mason, J.S. *et al.* (1999) New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem.* 42, 3251–3264
- 18 Drewry, D.H. and Young, S.S. (1999) Approaches to the design of combinatorial libraries. *Chemometrics Int. Lab. Sys.* 48, 1–20
- 19 Stahura, F.L. *et al.* (2002) Methods for compound selection focused on hits and application in drug discovery. *J. Mol. Graph. Model.* 20, 439–446
- 20 Savchuk, N.P. *et al.* (2004) Exploring the chemogenomic knowledge space with annotated chemical libraries. *Curr. Opin. Chem. Biol.* 8, 412–417
- 21 Schreiber, S.L. (2005) Small molecules: the missing link in the central dogma. *Nat. Chem. Biol.* 1, 64–66
- 22 Dobson, C.M. (2004) Chemical space and biology. *Nature* 432, 824–828
- 23 Lipinski, C.A. *et al.* (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26
- 24 Lien, E.J. and Wang, P.H. (1980) Lipophilicity, molecular weight, and drug action: reexamination of parabolic and bilinear models. *J. Pharm. Sci.* 69, 648–650
- 25 Stockwell, B.R. (2000) Frontiers in chemical genetics. *Trends Biotechnol.* 18, 449–455
- 26 Gassner, N.C. *et al.* (2007) Accelerating the discovery of biologically active small molecules using a high-throughput yeast halo assay. *J. Nat. Prod.* 70, 383–390
- 27 Giaever, G. *et al.* (1999) Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat. Genet.* 21, 278–283
- 28 Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29
- 29 Madan Babu, M. (2004) Introduction to microarray data analysis. In *Computational Genomics: Theory and Application* (Grant, R.P., ed.), pp. 225–249, UK, Horizon Press
- 30 Cheng, Y. and Church, G.M. (2000) Biclustering of expression data. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 8, 93–103
- 31 Madeira, S.C. and Oliveira, A.L. (2004) Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 1, 24–45
- 32 Dudley, A.M. *et al.* (2005) A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol. Syst. Biol.* 1, 2005.0001
- 33 Xu, R. and Wunsch, D., 2nd (2005) Survey of clustering algorithms. *IEEE Trans. Neural Netw.* 16, 645–678
- 34 D'Haeseleer, P. (2005) How does gene expression clustering work? *Nat. Biotechnol.* 23, 1499–1501
- 35 Scherf, U. *et al.* (2000) A gene expression database for the molecular pharmacology of cancer. *Nat. Genet.* 24, 236–244
- 36 Sharom, J.R. *et al.* (2004) From large networks to small molecules. *Curr. Opin. Chem. Biol.* 8, 81–90
- 37 Yildirim, M.A. *et al.* (2007) Drug-target network. *Nat. Biotechnol.* 25, 1119–1126
- 38 Haggarty, S.J. *et al.* (2003) Multidimensional chemical genetic analysis of diversity-oriented synthesis-derived deacetylase inhibitors using cell-based assays. *Chem. Biol.* 10, 383–396
- 39 Ma'ayan, A. *et al.* (2007) Network analysis of FDA approved drugs and their targets. *Mt. Sinai J. Med.* 74, 27–32
- 40 Wuster, A. (2008) Networking with Drugs. *Mol. Biosyst.* 4, 14–17
- 41 Bredel, M. and Jacoby, E. (2004) Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat. Rev. Genet.* 5, 262–275
- 42 Luesch, H. *et al.* (2006) A functional genomics approach to the mode of action of apratoxin A. *Nat. Chem. Biol.* 2, 158–167
- 43 Hohmann, S. (2005) The Yeast Systems Biology Network: mating communities. *Curr. Opin. Biotechnol.* 16, 356–360
- 44 Spring, D.R. (2005) Chemical genetics to chemical genomics: small molecules offer big insights. *Chem. Soc. Rev.* 34, 472–482
- 45 Stockwell, B.R. (2004) Exploring biology with small organic molecules. *Nature* 432, 846–854
- 46 Florian, S. *et al.* (2007) Chemical genetics: reshaping biology through chemistry. *HSFP Journal* 1, 104–114
- 47 Balch, W.E. *et al.* (2008) Adapting proteostasis for disease intervention. *Science* 315, 916–919
- 48 Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730
- 49 Dolma, S. *et al.* (2003) Identification of genotype-selective antitumor agents using synthetic lethal chemical screening in engineered human tumor cells. *Cancer Cell* 3, 285–296
- 50 Rognan, D. (2007) Chemogenomic approaches to rational drug design. *Br. J. Pharmacol.* 152, 38–52
- 51 Raikhel, N. and Pirrung, M. (2005) Adding precision tools to the plant biologists' toolbox with chemical genomics. *Plant Physiol.* 138, 563–564
- 52 Gasch, A.P. *et al.* (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* 11, 4241–4257
- 53 Baba, T. *et al.* (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 2006.0008
- 54 Boutros, M. *et al.* (2006) Analysis of cell-based RNAi screens. *Genome Biol.* 7, R66
- 55 Paddison, P.J. *et al.* (2004) A resource for large-scale RNA-interference-based screens in mammals. *Nature* 428, 427–431
- 56 Horn, T. *et al.* (2007) GenomeRNAi: a database for cell-based RNAi phenotypes. *Nucleic Acids Res.* 35, D492–D497
- 57 Brummelkamp, T.R. *et al.* (2006) An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors. *Nat. Chem. Biol.* 2, 202–206
- 58 Bantscheff, M. *et al.* (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat. Biotechnol.* 25, 1035–1044
- 59 Keiser, M.J. *et al.* (2007) Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* 25, 197–206
- 60 Bakal, C. *et al.* (2007) Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science* 316, 1753–1756
- 61 Butcher, R.A. and Schreiber, S.L. (2005) Using genome-wide transcriptional profiling to elucidate small-molecule mechanism. *Curr. Opin. Chem. Biol.* 9, 25–30
- 62 de Bono, M. and Bargmann, C.I. (1998) Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* 94, 679–689
- 63 Burney, R.O. *et al.* (2007) A transgenic mouse model for high content, cell cycle phenotype screening in live primary cells. *Cell Cycle* 6, 2276–2283
- 64 Lokey, R.S. (2003) Forward chemical genetics: progress and obstacles on the path to a new pharmacopoeia. *Curr. Opin. Chem. Biol.* 7, 91–96
- 65 Crick, F. (1970) Central dogma of molecular biology. *Nature* 227, 561–563