Genomics in target and drug discovery

M. van Duin1, H. Woolson, D. Mallinson and D. Black
Organon Research Laboratories, Department of Target Discovery, Newhouse, Lanarkshire ML1 5SH, Scotland, U.K.

Abstract
Genomics-based discovery of novel therapeutic drug targets requires the design of well-controlled biological or pharmacological experiments with experimental questions and hypotheses that relate to the therapeutic area of interest. This will aid the validation level of differentially expressed genes and hence facilitate the de-selection of the genes that are identified in microarray experiments. We here provide an example of how this approach is followed in the manipulation of human macrophage foam cells towards the discovery of novel drug targets for treatment of atherosclerosis.

Introduction
One of the big challenges that the pharmaceutical industry is facing is to take advantage of the massive amount of information that can now be gathered through the technological developments of the last few decades. This has resulted in an ‘…omics era’ in which genomics is defined as the elucidation and functional understanding of the genes encoded by the human genome. Likewise, related fields have emerged such as transcriptomics, proteomics and even metabolism-related metabolomics. Together these fields cover the various aspects underlying the molecular mechanisms of biological and pathological processes. More recently, the awareness has grown that this requires the full integration of data and knowledge, which is now also referred to as systems biology [1].

The expectations regarding the potential of the various ‘…omics’ areas for the modern drug industry were particularly fuelled by the observation that to date only less than 500 drug targets act in the mechanism of action of currently known drugs [2,3]. Further analysis of that information revealed that families of particular gene products in subsets of drug targets were more amenable as drug targets than others. This has led to the principle of druggability and indeed the ‘ear marking’ of the gene families encoding G-protein-coupled receptors [4], nuclear receptors [5], ion channels [6] and kinases [7] and other enzymes as prime-candidate drug targets in current target-hunting research.

The recent elucidation of the human genome [8,9] has now opened the possibility to analyse how many of these potentially interesting drug targets are encoded by the 30 000–35 000 genes on the human chromosomes. A summary of the number of human genes in the various drug-friendly gene families was recently provided by Hopkins and Groom [10]. Altogether, the members of the various drug-target families now amount to about 3000 genes. Interestingly, further analysis of these data, including the question whether pharmaceutical compounds actually bind to the presumed target, revealed that the frequently quoted number of druggable genes [2,3] is in fact much lower and possibly only about 25% of the 483 targets originally identified [2,3]. Identification of the binding pockets for the existing ligands in these real drug targets has the potential to further predict the success in high-throughput screening efforts with novel drug targets. Furthermore, this information may be used to further steer and optimize the nature of the chemical matter in compound libraries and to progress strategies to arrive at gene-family-specific or at least focused chemical libraries.

Whichever analysis of the number of actual drug targets is correct, the available data strongly suggest that the present number of known and well-validated drug targets is still relatively small. Simply looking at the number of identified G-protein-coupled receptors, nuclear receptors, ion channels and various types of enzyme, the validity of at least 2500 candidate drug targets remains to be elucidated. Future target-validation efforts are required to link them to the aetiology of known diseases and/or to demonstrate that the novel targets have relevant therapeutic potential.

The feasibility by which novel drug targets can be (de-)selected in genomics-driven discovery experiments is dependent on (i) a good hypothesis that can be addressed through experimental manipulation and (ii) the availability of good biological models. Regarding the latter, this includes either in vitro- or in vivo-generated samples from human and/or animal origin. The extent to which these materials will be available is very much dependent on the therapeutic area of interest. Inflammation genomics research by manipulating human lymphocytes is clearly much more feasible than assessing gene-expression differences in the human brain in the case of psychiatric research towards novel anti-depressants. Although it is considered a great advantage to have unlimited access to human biological material, the value of that material for proper and well-controlled experiments can also be questionable. Intrinsic problems with the use of some human tissues are that the biological material has been obtained post-mortem and in addition there is variation between subjects that are used for sampling. Whether or not samples have been taken from living donors or have been obtained post-mortem, there will always be

Key words: atherosclerosis, genome, microarrays, target discovery
Abbreviations used: LDL, low-density lipoprotein; acLDL, acetylated LDL.
1To whom correspondence should be addressed (e-mail marcel.vanduin@organon.com).
a large variation in terms of genetic background, medical history, diet and behavioural aspects for which the design of controls is a major challenge. Of course, several of these issues can be overcome by increasing the number of samples or careful patient selection; however, in many cases this is not a realistic option. Therefore, it seems fair to conclude that the use of human brain samples for experiments that are directly focused to discover differentially expressed genes should be done with care. In fact, the gene-expression data from such types of material seem to be much more suited to provide a gene-expression reference database. The gene-expression information in this database is likely to be used for the formulation of hypotheses that require subsequent experimental follow up. Alternatively, the expression data from human samples that are difficult to validate and standardize may be used in the bioinformatical analysis and/or verification of data that have been obtained in hypothesis-driven microarray discovery studies with cellular and/or animal models.

**Genomics-driven discovery in atherosclerosis**

In the case of cardiovascular research there are ample opportunities to design genomics experiments that are tailored towards the identification of novel drug targets in vascular diseases. Coronary atherosclerosis is the fifth leading cause of global disease burden, the leading cause in developed societies [11] and a research area with a clear and unmet medical need. The causal relationship between high plasma concentrations of cholesterol, particularly low-density lipoprotein (LDL) cholesterol, as one of the principal risk factors is now well established [12,13]. However, despite changes in lifestyle and the use of pharmacological approaches to lower plasma cholesterol, cardiovascular disease is still the principal cause of death in the U.S.A., Europe and much of Asia [12,13]. It is now well established that the development of atherosclerotic lesions in human arteries can be regarded as a form of chronic inflammation in response to injury of the vascular endothelium [12]. Hence, atherogenesis can be viewed as an inflammatory disease, which is initiated and progresses in the context of hypercholesterolaemia [13,14].

Macrophages are a hallmark of atherosclerosis at every stage of the disease [15] and as such can be regarded as a key cellular player in the generation, progression and, ultimately, rupture of atherosclerotic plaques. The initial stage of atherosclerosis, the fatty streak, is made up exclusively of macrophage foam cells, which are formed by the uptake of modified (e.g. oxidized) LDL by macrophages. Macrophage foam cells are also present in both early and late lesions [16,17]. Macrophages are also a dominant feature of unstable atherosclerotic plaques, the major risk factor in atherosclerosis. Such plaques contain many activated macrophages, especially within the shoulder regions of the plaque [16–18]. Hence, macrophage foam cells, macrophages and activated macrophages can be considered important and relevant cellular targets for therapeutic intervention [15–19]. Formation of macrophage foam cells and the activation of macrophages are mediated by changes in gene expression in response to various stimuli. Provided that experiments can be designed in which macrophages can be activated in a manner that is hypothesized to be akin to the pathological situation, the use of microarrays in genomics experiments allows the analysis of global changes in gene expression in a high-throughput manner. As such, genomics experiments with macrophages designed to assess global changes in gene expression mediated by modified LDL and inflammatory mediators can potentially lead to the identification of candidate molecular drug targets. Below, an example is given of a genomics experiment using a predictive in vitro cellular model of human macrophage foam cell formation.

**Experimental design**

Macrophage foam cells can generated in vitro by treatment of monocyte-derived macrophages or monocyte/macrophage cell lines such as THP-1 cells with oxidized LDL or acetylated LDL (acLDL) [20,21]. Accumulation of these forms of modified LDL have been demonstrated to have a diverse range of effects on macrophage function at the mRNA, protein and cellular function levels, and these cellular model systems appear to be predictive for events occurring in lesions in vivo [20–24]. However, using these model cellular systems there are only a limited number of studies examining changes in gene expression in a large-scale manner [23,25], and this is even more limited when used in conjunction with microarrays [22]. Therefore, an experiment was designed with Affymetrix® human GeneChips to assess global changes in gene expression using an in vitro foam cell model system. In two independent experiments THP-1 macrophages were treated with acLDL for periods from 6 to 72 h to allow gene expression profiling in a time-series experiment. Throughout the experiment accumulation of lipid was monitored and confirmed by Oil Red O staining of parallel culture slides. This demonstrated that the acLDL-treated cells accumulated significant lipid from 24 h onwards (see a typical example of the stained cells in Figure 1). Total RNA was extracted for each time point for control and treated cultures using Trizol® (Invitrogen), and then passed through RNasey® (Qiagen) spin columns. RNA was quantified by Ribogreen® (Molecular Probes) and was quality-controlled by analysis on a Bioanalyser 2100 (Agilent). Each RNA sample was divided into three separate tubes and each hybridized to an Affymetrix® HG_133 array A. Established Affymetrix® protocols were used for cDNA and cRNA synthesis as well as for the hybridization, washing and staining steps. Raw array data were captured by array scanner and quality-controlled by Affymetrix® MAS5 software. The generated data were subsequently exported into microarray data-storage and analysis software Resolver® (Rosetta). Triplicate hybridizations were analysed for variation using this software and the results of each of these triplets were combined for further analysis. As an example of how the array data can be further analysed, Resolver® software was used to identify genes that were significantly differentially...
expressed at the 72 h time point using a stringent cut off ($P \leq 0.0001$; see Figure 2). A data set of genes differentially expressed in both biological replicates has been generated. A total of 183 significantly differentially expressed genes were identified for further analysis. Using the Gene Ontology Index these genes could be grouped according to function. Among the genes that were significantly down-regulated (circled in Figure 2) was a group concerned with cholesterol biosynthesis and associated regulatory proteins. This is the expected normal cellular response to high intracellular cholesterol levels [19] and similar results have been reported from a microarray study using oxidized LDL and THP-1 macrophages [22]. This result confirms and validates THP-1 macrophages loaded with acLDL as a predictive model...
cellular system for foam cells and hence a valuable biological tool in research underlying atherosclerotic plaque formation.

Concluding remarks
Following the bioinformatics-driven de-selection process of differentially expressed genes, using for example criteria such as gene annotation and function, additional criteria need to be applied in the target identification and validation process. From a target- and drug-discovery point of view aspects like assayability and chemical tractability based on membership of a family of well-known drug targets are of paramount importance in this selection. The selected genes fulfilling these criteria can be followed up by confirming differential expression by quantitative PCR and subsequently by confirming expression and tissue distribution in human atherosclerotic plaques by quantitative PCR, immunohistochemistry and in situ hybridization. Other ways in which array data can be utilized include analysis of the time series, which allows one to follow changes in gene expression at different time points and also over time. Such data can be clustered to allow identification of biochemical pathways, which may be therapeutically exploited and/or lead to further hypotheses as the basis for follow-up experiments. Indeed, microarray-based genomics experiments cannot only identify novel candidate molecular targets and biochemical pathways that may be therapeutically exploited, but also increase our understanding of the biology of a disease process, which may ultimately lead to the identification of other potential drug targets.

We are grateful to David Pulford, Mark Anderson and Julie McPhee for the microarray work and thank Donald Dunbar and Fiona Sinclair for bioinformatics support.

References
20 Kruth, H.S. (2001) Frontiers Biosci. 6, d429-d455

Received 20 November 2002