Drug Discovery Today • Volume 00, Number 00 • December 2013 PERSPECTIVE

feature

Connections in pharmacology: innovation serving translational medicine

Trinidad Montero-Melendez, t.monteromelendez@qmul.ac.uk and **Mauro Perretti**

There is a paucity of molecules that progress through the drug development pipeline, making the drug discovery process expensive and frustrating. Innovative approaches to drug development are therefore required to maximise opportunities. Strategies such as the Connectivity Map (CMap), which compares >7000 gene expression signatures generated from more than 1000 drugs, can produce associations between currently unrelated therapeutics, unveiling new mechanisms of action and favouring drug repositioning. Here, we discuss these opportunities that could aid the drug development process and propose rigorous publication of 'omics data with open access and data sharing. We, pharmacologists of the third millennium, must aim towards maximising knowledge in an unbiased and cost-effective manner, to deliver new drugs for the global benefit of patients.

Introduction

As learnt from Darwin's *Origin of Species*, it is not the strongest nor the most intelligent of the species that survives but the one that is the most adaptable to change. We could extrapolate this statement to the current situation of the pharmaceutical industry, which seems unable to sustain its own growth, owing to the worldwide challenging economical climate and current research strategies, perhaps too much seduced by technology and forgetting the unpredictable nature of research discoveries [1,2]. There is an unquestionable need for change and a reinvention of the drug development process to guarantee, in a cost-effective manner, the transition from basic research to patient benefit [3].

We now know that patients are not all the same, even if they receive the same diagnosis [4]. They can belong to a particular disease subtype that might require a specific therapy. The so-called 'omics (a suffix etymologically derived from Greek, meaning the totality of something) represent one of the best strategies to reveal differences between patients, because the study of the totality of the genome, transcriptome, proteome, lipidome or metabolome does not require previous knowledge on the nature of these differences. Genomics, however, can contribute not only to patient stratification [5] but can also impact the entire drug development process [6], including target identification, deciphering the mechanism of action of drugs, implementation of individualised medicines to seek optimal benefit for each patient and monitoring drug response and toxicity. In this article, we will discuss innovative whole-genome-based strategies that contribute to drug discovery and development by: (i) identification of novel treatments for a specific disease; (ii) discovery of mechanisms of action of novel or known compounds; and, finally, (iii) drug repositioning studies. We will also highlight the need for more standardised methods and data-sharing policies to ensure full exploitation of these findings into genuine clinical benefit.

Emerging strategies for drug discovery and drug repositioning

The pharmaceutical industry needs to adapt according to the current economical situation. A reinvention of the innovation process is necessary, because technological innovation has not been proportionally translated into scientific innovation. Therefore, besides new instruments, new concepts are needed to improve the efficiency of drug discovery [1,2]. One of the main consequences of a genome-wide study is the

1359-6446/06/$ - see front matter http://dx.doi.org/10.1016/j.drudis.2013.11.022

massive amount of information that is generated. Whereas analyses of multiple hits can be more sophisticated than simple listing (up- and down-regulated genes), current approaches tend to follow a more integrated interpretation from a systems-oriented perspective [7–9].

A novel and powerful opportunity derives from the Connectivity Map (CMap) [10–12]. CMap is open-source software that allows a new interpretation of microarray data by comparing gene expression profiles of interest with those obtained for hundreds of bioactive small molecules, most of which are FDA-approved drugs. The most recent version (build 02; http://www.broadinstitute.org/cmap/) of this database contains 7056 gene expression profiles from 1309 bioactive compounds in five different human cell lines. The signatures contained in the database can be compared with any gene-expression profile of interest following two approaches: a disease-centred approach, when we use the gene expression profile of a disease; and a drug-centred approach, when we use the gene expression profile of another drug of interest. As a result, the 1309 CMap drugs will be ranked according to the similarity with the gene signature of interest. Therefore, drugs with a negative score (i.e. they present opposite profiles to the signature of interest) might have the potential as new treatments for specific diseases, whereas drugs with a positive score (i.e. they have similar gene expression profiles) could be useful for identification of novel actions of existing drugs or to unravel drug mechanisms of action [10] (Fig. 1). Active efforts are currently being made to increase the capabilities of the CMap. The new forthcoming version (http://lincscloud.org/) will represent a dramatic expansion of the database and will contain almost one million gene expression profiles. In addition to the expansion in the number of pharmacological perturbagens (over 5000 compounds), one of the major novelties of the new CMap will be incorporation of genetic perturbations, that is gene expression profiles obtained by upregulation or downregulation using shRNA of specific genes, including drug targets and candidate disease genes. Thus, the query of the CMap could be used for drug repositioning; that is, giving novel indications for an existing drug [13,14]. For example, the anticonvulsant drug topiramate was linked (with a negative score) with the gene expression signature of inflammatory bowel disorder (IBD) [15]. This prediction was experimentally assessed using the trinitrobenzenesulfonic (TNBS)-acid-induced colitis model, in which the administration of topiramate significantly reduced intestinal inflammation. Using a similar approach, the histone deacetylase inhibitor vorinostat was predicted as a candidate therapeutic drug for gastric cancer, soliciting a series of *in vitro* investigations to explore this functional association [16]. It is worth noting that the CMap was proposed as a 'hypothesis-generating tool', which means that confirmation studies are an absolute requirement to validate initial predictions. Hassane *et al.* queried the CMap with the gene expression signature produced by the drug parthenolide on acute myelogenous leukemia (AML) cells. This drug was previously shown to ablate these cancer cells, and the predictions made with the CMap led to the identification of novel agents (celastrol and 4-hydroxy-2-nonenal) that could also markedly affect AML cells [17]. A CMap analysis also enabled Zhong *et al.* to propose a combination with angiotensin-converting enzyme inhibitors and histone deacetylase inhibitors as a reno-protective therapy [18].

Interrogation of the CMap can also serve for the identification of novel mechanisms of action of drugs. Hypoxia-inducible factor (HIF)2a inhibitors were found by the CMap to be associated (positive score) with the anti-inflammatory prostaglandin PGJ2 [19]. This finding incited subsequent experiments that showed how PGJ2 was acting as an endogenous regulator of HIF2a translation, suggesting this action as part of the anti-inflammatory effects of the prostaglandin. The CMap approach has also facilitated identification of novel classes of drugs including heat-shock protein (HSP)90 inhibitors [20], and dissection of the mechanism of action of a traditional Chinese medicinal herbal formula [21].

We have recently queried the CMap using the gene expression signature produced by the endogenous pro-resolving mediator Annexin A1 (AnxA1) [22]; although this analysis produced predictable associations (e.g. with nonsteroidal anti-inflammatory drugs and glucocorticoids), unexpected associations also emerged. In particular, the positive association with histone deacetylase inhibitors (HDACIs) led us to investigate whether a functional and mechanistic link between AnxA1 and HDACIs could exist. Further experimentation made us conclude that AnxA1 contributes to the anti-inflammatory mechanism of action of HDACIs [23].

Although innovative and promising, the CMap strategy is however, not devoid of limitations, but the new version discussed above might resolve some of them. Firstly, pharmacologically relevant effects do not necessarily need to be reflected at the transcriptional level. Secondly, the database was generated with a limited number of compounds and cell lines. For example, the under-representation of certain drug classes, such as kinase inhibitors, in the current version (build 02) might bias the results. Thirdly, gene expression signatures of interest are often not measured in the same cells and/or tissues as those used in the CMap. In addition, different treatment durations can lead to different results as a result of feedback regulation of the target, for example when studying G-protein-coupled receptors. Other nonbiological phenomena such as the 'batch effect', which affects the microarrays, compounds and cell type used, can also impact the accuracy of the predictions [24]. Finally, as mentioned before, the CMap has to be considered a hypothesis-generating tool where results need to be validated by further experimentation. In any case, its potential could be significant and, indeed, similar approaches for connecting drugs and genes are starting to emerge. For example, the tool mode of action by network analysis (MANTRA) allows analysis of the CMap data with an innovative approach that takes into consideration the variability in the transcriptional responses to the drug caused by cell line specific effects, different concentrations of drug applied and distinct experimental conditions [25]. Another example is drug versus disease (DvD), a new tool that combines the data from the CMap and the public microarray repositories Gene Expression Omnibus and Array Express [26]. In addition to new analytical tools, new powerful technologies such as next-generation sequencing (NGS), currently generating data faster than they can be analysed, might be incorporated and applied to drug discovery and development [27].

Successful translational research: importance of data sharing and replication

Despite the large number of studies using these powerful high-throughput 'omics analyses conducted over the past decade, the low number of discoveries that have been translated into practice is striking and concerning. To improve these odds, it is absolutely fundamental that research discoveries are reproduced and validated in independent studies. A recent analysis of 18 microarray studies showed that only two were fully reproduced by independent researchers [28]; the main reason for failure was the unavailability of the data necessary to reproduce the published results. Similarly, analysis of the top 50 journals with the highest impact factors revealed that only 70% require a mandatory public deposition of microarray data to guarantee publication. More surprisingly, even