DRUG DISCOVERY TODAY TARGETS

The transcriptome's drugable frequenters

Koen J. Dechering

Microarray studies are widely employed in the exploratory phase of the drug discovery process. Expectations raised by the genomics revolution led to the belief that they would rapidly lead to the identification of novel drug targets. However, a few basic questions were often overlooked. Are members of drugable gene families properly represented in the transcriptome? Or are they poorly expressed and below the detection limit of the microarray technology? This review explores the representation of drug targets and components of downstream cellular signaling pathways in the transcriptome. It appears that members of drugable gene families are underrepresented in the transcriptomes of non-pathological human tissues. But, they are represented at or above the expected frequency in the differential transcriptome (i.e. the set of genes that changes expression upon a change in cellular environment). Analysis of differential gene expression on a genome-wide scale will therefore give a comprehensive overview of cellular pathways and possible drug targets.

The pharmaceutical industry embraced DNA microarray technology when it was first introduced in the mid 1990s [1]. This can be explained by the gene-consuming nature of the modern drug discovery process: the cumulative attrition rate of this process is so high that a large number of genes has to be fed into the discovery pipeline in order to introduce a single new chemical entity on the market. Typically, a small molecule discovery pipeline consists of target identification and validation, lead identification, lead optimization, preclinical development, three phases of clinical development and, finally, registration. A recent analysis showed that each step in the process has an approximate success rate of 60% [2]. Consequently, the cumulative success rate amounts to $(0.6)^8 = 0.017$. In other words, for every 60 drug targets identified only one new chemical entity enters the market. It is therefore not surprising that high-throughput research techniques are highly welcomed by the pharmaceutical industry. Following the introduction of DNA microarray technology by academic groups, the pharmaceutical industry rapidly incorporated this high-throughput technology into the drug discovery process. This is reflected in the relative number of publications on microarrays (Figure 1). In the percentage of publications featuring microarrays, the top ten pharmaceutical companies took a head start compared with academia. For example, ~1% of the publications published by pharmaceutical companies in 2001 contained the word microarray in the title or abstract; in that year, ~0.08% of the publications originating from academia described the use of microarray technology. It should be noted, however, that there is a tremendous difference in absolute numbers of publications produced by the two institutional groups. Since 1995, 10 560 papers featured microarray technology. Of these, 169 were contributed by the top ten pharmaceutical companies.

The vast number of publications show that microarrays have become commodities and are

K.J. Dechering
N.V. Organon,
Department of Target,
Discovery, Room RE2106,
PO Box 20,
5340 BH Oss,
The Netherlands
e-mail:
koen.dechering@organon.com

1359-6446/05/$ – see front matter PII: S1359-6446(05)03451-3

widely applied by the scientific community. One area where microarrays have had an enormous impact is in disease classification and prediction of clinical outcome [3,4]. Usually, the expression of a single marker is not predictive for prognosis of a complex disease, but predictive power greatly improves when a set of genes is analyzed. Advanced statistical techniques have been used to identify gene signatures that correlate with disease outcome [5,6]. Even complex traits such as behavior associate with specific gene expression patterns in the brain and could be predicted from microarray data [7].

In drug target discovery, microarrays are used to identify genes that are expressed under relevant (pathological) conditions [8]. The dynamics of the transcriptome are often studied to unravel signal transduction pathways or gene function (see Box 1 for a definition of 'omes'). It has been convincingly shown that the dynamics of gene expression contain information about gene function [9,10]. Using mathematical techniques to group genes according to the similarity of their expression profile yields clusters that are enriched for genes involved in a specific process. Vice versa, this principle can be used to predict functions for 'novel' genes that reside in a cluster enriched for certain gene functions [11]. Predicting gene function from studying the dynamics of gene expression may not work for all classes of genes. Some genes are constitutively expressed in all tissues and under all (patho)physiological conditions. Such genes are referred to as housekeeping genes, although it appears that the 'traditional' housekeeping genes such as glyceraldehyde-4-phosphate dehydrogenase, β-actin and β-tubulin do change expression under some conditions [12]. Nevertheless, a genome-wide study of expression levels showed that the concept of housekeeping genes still holds, as a subset of all genes encoded by the genome is indeed constitutively expressed [13]. On the other side of the spectrum reside 'noise' genes that show erratic expression levels apparently not related to cellular state. This was illustrated in a study of yeast in which gene expression was measured in 63 identical control cultures. These analyses showed that at least 278 out of the 5835 genes showed transcriptional fluctuations unrelated to culture or experimental conditions [11]. These 278 'noise' genes did not belong to a single gene family, but instead encoded a variety of protein functions.

This review takes a closer look at expression of drug targets and explores the frequencies of genes from different functional categories in transcriptome data derived from microarray studies.

The transcriptome versus the proteome

Drug target discovery often starts with a transcriptomics study to identify genes that show altered expression during, for instance, a disease condition. Similarly, proteomics monitors alterations in the protein content of samples. Several studies have addressed the global relationship between protein and cognate messenger RNA (mRNA) levels by regression analysis or by calculation of a correlation coefficient such as Pearson's r [14–16]. The results showed that protein and mRNA levels hardly correlate, if they correlate at all. This is surprising as proteins are translated from mRNA and, therefore, one might assume some kind of correlation between the two entities. A possible problem related to studies comparing protein and mRNA levels is that the methods used to study the translatome and the transcriptome often have different sensitivities. For example, correlation is lost when a group of gene products is above the detection limit of one technology but fails to be reliably detected by the other. Chen and coworkers investigated whether levels of proteins that are well above the detection limit show a good correlation with their cognate mRNA levels [17]. To this end, they investigated the correlation between mRNA and protein levels for separate subsets of proteins with differing levels of abundance. The result showed poor correlations for all subsets.

The lack of correlation between mRNA and protein levels is not surprising. Both

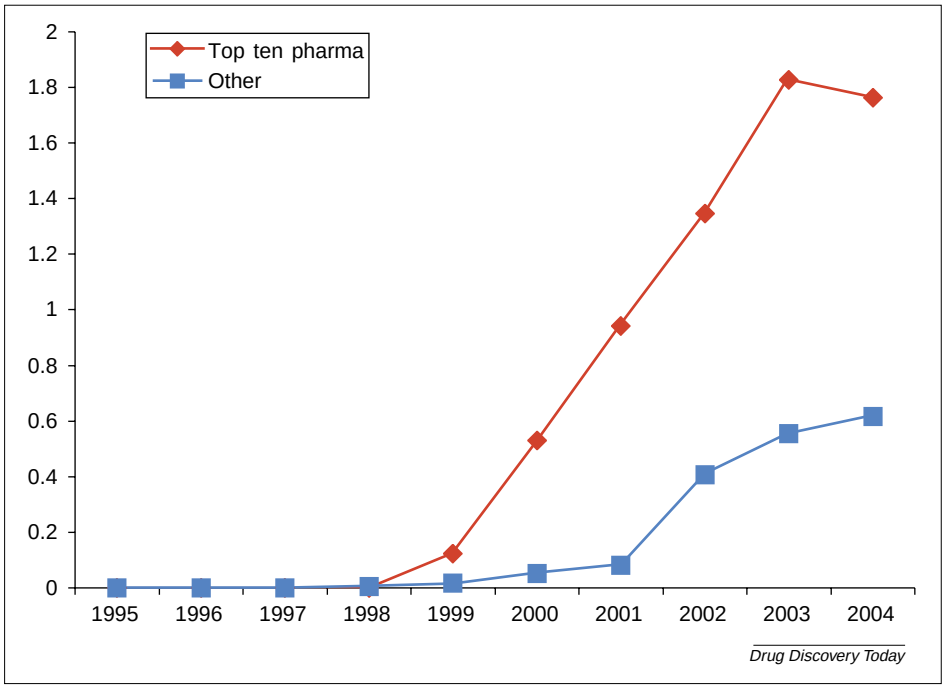

FIGURE 1

Percentages of publications by the top ten pharmaceutical companies and other institutions featuring microarray technology. The number of publications on microarrays is expressed as a percentage of the total number of publications produced by an institutional group (top ten pharma or 'other'). Publication numbers per year were determined from the Medline database. Publications originating from the top ten pharmaceutical companies (Pfizer Inc., Johnson and Johnson, GlaxoSmithKline Plc, Novartis AG, Roche Group, Merck and Co., Inc., Bristol-Myers Squibb Company, Sanofi-Aventis, Abbott Laboratories, and AstraZeneca Plc) were identified through the 'institution' field. Publications on microarrays were identified using 'microarray*' as a query in the 'text word' field.

BOX 1

'Omes'

Definitions for the various 'omes' vary. This paper uses the definitions as presented on http://bioinfo.mbb.yale.edu/what-is-it/omes

Transcriptome: the population of mRNA transcripts in the cell, weighted by their expression levels.
Proteome: the set of protein-coding regions in the genome.
Translatome: the population of proteins in the cell, weighted by their expression levels.

In addition, this paper distinguishes between the static transcriptome and the differential transcriptome. The differential transcriptome represents the set of genes that are differentially expressed during a cellular transition. The static transcriptome is the set of genes that does not change expression under that particular condition.

are controlled by individual rates of synthesis and turnover. The synthesis route from RNA to protein consists of a set of biochemical equilibrium reactions that are determined by cognate equilibrium constants. These constants are unique for every protein synthesis route as they are determined by, for example, gene-specific chromatin structure, sequence context and binding of accessory proteins [18]. Furthermore, gene expression is driven by nanoscale reactions with low concentrations of reactants and is, therefore, of a stochastic nature [19]. As a consequence, gene expression occurs in bursts of transcription and translation and is intrinsically noisy [20,21]. Therefore, comparing the translatome and the transcriptome assuming a simple linear relationship with identical reaction constants for all genes is probably not appropriate. Models could be improved by studying the dynamics of gene expression. Such studies may provide insight into the individual equilibrium constants that determine protein and mRNA concentrations. Indeed, when studying the relationship between mRNA and protein levels, Lian and co-workers [22] saw a poor correlation using single time point data but a strong correlation when multiple time points were used to build a model. Apparently, steady-state levels of transcripts and proteins do not correlate, whereas the timing in changes of levels does correlate. This is consistent with the notion that the majority of changes in protein levels result from a change in mRNA levels.

In a sense, the discussion on expression levels and the comparison between RNA and protein levels is meaningless. There is no documented relationship between expression level and functional importance. Inspection of the Mouse Genome Informatics database of alleles and phenotypes (http://www.informatics.jax.org) reveals a variety of phenotypes linked to genes with varying expression levels, with no obvious relationship between gene expression level and phenotype. For example, complete ablation of the highly expressed genes LECT1 or CCR1 has no phenotypic effect [23,24], whereas for genes such as RUNX2 or DLL4 a twofold decrease in expression already leads to a severe phenotype [25,26]. The issue of expression levels is a technological rather than a biological issue. It only becomes relevant when expression of a gene product is below the detection limit of the current technology; it then escapes from our attention and subsequent functional analysis.

Instinctively, proteins are regarded as being functionally more relevant than mRNAs as the proteins contribute more directly to the phenotype. However, in whole genome approaches that study hundreds of gene products at the same time it is debatable whether one would obtain fundamentally distinct information from studying either the transcriptome or the translatome. Another way of comparing the translatome and proteome is to look at sets of genes grouped into functional categories, rather than analyzing individual gene products [27]. It turns out that when gene products are categorized in broader functional categories, the translatome and transcriptome show identical compositions whereas individual gene products may show different representations [28–30]. It can thus be argued that in a study aimed at the identification of novel biological functions or pathways the transcriptome provides as good a starting point as the translatome.

Microarrays in drug discovery

Drug discovery aimed at the development of low molecular weight chemical compounds specifically focuses on the part of the genome that is amenable to pharmacological intervention. Drugs currently on the market target a relatively small number of gene families, termed the 'drugable genome'. This subset of the genome is dominated by the G-protein-coupled receptor, kinase, protease, ion channel and nuclear hormone receptor gene families [31,32]. One strategy to identify novel drug targets is to mine the human genome for new members of the established families of drug targets [33,34]. Such gene identification efforts based on homology searches can then be integrated with other approaches (e.g. microarray studies) to accumulate evidence that the presumed drug target is indeed a disease-modifying gene product [35,36]. Alternatively, microarrays can be used *ab initio* to identify genes associated with pathology. This strategy may also be used to implicate genes of unknown annotation or function in a disease process [11]. The quality of the resulting candidate drug target is not easy to define. In the analysis of the research performance of the pharmaceutical industry, McKinsey & Company defined quality of a drug target as the chance that a drug directed against the target will reach the market, multiplied by the revenues it will generate (M. Baudouin, pers. commun.). In my view, this definition makes sense. It incorporates aspects of novelty, as this translates into revenues, drugability, and soundness of the biological concept, as this translates into attrition rates.

What is the impact of microarray technology on the quality of the candidate drug target? It is probably too

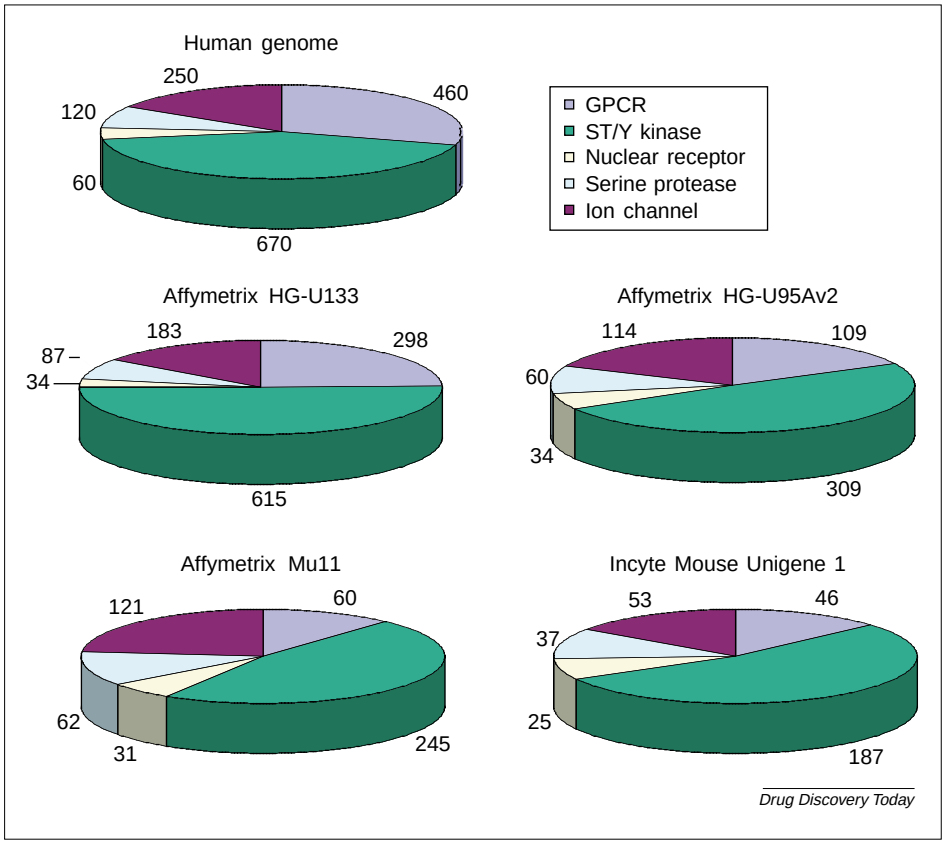

FIGURE 2

The distribution of drugable gene families in the human genome and on microarrays. The figure depicts the number of genes belonging to the gene families of typical drug targets. Data on representation in the human genome were taken from a published analysis of drugable genes [31]. Family members represented on the Affymetrix arrays were identified on the basis of their Interpro ID or Gene Ontology Molecular Function classification. Identification of drug target family members on the Incyte array was performed on the basis of the Protein Functional Hierarchy classification provided by Incyte Genomics. Abbreviations: GPCR, G-protein-coupled receptor (only the rhodopsin-like GPCRs, the secretin-like GPCRs, and the metabotropic glutamate receptor families were taken into account); ST/Y kinase, serine-threonine/tyrosine kinase.

early to answer this question. Research using microarray technology only took off in the late 1990s. Given the current development timelines in the pharmaceutical industry, the first drugs targeting gene products identified by microarray technology have yet to hit the market [37]. Therefore, an analysis of revenues cannot be made yet. Also, a proper assessment of attrition rates is hard to make as the statistics are too poor. We can, however, address a more fundamental issue. Do microarrays have the potential to identify novel drug targets? Expectations raised by the genomics revolution provided a simple yes to this question. However, a few basic questions are often overlooked. Are members of drugable gene families properly represented in the transcriptome? Or are they poorly expressed and below the detection limit of the current technology?

Gene frequencies in transcriptome data

Drug target representation on microarrays has improved over the years and modern microarrays such as the Affymetrix HG-U133 array provide a comprehensive view on drugable gene families (Figure 2). Both commercial and public efforts have been undertaken to create comprehensive repositories of microarray data [38–40]. These databases allow the analysis of gene frequencies in various transcriptomes [41,42]. Figure 3 shows the results of an analysis of the GeneExpress database containing Affymetrix HG-U133 genechip data that were generated on samples derived from normal (non-diseased) human tissues. Observed gene frequencies were compared with expected frequencies assuming a random distribution of genes belonging to a particular functional class across the set of expressed genes. For example, when 60% of all genes were detected above background, it was expected that 60% of the kinases on the array would also be detected above background. Data were expressed as an observed/expected ratio as indicated in the figure legend. A deviation from 1 indicates a deviation from the null hypothesis that all gene classes are equally represented in the set of expressed genes. Figure 3a shows the ratio of observed over expected frequencies for five classes of drugable gene families and, for comparison, transcription factors. The analysis shows that the serine-theonine/tyrosine kinases (ST/Y kinases) are represented at the expected frequency, whereas all the other gene groups are underrepresented. The G-protein-coupled receptors (GPCRs), ion channels and serine proteases are particularly infrequent in the transcriptomes analyzed here. This indicates that genes encoding these drug targets are less frequently expressed than the average gene in the transcriptome. This can be explained by the fact that some genes (e.g. housekeeping genes) are constitutively expressed under all conditions, whereas other genes are only expressed upon changes in cellular environment or in specific cells [13].

With respect to genes that are part of pathways that are targeted for drug intervention, most of these pathways are represented at or above the expected frequency (Fig. 3b). A notable exception is formed by the G-protein signaling pathway components that are underrepresented. Thus, with respect to G-protein signaling both the protein receptors and the downstream pathways are underrepresented. These findings seem to indicate that microarrays are not suited to the study of this important signal transduction route.

Gene frequencies in differential transcriptome data

Analyzing the dynamics of gene expression provides more information than the analysis of a single time point. Therefore, most microarray studies examine changes in

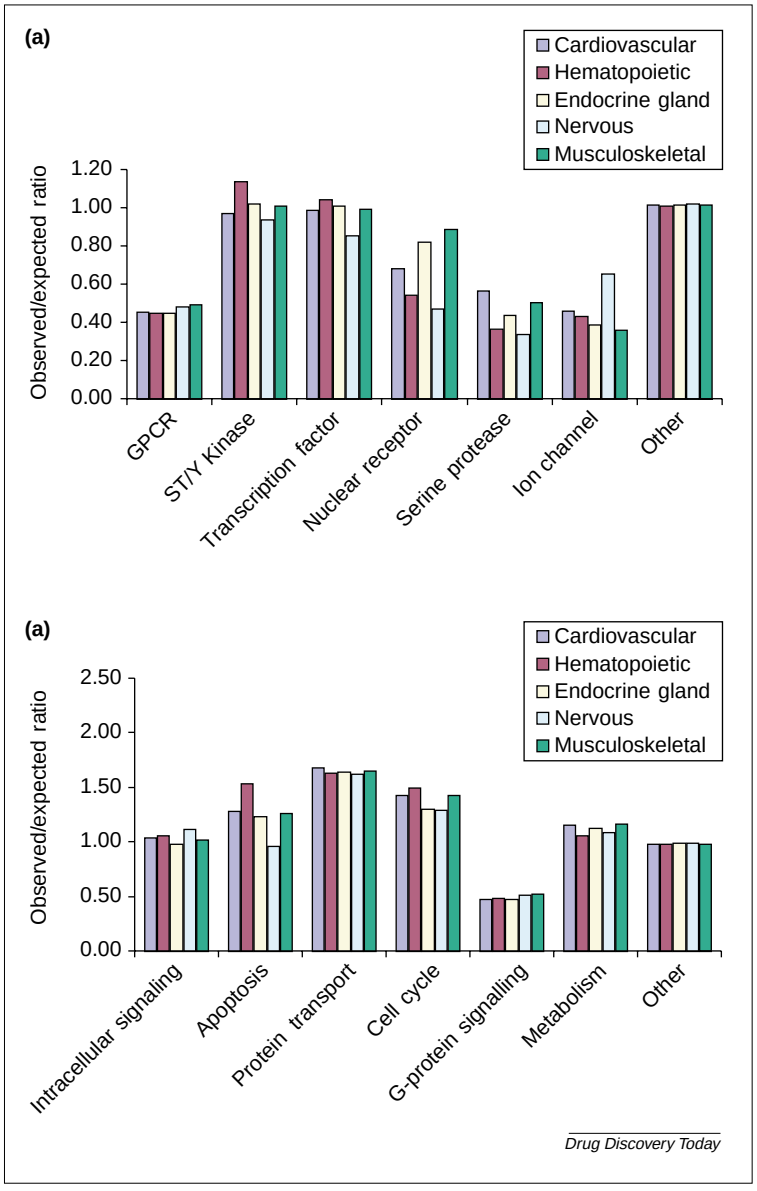

FIGURE 3

Frequencies of drug targets and pathway components in the static transcriptome. Data generated with the Affymetrix HG-U133 array was used to evaluate drug target frequencies in different transcriptomes. The figure shows frequencies of **(a)** drug targets and, for comparison, transcription factors and **(b)** pathway components. The analysis was based on Affymetrix Genechip HG-U133 data that were taken from the GeneExpress database (Genelogic, Gaithersburg, MD). Classification of tissue categories was performed based on the Systemized Nomenclature for Medicine (http://www.snomed.org). Genes were considered to be expressed when they were detected above background in at least 75% of the samples within a tissue category. Observed gene frequencies were compared with expected frequencies assuming a random distribution of genes belonging to a particular functional class across the set of expressed genes. Therefore, the expected frequency was defined as the [number of detected genes] / [total number of genes represented on the array]. Data are plotted as an observed/expected ratio as explained in the main text. The observed frequency for a particular class or subset of genes was defined as the [number of detected genes in the subset] / [total number of genes from the subset represented on the array]. Drug target family members were identified as indicated in the legend to Figure 2. Pathway components were identified based on Gene Ontology Biological Process classification. Abbreviations: GPCR, G-protein-coupled receptor; ST/Y kinase, serine-threonine/tyrosine kinase.

gene expression in response to an altered cellular environment or as a function of time. For instance, by comparing pathological conditions with normal conditions those genes can be selected that show a change in expression correlating with the disease condition. Therefore, in many studies the differential part of the transcriptome is more relevant than the static part. The differential part of the transcriptome may be defined as the set of genes that are differentially expressed during a cellular transition, whereas the static part does not change in expression. Are all genes and biological functions equally represented in this differential part of the transcriptome or do some functions remain hidden? The latter seems conceivable as, for instance, protein phosphorylation events in a signal transduction cascade are not reflected at the mRNA level. However, if the genes that are part of this signal transduction pathway alter in expression, a transcriptome analysis might be able to identify this pathway as relevant for the process under study. A similar question is relevant for drug target discovery: can all potential drug targets be identified equally well by microarray analysis, or are some targets not regulated at the RNA level? Again, the latter is conceivable. For instance, in the case of a receptor-mediated process the rate-limiting step for activation of the pathway might occur at the level of ligand availability and not at the level of expression of the receptor.

Figure 4 shows the analysis of gene frequencies in five different differential transcriptomes. Data were analyzed from five published studies (Table 1). These studies comprise time course analyses of changes in gene expression in response to external stimuli or during differentiation of cells. All studies used rigid statistical criteria for selection of differentially expressed genes and these gene sets, as defined by the authors, were included in this study. The different gene families show varying degrees of variation in the ratios of observed over expected frequencies among the different studies (Figure 4). The ST/Y kinases are relatively constantly represented at the expected frequency. Although the activity of many kinases is regulated by phosphorylation events, these data indicate that there is an additional level of control at the transcriptional level. In contrast to the transcriptome described in Figure 3, GPCRs are well presented in the differential transcriptome. They show overrepresentation in two published studies and mild underrepresentation in the other three datasets studied here. The groups of nuclear receptors, proteases and ion channels show more variation and are overrepresented in some studies yet underrepresented in others. In terms of biological processes, the response of fibroblasts to a pH change and the response of skeletal muscle to tumor necrosis factor α show a preponderance for drug-target-like molecules as these responses lead to an overrepresentation of such molecules in the differential transcriptome. Conversely, the differentiation of lymphocytes shows a general underrepresentation of drug targets. The variation in gene frequencies is also apparent in an analysis of pathway components (Fig. 4b). The differential transcriptomes show distinct over- or underrepresentation of transcripts

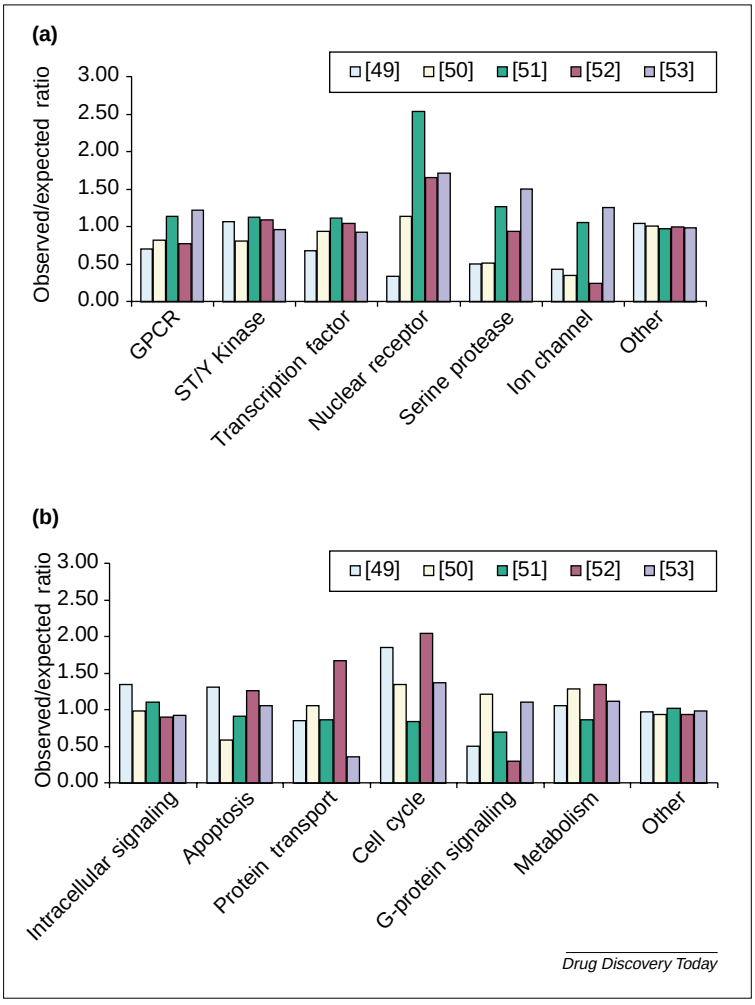

FIGURE 4

Frequencies of drug targets and pathway components in the differential transcriptome. Data from five published studies (Table 1 and references in the figure) were analyzed for gene frequencies. The figure shows frequencies of **(a)** drug targets and, for comparison, transcription factors and **(b)** pathway components. Frequencies are plotted as observed/expected ratios. Here the expected frequency was defined as the [size of the differential transcriptome] / [total number of genes represented on the array]. The observed frequency for a particular class or subset of genes was defined as the [number of genes from the subset present in the differential transcriptome] / [total number of genes from the subset represented on the array]. Drug target family members and pathway components were identified as indicated in the legends to Figures 2 and 3. Abbreviations: GPCR, G-protein-coupled receptor; ST/Y kinase, serine-threonine/tyrosine kinase.

encoding signaling components. This is not surprising, as the five studies discussed here studied a wide variety of cellular processes that all have distinct requirements for distinct gene functions. It appears, however, that none of the functional categories of genes shows a general absence in microarray datasets. All types of drug targets or signaling components are overrepresented in some type of study. This indicates that the type of study rather than the technological limits or intrinsic transcriptome properties determine the presence of drug targets in microarray datasets.

Pathway discovery

The meta-analysis of differential transcriptome data shows that all pathway components and classes of drug targets are represented in some study or another. Even cytoplasmic signaling leaves its fingerprint in the transcriptome. The mitogen-activated protein kinase (MAPK) signaling pathway is a classical example of a cytoplasmic phosphorylation cascade. Yet, Roberts and co-workers [43] were able to reveal the circuitry of MAPK signaling in yeast responding to pheromone using microarray expression profiling. Such studies indicate that DNA microarray analysis is well suited for exploratory studies aimed at the identification of signaling pathways underlying disease. There is an increasing belief that studying pathways rather than individual genes is a more fruitful way to identify novel drug targets [35,36]. Once a pathway is identified as being relevant for the disease, it is then possible to assess the individual components of the pathway for their drugability and disease-modifying properties. For many drugs on the market it is not entirely clear which molecular pathways they affect. Chemical genomics approaches may clarify the targeted pathway [44,45]. Such identification would then enable a 'son-of-target' approach in which downstream components are assessed as new targets in a pathway that has been clinically-proven to be an effective disease modulator [46].

Microarray analysis is of a descriptive nature and follow-up validation experiments are required to confirm gene function [35]. The aim of the microarray experiment is to help the researcher to make an educated guess and to reduce the attrition rates in subsequent functional studies as much as possible. New bioinformatics approaches could improve data analysis and functional predictions. An exciting avenue is the study of genetic networks [47]. These analyses employ algorithms that deduct causative relationships between gene activation events from microarray data [48]. It is likely that these approaches will greatly increase the quality of the hypotheses deducted from microarray analyses.

Conclusions

Genomics technology is widely applied and has delivered appealing examples of functional discoveries and predictive classifications. The present analysis of published microarray studies indicates that all functional classes of genes are represented in transcriptome data. The degree of representation depends on the type of study. It appears that studying the dynamics of gene expression might increase the chance of identifying a drugable gene product. No functional gene class shows a consistent underrepresentation in the current meta-analysis of the differential transcriptome. Therefore, the differential transcriptome provides a comprehensive view of the gene functions and cellular processes encoded by the genome.

Clearly, translatome and transcriptome data show discrepancies; however, these discrepancies arise at the level of individual genes and not at the level of broader functional categories. Parallel transcriptome and translatome studies are likely to identify the same pathway, although

TABLE 1

Datasets used to analyze the frequencies of drug targets and pathway components in the differential transcriptome

Type of study	Array	Size of differential transcriptome (number of genes)	Refs
Lymphocyte development	Affymetrix Mu11	1304	[49]
Osteoblast development	Incyte Mouse Unigene 1	1054	[50]
Response of skeletal muscle to TNFα	Affymetrix Mu11	1037	[51]
Senescence of papillomavirus-positive cervical cancer cells	Affymetrix HG-U95Av2	703	[52]
Response of fibroblasts to extracellular pH change	Affymetrix HG-U95Av2	2068	[53]

Data were selected from five published studies that contained large sets of differentially expressed genes. All studies used rigid statistical criteria for the selection of differentially expressed genes and these gene sets, as defined by the authors, were included in this study.
Abbreviation: TNFα, tumor necrosis factor α.

the individual pathway components they identify may differ. In an exploratory phase where the choice for a specific gene product still has to be made, the choice for either technology seems to be irrelevant. So far, there are no good examples where translatome versus transcriptome analyses have uncovered distinctively different pathways. The choice for either approach should merely be driven by the nature and availability of sample material and the throughput that is needed. Currently, DNA microarray technology has a higher throughput than proteomics technologies. Given the gene-consuming nature of the drug discovery process, microarray analysis seems a logical initial step in the identification of novel drug targets, pathways and biological concepts.

Acknowledgements

The author wishes to thank Susanne Bauerschmidt and Edwin van den Heuvel for helpful discussions, Susan Wells for sharing microarray data and Mark Gerstein for suggesting the term 'differential transcriptome'. Maarten Sollewijn Gelpke is gratefully acknowledged for SNOMED classification of the GeneExpress database.

References

1 Schena, M. *et al.* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470
2 Brown, D. and Superti-Furga, G. (2003) Rediscovering the sweet spot in drug discovery. *Drug Discov. Today* 8, 1067–1077
3 van't Veer, L.J. *et al.* (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536
4 Burczynski, M.E. *et al.* (2005) Transcriptional profiles in peripheral blood mononuclear cells prognostic of clinical outcomes in patients with advanced renal cell carcinoma. *Clin. Cancer Res.* 11, 1181–1189
5 Golub, T.R. *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537
6 Shipp, M.A. *et al.* (2002) Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.* 8, 68–74
7 Whitfield, C.W. *et al.* (2003) Gene expression profiles in the brain predict behavior in individual honey bees. *Science* 302, 296–299
8 van Duin, M. *et al.* (2003) Genomics in target and drug discovery. *Biochem. Soc. Transact.* 31, 429–432
9 Eisen, M.B. *et al.* (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14863–14868
10 Gibbons, F.D. and And Roth, F.P. (2002) Judging the quality of gene expression-based clustering methods using gene annotation. *Genome Res.* 12, 1574–1581
11 Hughes, T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126
12 Ferguson, R.E. *et al.* (2005) Housekeeping proteins: a preliminary study illustrating some limitations as useful references in protein expression studies. *Proteomics* 5, 566–571
13 Yanai, I. *et al.* (2005) Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21, 650–659
14 Izzotti, A. *et al.* (2004) Proteomic analysis as related to transcriptome data in the lung of chromium(VI)-treated rats. *Int. J. Oncol.* 24, 1513–1522
15 Greenbaum D. *et al.* (2003) Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* 4, 117.1-117.8
16 Lorenz, P. *et al.* (2003) From transcriptome to proteome: differentially expressed proteins identified in synovial tissue of patients suffering from rheumatoid arthritis and osteoarthritis by an initial screen with a panel of 791 antibodies. *Proteomics* 3, 991–1002
17 Chen, G. *et al.* (2002) Discordant protein and mRNA expression in lung adenocarcinomas. *Mol. Cell. Proteomics* 1, 304–313
18 McAdams, H.H. and And Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 94, 814–819
19 Salwinski, L. and Eisenberg, D. (2004) In silico simulation of biological network dynamics. *Nat. Biotechnol.* 22, 1017–1019
20 Ozbudak, E.M. *et al.* (2002) Regulation of noise in the expression of a single gene. *Nat. Genet.* 31, 69–73
21 Thattai, M. and van Oudenaarden, A. (2001) Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8614–8619
22 Lian, Z. (2002) Genomic and proteomic analysis of the myeloid differentiation program: global analysis of gene expression during induced differentiation in the MPRO cell line. *Blood* 100, 3209–3220
23 Brandau, O. *et al.* (2002) Chondromodulin I is dispensable during enchondral ossification and eye development. *Mol. Cell. Biol.* 22, 6627–6635
24 Kaesler, S. *et al.* (2004) The chemokine receptor CCR1 is strongly up-regulated after skin injury but dispensable for wound healing. *Wound Repair Regen.* 12, 193–204
25 Gale, N.W. *et al.* (2004) Haploinsufficiency of delta-like 4 ligand results in embryonic lethality due to major defects in arterial and vascular development. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15949–15954
26 Otto, F. *et al.* (1997) Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell* 89, 765–771
27 Greenbaum, D. *et al.* (2001) Interrelating different types of genomic data, from proteome to secretome: 'oming in on function. *Genome Res.* 11, 1463–1468
28 de Nobel, H. *et al.* (2001) Parallel and comparative analysis of the proteome and transcriptome of sorbic acid-stressed *Saccharomyces cerevisiae*. *Yeast* 18, 1413–1428
29 Jang, W.G. *et al.* (2004) Analysis of proteome and transcriptome of tumor necrosis factor α stimulated vascular smooth muscle cells with or without α lipoic acid. *Proteomics* 4, 3383–339
30 Greenbaum, D. *et al.* (2002) Analysis of mRNA expression and protein abundance data: an approach for the comparison of enrichment of features in the cellular population of proteins and transcripts. *Bioinformatics* 18, 585–596
31 Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730
32 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–1964
33 Hsu, S.Y. and And Hsueh, A.J. (2000) Discovering new hormones, receptors, and

signaling mediators in the genomic era. *Mol. Endocrinol.* 14, 594–604
34 Willson, T.M. and And Moore, J.T. (2002) Genomics versus orphan nuclear receptors – a half-time report. *Mol. Endocrinol.* 16, 1135–1144
35 Kramer, R. and Cohen, D. (2004) Functional genomics to new drug targets. *Nat. Rev. Drug Discov.* 3, 965–972
36 Slonim, D.K. (2002) From patterns to pathways: gene expression data analysis comes of age. *Nat. Genet.* 32, 502–508
37 Tufts Center for the Study of Drug Development (2004) Outlook 2004. URL: http://csdd.tufts.edu/InfoServices/OutlookReports.asp
38 Brazma, A. *et al.* (2003) ArrayExpress – a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* 31, 68–71
39 Edgar, R. *et al.* (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210
40 Sherlock, G. *et al.* (2001) The Stanford Microarray Database. *Nucleic Acids Res.* 29, 152–155
41 Shou, J. *et al.* (2002) Expression profiling of a human cell line model of prostatic cancer reveals a direct involvement of interferon signaling in prostate tumor progression. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2830–2835
42 Velasco, A.M. *et al.* (2004) Identification and validation of novel androgen-regulated genes in prostate cancer. *Endocrinology* 145, 3913–3924
43 Roberts, C.J. *et al.* (2000) Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* 287, 873–880
44 Ding, S. and Schulz, P.G. (2004) A role for chemistry in stem cell biology. *Nat. Biotechnol.* 22, 833–840
45 Clarke, P.A. *et al.* (2004) Gene expression microarray technologies in the development of new therapeutic agents. *Eur. J. Cancer* 40, 2560–2591
46 Marton, M.J. *et al.* (1998) Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat. Med.* 4, 1293–1301
47 Schmitt, W.A., Jr *et al.* (2004) Elucidation of gene interaction networks through time-lagged correlation analysis of transcriptional data. *Genome Res.* 14, 1654–1663
48 van Someren, E.P. *et al.* (2002) Genetic network modeling. *Pharmacogenomics* 3, 507–525
49 Hoffmann, R. *et al.* (2003) Rules for gene usage inferred from a comparison of large-scale gene expression profiles of T and B lymphocyte development. *J. Immunol.* 170, 1339–1353
50 Vaes, B.L. *et al.* (2002) Comprehensive microarray analysis of bone morphogenetic protein 2-induced osteoblast differentiation resulting in the identification of novel markers for bone development. *J. Bone Miner. Res.* 17, 2106–2118
51 Alon, T. *et al.* (2003) Cytokine-induced patterns of gene expression in skeletal muscle tissue. *J. Biol. Chem.* 278, 32324–32334
52 Wells, S.I. *et al.* (2003) Transcriptome signature of irreversible senescence in human papillomavirus-positive cervical cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 100, 7093–7098
53 Bumke, M.A. *et al.* (2003) Modulation of gene expression by extracellular pH variations in human fibroblasts: a transcriptomic and proteomic study. *Proteomics* 3, 675–688