Connections in Pharmacology: innovation serving translational medicine

Trinidad Montero-Melendez and Mauro Perretti

The William Harvey Research Institute, Barts and The London School of Medicine, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, United Kingdom

Corresponding author: Montero-Melendez, T. (t.monteromelendez@qmul.ac.uk)

Phone no: +44-207-882 5654; Fax no: +44-207-8826076

Keywords: Connectivity Map, drug repositioning, microarrays, pharmacology, gene expression.

Teaser: Innovative approaches like the CMap offer new opportunities for drug repositioning and discovery of new treatments and mechanisms of action, aiding the drug development process in a cost-effective manner.

Abstract

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 like the Connectivity Map (CMap), which compares >7,000 gene expression signatures generated from more than 1,000 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.

Main text

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, due 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 re-invention 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 may belong to a particular disease subtype that might require a specific therapy. The so-called 'omics' (a suffix etymologically derived from the Greek, meaning the totality of something) represent one of the best strategies to reveal differences between patients, as 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 drugs mechanisms of action, implementation of individualized medicines to seek optimal benefit for each patient and to monitor 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) for drug repositioning studies. We will also highlight the need for more standardized 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 re-invention of the innovation process is necessary, as 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 any genome-wide study is the massive amount of information that is generated. Whilst 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 an 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 7,056 gene expression profiles from 1,309 bioactive compounds in 5 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-centered* approach, when we use the gene expression profile of a disease, and a *drug-centered* approach, when we use the gene expression profile of another drug of interest. As a result, the 1,309 CMap drugs will be ranked according to the similarity with the gene-signature of interest. Therefore, drugs with negative score (i.e. they present opposite profiles to the signature of interest) might have the potential as new treatments for specific diseases while drugs with 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] (Figure 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 of gene expression profiles. In addition to the expansion in the number of pharmacological perturbagens (over 5,000 compounds), one of the major novelties of the new CMap will be incorporation of genetic perturbations, that is gene expression profiles obtained by up-regulation or down-regulation 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 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 allowed Zhong *at al* to propose a combination with angiotensin-converting enzyme inhibitors and histone deacetylase inhibitors as a renoprotective 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 PGJ_2 [19]. This finding incited subsequent experiments that showed how PGJ_2 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 HSP90 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]; whilst this analysis produced predictable associations, e.g. with non-steroidal anti-inflammatory drugs and glucocorticoids, unexpected associations also emerged. In particular, the positive association with histone

deacetylase inhibitors (HDACIs) brought us to investigate whether a functional and mechanistic link between AnxA1 and HDACIs could exist. Further experimentation made us conclude that AnxA1 contribute to the anti-inflammatory mechanism of action of HDACIs [23].

Though innovative and promising, the CMap strategy is however not devoid of limitations, although 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/tissues as those used in the CMap. In addition, different treatment durations can lead to different results due to feedback regulation of the target, for example when studying G-protein coupled receptors. Other non-biological phenomenon such as the "batch effect", which affects the microarrays, compounds and cell 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 MANTRA (Mode of Action by NeTwoRk Analysis) allows analysis of the CMap data with an innovative approach that takes into consideration the variability in the transcriptional responses to the drug due to cell-line specific effects, different concentrations of drug applied and distinct experimental conditions [25]. Another example is DvD (Drug versus Disease), a new tool that combines together 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 analyzed, 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' analysis conducted over the last decade, it is striking and concerning the low number of discoveries that have been translated into practice. 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 2 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 highest impact factors revealed that only 70% require a mandatory public deposition of microarrays data to guarantee publication. More surprisingly, even if journals were subjected to data availability policies, 59% of the articles analysed did not fully adhere to their requirements [29]. Scientific journals should fully adhere to data-sharing policies to ensure reproducibility as a cornerstone of the scientific process. Because CMap studies are based on a selection of a number of up- and down-regulated genes obtained from previously conducted microarray analyses, the selection criteria and the list of genes used for the analysis should be available to ensure transparency and reproducibility.

Other publication practices might also be considered, such as the general tendency to publish the more spectacular results, which might be not fully representative of the true 'real-life' result. Journals should allow and promote publication of independent re-analysis and confirmation studies, not only initial evidence, as replication is essential for the consolidation of scientific knowledge and its eventual translation. In addition, underestimation and general refusal of negative data also distorts the real picture [30, 31]. From the bench side, a more accurate communication of microarray data is needed, although this aspect has improved thanks to MIAME (minimum information about a microarray experiment), consisting of a number of recommendations on the information that needs to be provided to enable the unambiguous interpretation of microarray-based experimental results [32].

Challenges and future directions

Despite its slow starting, we truly believe that integration of "omics' into the drug development process and clinical practice will become a reality in future years. Innovative tools and

databases promoting the re-use of publicly available information provide new opportunities for drug development at a low cost [33]. Initiatives like the Connectivity Map described here provide publicly available tools to extract useful information from whole-genome studies, often not fully exploited in part due to the difficulty associated to the analysis of large amount of information. Addition of more gene expression signatures representing more drugs and more cell lines, as it will happen with forthcoming CMap versions, would increase its usefulness. Data-sharing policies should be fully implemented and Journals should encourage authors to submit sufficient details to allow independent assessments of their findings. This transparency is of vital importance for the performance of meta-analysis, which might help to overcome the variation between individual studies.

In conclusion, costs and objective difficulties associated with the drug discovery process require innovative approaches, where the benefits of available information is maximised. In this sense, drug repositioning and identification of new mechanisms represent a low-cost process since making use of already developed therapeutics: these have often been used in humans, therefore facilitating rapid testing in clinical settings and rapid completion of drug repositioning. The CMap can be of great help for this, even more if potentiated with more meaningful protocols (e.g. use of primary cells). On the other hand, an organized multi-disciplinary effort is needed, from basic scientists, clinicians, research journals and regulatory bodies, to make the concept of translational medicine a reality and not a future perspective. An effort by bio-informatics to make these powerful tools easy to use and to interpret by basic scientists (biologists, pharmacologists...) will also be desirable. This must be our priority considering that the ultimate goal of drug development is improvement of the quality of life of patients. And sooner or later, we all will be patients!

Conflict of interest

The authors declare no conflict of interest.

References

1 Kaitin, K.I. (2010) Deconstructing the drug development process: the new face of innovation. *Clinical pharmacology and therapeutics* 87, 356-361

2 Kaitin, K.I. (2010) The Landscape for Pharmaceutical Innovation: Drivers of Cost-Effective Clinical Research. *Pharmaceutical outsourcing* 2010

3 Milne, C.P. and Kaitin, K.I. (2009) Translational medicine: an engine of change for bringing new technology to community health. *Science translational medicine* 1, 5cm5

4 Loscalzo, J., *et al.* (2007) Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Molecular systems biology* 3, 124

5 Alizadeh, A.A., *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503-511

6 Roses, A.D. (2004) Pharmacogenetics and drug development: the path to safer and more effective drugs. *Nature reviews. Genetics* 5, 645-656

7 Dudley, J.T., *et al.* (2010) Drug discovery in a multidimensional world: systems, patterns, and networks. *Journal of cardiovascular translational research* 3, 438-447

8 Al-Shahrour, F., *et al.* (2007) From genes to functional classes in the study of biological systems. *BMC bioinformatics* 8, 114

9 Joyce, A.R. and Palsson, B.O. (2006) The model organism as a system: integrating 'omics' data sets. *Nature reviews. Molecular cell biology* 7, 198-210

10 Lamb, J., *et al.* (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929-1935

11 Lamb, J. (2007) The Connectivity Map: a new tool for biomedical research. *Nature reviews. Cancer* 7, 54-60

12 Gullans, S.R. (2006) Connecting the dots using gene-expression profiles. *The New England journal of medicine* 355, 2042-2044

13 Lussier, Y.A. and Chen, J.L. (2011) The emergence of genome-based drug repositioning. *Science translational medicine* 3, 96ps35

14 Sirota, M., *et al.* (2011) Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Science translational medicine* 3, 96ra77

15 Dudley, J.T., *et al.* (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Science translational medicine* 3, 96ra76

16 Claerhout, S., *et al.* (2011) Gene expression signature analysis identifies vorinostat as a candidate therapy for gastric cancer. *PloS one* 6, e24662

17 Hassane, D.C., *et al.* (2008) Discovery of agents that eradicate leukemia stem cells using an in silico screen of public gene expression data. *Blood* 111, 5654-5662

18 Zhong, Y., *et al.* (2013) Renoprotective effect of combined inhibition of angiotensin-converting enzyme and histone deacetylase. *Journal of the American Society of Nephrology : JASN* 24, 801-811

19 Zimmer, M., *et al.* (2010) The connectivity map links iron regulatory protein-1-mediated inhibition of hypoxia-inducible factor-2a translation to the anti-inflammatory 15-deoxy-delta12,14-prostaglandin J2. *Cancer research* 70, 3071-3079

20 Hieronymus, H., *et al.* (2006) Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer cell* 10, 321-330

21 Wen, Z., *et al.* (2011) Discovery of molecular mechanisms of traditional Chinese medicinal formula Si-Wu-Tang using gene expression microarray and connectivity map. *PloS one* 6, e18278

22 Renshaw, D., *et al.* (2010) Downstream gene activation of the receptor ALX by the agonist annexin A1. *PloS one* 5

23 Montero-Melendez, T., *et al.* (2013) Gene expression signature-based approach identifies a pro-resolving mechanism of action for histone deacetylase inhibitors. *Cell death and differentiation* 20, 567-575
24 Iskar, M., *et al.* (2010) Drug-induced regulation of target expression. *PLoS computational biology* 6
25 Iorio, F., *et al.* (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proceedings of the National Academy of Sciences of the United States of America* 107, 14621-14626
26 Pacini, C., *et al.* (2013) DvD: An R/Cytoscape pipeline for drug repurposing using public repositories of gene expression data. *Bioinformatics* 29, 132-134
27 Woollard, P.M., *et al.* (2011) The application of next-generation sequencing technologies to drug discovery and development. *Drug discovery today* 16, 512-519
28 Ioannidis, J.P., *et al.* (2009) Repeatability of published microarray gene expression analyses. *Nature genetics* 41, 149-155
29 Alsheikh-Ali, A.A., *et al.* (2011) Public availability of published research data in high-impact journals. *PloS one* 6, e24357
30 Young, N.S., *et al.* (2008) Why current publication practices may distort science. *PLoS medicine* 5, e201
31 Ioannidis, J.P. (2005) Why most published research findings are false. *PLoS medicine* 2, e124
32 Brazma, A., *et al.* (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature genetics* 29, 365-371
33 Rung, J. and Brazma, A. (2013) Reuse of public genome-wide gene expression data. *Nature reviews. Genetics* 14, 89-99

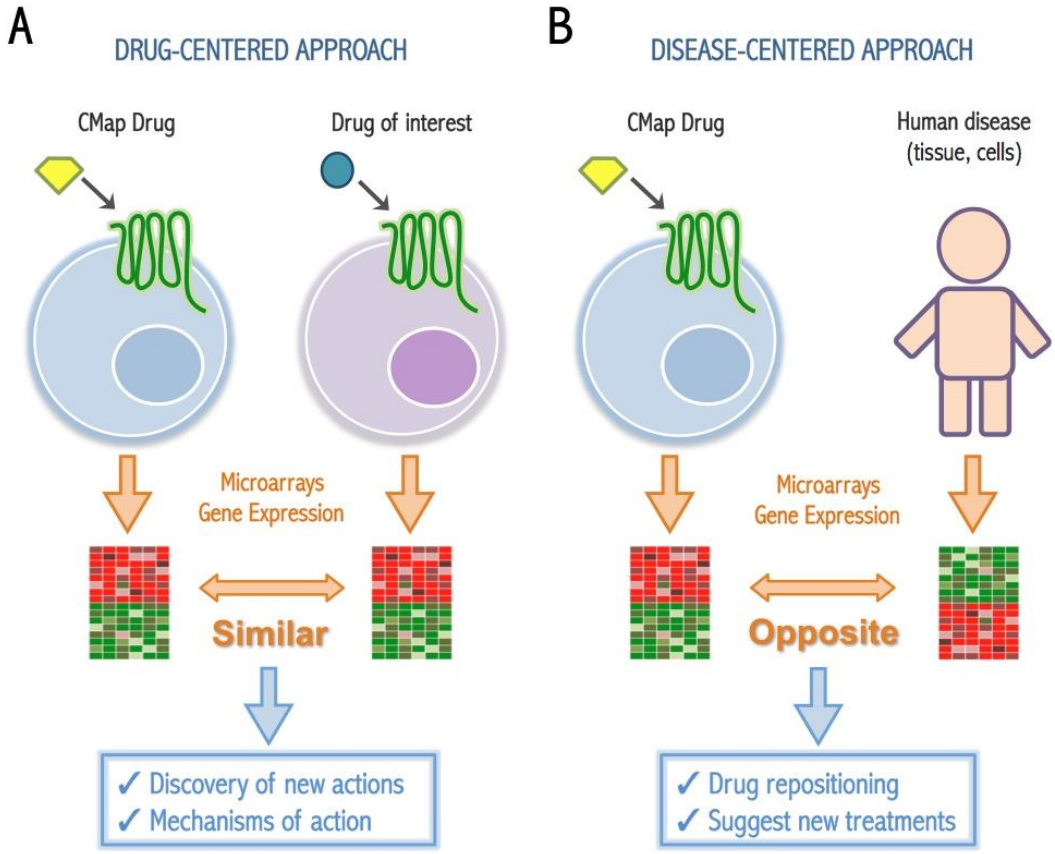
Figure Legends

Figure 1. The Connectivity Map concept. The Connectivity Map (CMap build 02) is a database that contains the gene expression signatures (obtained with the Affymetrix Genechip HG-U133A) of more than 1,300 bioactive molecules. Differentially expressed genes were identified by comparing cells treated with each distinct drug with untreated cells. A gene expression signature of interest (e.g. of a drug on a particular cell type (A) or a disease (B)) can be compared with those contained in the CMap database. If the signatures compared are similar (that will be identified by a 'positive' score), this could potentially be used to predict novel actions or suggest mechanism of actions of known or novel compounds. On the other hand, comparisons with a disease signature and identification of a 'negative' score (i.e. the gene signatures are the opposite) could be used for drug repositioning studies or to suggest new

294 treatments for that disease. Experimental validation is further required to confirm hypothesis or
295 predictions furnished by the CMap.
296

297 Figure 1

298



299