Differential gene expression technologies for identifying surrogate markers of drug efficacy and toxicity

Joseph A. Rininger, Vincent A. DiPippo and Bonnie E. Gould-Rothberg

Advances in the rapidly evolving discipline of pharmacogenomics have forced the biotechnology and pharmaceutical industries to integrate differential gene expression profiling into their drug discovery and development strategies. Here we highlight the use of differential gene expression technologies for the elucidation of both drug efficacy and toxicity as well as novel candidate genes for pharmacogenetic analyses to assess individual variability to drug response. This will include an overview of the different technologies created to facilitate pharmacogenomic analyses and to highlight advantages and disadvantages of these emerging methodologies. Two high-throughput differential gene expression technologies, microarrays and GeneCalling®, will be presented in detail.

The application of pharmacogenomics is rapidly becoming standard practice in both the preclinical and clinical phases of drug development. The underlying principle for the rapid development of pharmacogenomics is that drug efficacy and/or toxicity is linked to the differential expression of target genes[1]. Additionally, pharmacogenomics can be divided into two discrete disciplines: pharmacogenomics and pharmacogenetics. Pharmacogenomics can be defined as the quantification of differential gene expression (DGE) induced by a test agent in either an *in vitro* or an *in vivo* model. From such analysis, molecular mechanisms driving individual drug pharmacodynamics can be derived and yield a subset of gene-based markers correlative with and predictive of drug efficacy and/or toxicity. Pharmacogenomics represents a bridge between molecular biology and medicinal chemistry, allowing any small-molecule drug or biologic to be represented by its association with a unique set of differently modulated genes in key target tissues[2]. Successful implementation of pharmacogenomic studies is contingent not only on a robust technology platform but also on bioinformatics support that can assimilate and integrate the vast amounts of data produced by these studies.

Technology

As stated, DGE profiling forms the underpinnings of pharmacogenomics. Consequently, the validity of any pharmacogenomics analysis is reflective of the robustness of the DGE profiling technology used to generate the raw data. This can be objectively measured using the following parameters:

(1) resolution – how clearly does the technology differentiate any given gene from its nearest neighbor (both physically on the platform and virtually at the level of sequence identity);
(2) sensitivity – the limit of the *n*-fold change in expression that can be statistically detected by the given method;

Joseph A. Rininger, **Vincent A. DiPippo** and * **Bonnie E. Gould-Rothberg**, CuraGen Corporation, 555 Long Wharf Drive, New Haven, CT 06511, USA. *tel: +1 203 871 4219, fax: +1 203 481 1102, e-mail: bgould@curagen.com

 1359-6446/00/$ – see front matter PII:S1359-6446(00)01597-7

(3) coverage – percentage of expressed genes for a given species or tissue that are assayed reliably and reproducibly by the technology in successive iterations;
(4) false-positive rate – the amount of genes erroneously detected as differentially expressed by each platform; and
(5) false-negative rate – the number of genes truly differentially expressed that are not detected by the platform.

DGE technologies can be subdivided into two major groups: 'closed architecture' systems and 'open architecture' systems. Closed architecture, which requires *a priori* sequence knowledge of each gene or clone to be assayed and only measures DGE in a limited set of selected or known genes, is best represented by DNA microarrays and quantitative RT-PCR (TaqMan®, Hoffmann-La Roche, Basel, Switzerland) strategies[3–9]. Conversely, open architecture technologies do not assume prior sequence knowledge and can, theoretically, survey all transcripts in any selected tissue and identify novel genes in response to drug treatment or to fully characterize disease models[10–12]. As a result, open architecture systems are particularly suited for novel gene discovery and for generating primary pharmacogenomic profiles[13]. Examples of open architecture systems include:

(1) differential display[10];
(2) serial analysis of gene expression (SAGE)[14];
(3) representational differential analysis (RDA)[15];
(4) subtractive hybridization[6];
(5) GeneCalling®[11]; and
(6) the recently described total gene expression analysis (TOGA)[12].

Sensitivity and coverage can vary depending on the specific technology chosen with sensitivity ranging from 1:10,000 to 1:300,000 and coverage from 60% to 98% (Table 1)[10–12,14,15]. An additional benefit of open systems is the acquisition of transcript sequence information, therefore, splice variants and amino acid-altering mutations are identified readily.

Closed architecture methodology: DNA microarrays

DNA microarrays are designed for the simultaneous measurement of the expression of several thousand genes in a single hybridization procedure[16–18]. Microarrays are manufactured in a reproducible pattern of thousands of DNAs (primarily PCR products or oligonucleotides) attached to a solid support such as glass[19,20]. Fluorescently labeled DNA or RNA prepared from mRNA is then hybridized to their complementary DNA contained on the microarray and detected via laser scanning. Differences in labeling intensity are converted into a quantitative output of relative gene expression.

There are two principal types of microarrays: (1) spotted arrays that consist of pre-synthesized DNA affixed to glass slides and (2) high-density oligonucleotide arrays in which oligonucleotide sequences (~25 base pairs in length) are synthesized *in situ* on glass chip wafers using a photolithographic manufacturing process[3,4,21]. Both types of microarrays are available from commercial sources. However, spotted microarrays can be produced successfully in-house. Although these two different forms of microarrays both produce comprehensive expression profiles, there are significant differences between the two systems[16].

The DNA microarray technology platforms have two significant advantages: speed and sensitivity. Reported detection levels are 1:300,000 for low abundance mRNAs[4] (Table 1). The quantitative RT-PCR procedure, TaqMan®, has even greater sensitivity with transcripts successfully amplified from as little as 500 fg of starting RNA[8,9]. Therefore, TaqMan® represents a popular means of validating differentially expressed genes found by either closed or open architecture technologies.

The deficiencies of closed architecture platforms are the coverage of genes for a given experiment, the lot-to-lot variability of manufactured chips and the current deficiencies of bioinformatic support to successfully analyse generated data sets[1,16,17]. As new sequence information is added to the species' databases and novel genes are characterized, existing chips become outdated and new ones need to be constructed. Chip content cannot be updated in real-time. For the foreseeable future, there will be a discrepancy between the available coverage based on publicly (and privately) available database resources and that which is actually spotted on any given microarray. In the rat, for example, less than 30% of the genes spotted on genome-encompassing commercially available chips are fully characterized (based on GenBank accession numbers associated with the spotted genes), the remainder providing expressed sequence tag (EST) information[16]. Another technical limitation of DNA array and RT-PCR technology is the inability to distinguish between closely related members of the same gene family (e.g. cytochrome P450s, glucuronosyltransferases). Therefore, although physical resolution is not a significant concern, the ability to resolve polymorphisms, splice variants, insertions, deletions and so on, is significantly hampered when exclusively utilizing chip technology.

Open architecture methodology: GeneCalling®

GeneCalling® has been adapted specifically for high sample throughput. It is a quantitative open architecture DGE

Table 1. A comparison of representative open and closed DGE architecture systems

Platform	Differential gene expression technology	Resolution (variant gene detection)	Sensitivity[a]	Coverage[b] (%)	Clarifying comments	Refs
Closed						
	DNA microarray	No	1:300,000	Variable		3,4,16
	TaqMan®/RT-PCR	NA	<1:300,000	100		8,9
Open						
	Differential display	Yes	1:100,000	96	240 reactions	10,13
	SAGE	Yes	<1:10,000	92	300,000 tags or 10,000–15,000 sequencing reactions	13,14,46
	RDA	Yes	<1:300,000	NA		13,15
	GeneCalling®	Yes	1:125,000	95	96 restriction enzyme pair reactions	11,13
	TOGA	Yes	<1:100,000	60 to 98	60% coverage using one enzyme, 98% using four (4 × 256 reactions)	12,13

[a]Sensitivity, defined as *n* -fold difference detection limit, is directly correlated to the number of iterations of analysis performed.

[b]For many methodologies, the reported percentage coverage is not correlated to the abundance of mRNA. Coverage of rare transcripts could be significantly less.

Abbreviation: NA, no information available.

technology that assays transcript abundance by processing poly-A^+-mRNA-derived cDNA through restriction digestion utilizing optimized pairs of six-base-pair recognition restriction endonucleases (Fig. 1)[11]. Following restriction enzyme digestion, the resultant gene fragments are ligated and end-labeled with 5′-fluorescamine (FAM) and 3′-biotin adapters and are then subjected to PCR amplification. Subsequently, fragments specifically generated by double digestion are resolved by gel electrophoresis[11,22].

By combining the sequence knowledge derived from the two restriction enzymes with a differentially expressed fragment's length (in base pairs), this platform has the specific advantage of being able to predict or 'call' putative gene assignments for each fragment. This information is sufficient for database cross-referencing to determine the identity (or novelty) of the gene(s) for a particular species or tissue. To confirm the identity of each modulated gene fragment (and rule out false-positives), a subsequent gene-specific competitive PCR reaction, known as oligonucleotide poisoning, is performed[11]. The technology has demonstrated: resolution to 1:125,000 copies, the ability to detect *n*-fold changes as subtle as ± 1.5–2.0 fold and has shown equal sensitivity and coverage in a wide range of species including plants, fungi and mammals[11,23–27].

Differentially expressed fragments that do not yield any 'GeneCalls' (i.e. novel gene fragments) can be directly eluted from soft agarose gels and sequence isolated using standard cloning techniques. All resulting clones are assayed using oligonucleotide poisoning to definitively identify which clone co-migrates with the differentially expressed gene fragment. Pharmacogenomic studies of rat heart, brain and liver have been successful in uncovering not only previously undiscovered transcripts but also variant forms of previously identified genes (Ref. 28; B.E. Gould-Rothberg, unpublished).

The limitations of GeneCalling® and other open architecture technologies are that the analysis time is significantly longer than for microarrays (three to six months). However, it should be noted that the time component arises from the gene confirmation (poisoning) and isolation strategies of novel gene sequences. Further, although there is the potential of full transcriptome coverage, it is possible that the differentially expressed transcripts might not possess cleavage sites for the employed restriction enzymes. Therefore, certain differentially expressed genes would go undetected (false-negatives).

Bioinformatics support

Another major requirement for any pharmacogenomic strategy is a robust bioinformatics infrastructure that can collect and distill the massive datasets generated by both closed and open architecture studies, at the level of both workflow and data analysis. Bioinformatics support allows for visualization of binary comparisons and difference analysis. Higher-order analyses are executed through the use of 'array' (parallel visualization of multiple, related difference analyses), hierarchical clusterings and principle component analysis[16,17,29]. There are several software packages that have been designed for analysis of DGE-generated data including Spotfire Pro (Spotfire, Cambridge, MA, USA; http://www.spotfire.com), GeneCluster (http://wwwgenome.wi.mit.edu/mpr/software.html),

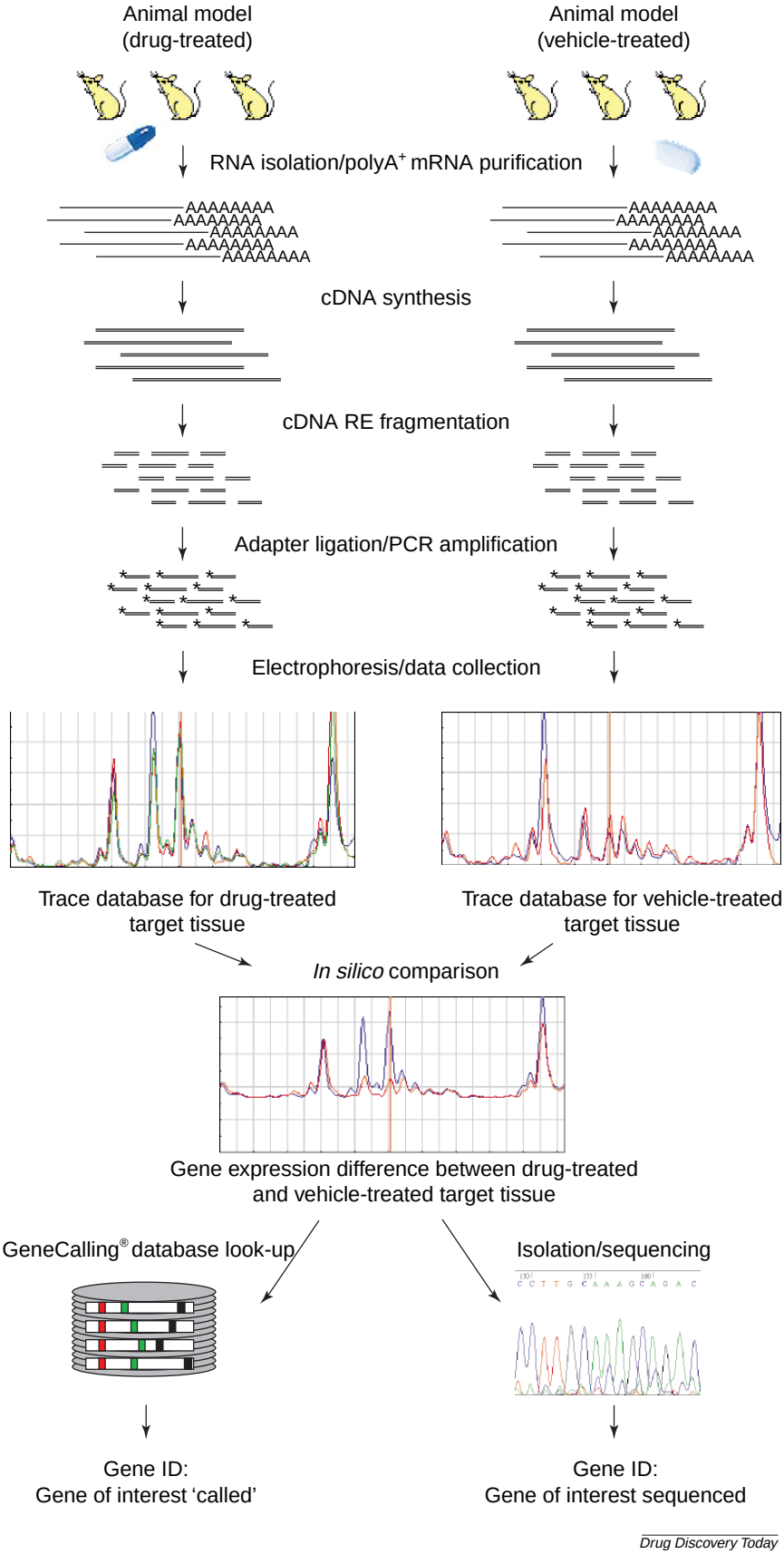

Figure 1. *Representation of the GeneCalling® process. Drug-treated or vehicle-treated rats were sacrificed and the target tissue removed. Poly-A+ mRNA is isolated and converted to cDNA. cDNA is digested by restriction endonuclease (RE) pairs, ligated to biotin and fluorescent-labeled adapters and processed through a limited number of PCR cycles. The fragments are resolved on thin gel electrophoresis and the resulting traces are entered into the GeneScape® database. Difference comparisons between the drug-treated and vehicle-treated traces are executed with differentially expressed peaks identified. Differentially expressed fragments are identified and confirmed by comparison against a virtual digest of the appropriate databases [GenBank; followed by oligonucleotide 'poisoning' (not shown)] and by isolation, cloning and sequencing.*

GenExplore (Applied Maths, Austin, TX, USA; http://www.applied-maths.com/home.html), GeneSpring™ (Silicon Genetics, Redwood City, CA, USA; http://www.sigenetics.com/index.html) and CuraTools™ (Curagen, Newhaven, CT, USA; www.curagen.com). These various DNA and protein sequence analysis tools facilitate characterization of novel gene fragments, coordination of electronic and wet-lab gene fragment 'contig' extensions, ortholog identification and classification of each gene in the gene ontology (GO) hierarchy[30]. The annotation of all genes is necessary in order for them to be appropriately 'binned' into categories related to specific metabolic or signaling pathways as well as enzymatic class. As DGE studies grow in use and acceptance, one of the greatest tasks confronting researchers will be the accurate characterization, annotation and organization of the data sets.

Experimental design and analysis strategy

The most common method of microarray data analysis consists of clustering for statistical relevance for either the up- or down-regulation of gene subsets[17,31,32]. However, the majority of published molecular studies using microarray technology are often conducted without replication[32,33]. A recent report evaluating the reproducibility of spotted microarrays found that output from any single microarray is subject to substantial variability with only one potential source of variation based on experimental design – spot location on the glass support[33]. These investigators reported that, in any single replicate experiment, the probability can be as large as 5% that the mRNA in the sample either fails to be represented as a probe or fails to be hybridized on the slide (false-negatives). In addition, a 10% probability of false-positives was determined[33]. In short, such experimental design (evaluation based on single hybridizations) resulted in the generation of different lists of expressed genes. The use of replicate samples and/or hybridizations should greatly reduce the variability and increase the statistical significance of findings generated via microarrays, especially in experiments containing multiple levels of complication adding to variation (animal-to-animal differences, percentage of responders, etc.)[33].

By contrast, GeneCalling® technology is typically performed with a minimum of three individual test samples, each run in triplicate, simultaneously[11,23,26,28]. This design provides greater confidence in the statistical relevance of the differentially expressed fragments generated and subsequently verified with TaqMan® analysis upon repetition of the experimental protocol[11,23]. However, binary comparisons are often insufficient to determine which sets of genes offer true predictive power of either pharmacological and/or toxicological significance[1,23]. Thus, it is necessary to compare several drugs or model systems that produce a desired physiological response with appropriate negative controls, and selectively triage the datasets for those genes that show statistical correlation with the desired phenotype[1].

Another frequently asked question is whether or not one can predict the unforeseen toxicity of a novel lead compound, the chemical structure of which has only minimal resemblance to previously approved drugs that act via the same mechanism. In this case, the experimental design is more complex. It is necessary to perform DGE on the novel lead compound, several representatives of each related chemical class, as well as representatives of each histopathological subtype in the selected organ, in addition to profiling a group of drugs devoid of toxicity. A detailed bioinformatic analysis of this massive dataset, including hierarchical clustering and principal component analysis is required to extract the limited sets of genes that define each histopathological subtype to the exclusion of all others and can then be used to evaluate the novel lead compound[1]. Subsequently, databases can be constructed for histopathological subtypes in each organ. Of course, follow-up experiments need to be devised to functionally validate the role of specific marker genes in the major categories of toxicological significance[1]. Because these individual experiments would not ignore novel genes, all genes including appropriate orthologs can be identified and contextualized. Furthermore, it is essential to perform time course and dose response studies to ensure that the DGE profiles are optimally representative.

DGE technologies applied to drug discovery and development

Pharmacogenomics offers the opportunity to obtain an alternative, often complementary, view to drug efficacy. In selected target tissues, drug efficacy can now be defined in terms of the selected metabolic and/or signal transduction pathways that are activated by drug administration. Benefits of this approach include the following:

(1) comprehensive characterization of drug efficacy for incompletely characterized drugs (e.g. lithium in the central nervous system);
(2) identification of clinical surrogate markers to monitor drug efficacy;
(3) indication of potential drug–drug interactions; and
(4) elucidation of novel indications.

Below are two examples for each DGE architecture that demonstrate the utility of DGE technologies in the drug development process.

Microarray technology using mammalian cells was employed in a study that profiled DGE responses induced

by serum in normal human fibroblasts that had been serum deprived for 48 h to initiate quiescence[32]. The microarray examined approximately 8600 distinct human genes (about half being unnamed ESTs) at 12 time points ranging from 15 min to 24 h post-stimulation. With the exception of the 8 h time point, only one replicate hybridization was performed per time point. Differential expression was determined based on a 2.2-fold change in expression in two or more time points or if the standard deviation measured for a specific gene during this time course (derived from the set of 13 arrays run with an expression ratio of $\log_2$) exceeded 0.7.

This experiment identified 517 genes whose expression changed substantially in response to serum. To manage this dataset, a clustering and display program was used to subdivide the genes into gene clusters based on the temporal expression pattern[17,32]. This strategy allowed for the categorization or predicted function of genes, including ESTs, that had little prior known biochemical characterization, based on a correlation with genes that have been characterized extensively possessing the same expression profile. This experiment found over 200 previously unknown genes, the expression of which was regulated in specific temporal patterns. For example, one cluster was enriched in genes known to mediate the G_2/M-phase transition of the cell cycle, suggesting a similar role for the ESTs grouped into that gene cluster.

A study that was more directed towards the elucidation of candidate genes for drug discovery and development using microarrays, focused on the evaluation of inflammatory gene expression patterns involved in rheumatoid arthritis and inflammatory diseases[34]. This work confirmed previous findings implicating certain cytokines in the development of rheumatoid arthritis. It also identified novel participants not previously reported, which included human matrix metallo-elastase (HME). The association of this known gene with a disease condition deems it a candidate drug target[34].

An example of using an open architecture technology (GeneCalling®) is the evaluation of genes regulated by obesity and leptin[23]. Since the discovery of leptin and its weight-loss effects, attempts to ascertain its mechanism of action have led to the discovery of central, direct and indirect peripheral effects[35–38]. In an attempt to identify genes relevant to the mechanism of action of leptin, GeneCalling® was performed on multiple target organs (pituitary, adipose, muscle, liver and hypothalamus) in both wild-type lean (C57Bl/6) and obese (ob/ob) mice treated with either leptin or vehicle control[23]. This study determined that up to 2% of the genes expressed within the tissues analysed were altered in response to obesity and that a one-week treatment with leptin returned only ~10% of these genes towards the normal state.

Among the gene expression changes in the pituitary, five differentially expressed genes were discovered (including one novel gene) that were altered by obesity and were responsive to leptin treatment. Subsequent biochemical validation of one of the known genes found to meet this criteria, proopiomelanocortin (POMC), revealed a novel regulatory loop between leptin and the pituitary hormone, adrenocorticotropic hormone (ACTH) – a cleavage product of POMC[23]. ACTH was shown to directly suppress leptin mRNA levels as well as leptin release in primary adipocytes[23]. Interestingly, this result implies that the regulation of leptin expression by ACTH can occur at multiple levels because one of the earliest findings regarding leptin was the ability of corticosteroids (which are directly regulated by ACTH) to induce leptin expression and reduce food intake and body weight[38]. This novel regulatory pathway could lead to additional clinical applications for leptin and identify additional drug targets situated within this feedback loop.

In another example of GeneCalling® applicability, the elucidation of small-molecule efficacy was demonstrated by the recent characterization of a novel peroxisome proliferator-activated receptor alpha (PPARα) ligand, GW9578, in rat liver[28]. Following oral dosing, GW9578 treatment resulted in 2.4% of the rat liver genes being differentially expressed. After confirming the sequence identity of 50 distinctly modulated genes, differentially expressed genes representative of multiple metabolic pathways were observed[28]. Consistent with previous literature reports associated with PPARα ligands, GW9578 induced an upregulation of genes involved in mitochondrial, peroxisomal and microsomal fatty acid oxidation[39]. In addition, pathways involved in xenobiotic detoxification and steroid modification were elucidated and suggested a potential novel indication for GW9578 in the treatment of X-linked adrenoleukodystrophy[28].

Toxicity

In addition to creating DGE profiles to address issues of efficacy, a second utility of pharmacogenomics is to supplement current preclinical and Phase I clinical protocols to identify and eliminate drug candidates that might elicit downstream toxicity from further development. Two discrete advantages can be gleaned from this approach.

The first benefit is most broadly applicable to all compounds entering preclinical evaluation. The skyrocketing costs of drug development have motivated the pharmaceutical industry to identify strategies that can streamline both the length and expense of this process. Assessment

of a compound's preclinical toxicity is one area that is amenable to intervention and to amelioration of the hurdles of time and cost. Pharmacogenomics, through the identification and development of appropriate surrogate markers, can improve the predictive accuracy of animal models to humans[1]. In this manner, compounds can be dosed in animal models and subsequently triaged using the appropriate set of gene expression markers in a much-abbreviated time frame[1]. This process will potentially not only improve the efficiency of lead optimization cycles and offer cost reduction, but will also ultimately allow a compound more marketing time before patent expiration to recuperate the costs of drug development.

The second benefit is directed at late-stage clinical and Phase IV aftermarket drug trials. Since 1997, a significant number of marketed compounds were withdrawn from the market owing to unanticipated aftermarket toxicity (http://www.fda.gov/opacom/archives.html). Rezulin™ (troglitazone), Posicor® (mibefradil) and Tasmar® (tolcapone) are just three examples of compounds that represented new drug classes and were expected to be extremely successful. In each case, the withdrawn drug was associated with a highly morbid toxicity occurring in a very narrow segment of the population. Only 20 of the 2,510 patients receiving troglitazone during its Phase III clinical trial discontinued therapy due to abnormal liver tests and none of these patients suffered irreparable hepatic injury[40]. When troglitazone was introduced to a larger patient population, 35 of 600,000 patients suffered liver damage (including one death and one transplant) in the first six months of the USA aftermarket and the FDA required a warning be issued with the drug label[41]. Troglitazone was finally withdrawn in March 2000 after it became apparent that similar toxicities were not seen in patients treated with competitor drugs rosiglitazone and pioglitazone[42]. In another example, mibefradil was withdrawn from the market only three months after its launch, following the pre-publication of data from the 2590 patient Mortality Assessment in Congestive Heart Failure (MACH-1) trial[43]. In this instance, the evidence suggested that cytochrome P4503A4 activity was dysregulated by mibefradil, resulting in drug–drug interactions with other commonly prescribed drugs administered to heart failure patients and increased mortality[43,44].

The most compelling observation regarding each of these failures is that the affected population represented less than 1% of all patients prescribed this therapeutic, whereas over 99% of the remaining patients had uneventful courses and were benefiting from the anticipated therapeutic effects. For other compounds like codeine and 6-mercaptopurine, it has been shown that individual variation in drug response is related to single nucleotide polymorphisms (SNPs) in focal genes regulating drug pharmacokinetics or pharmacodynamics (CYP2D6 for codeine and thiopurine methyltransferase for 6-mercaptopurine)[45]. In fact, genetic variations found in over 100 pharmacologically relevant genes have been definitively correlated with variability in drug response across individuals of a given population[45]. Yet, there are many idiosyncratic drug side-effects whose associative gene and causative polymorphisms have not been identified. The task of identifying SNPs within candidate drug-regulated genes and conducting population-based assays to identify those SNPs correlative with individual variability in drug response defines pharmacogenetics, the second branch of pharmacogenomics.

Pharmacogenetics might play a significant role in the case of cardiotoxicity associated with co-administration of Phentermine and Fenfluramine, 'Phen-fen'. The combined use of these drugs resulted in the development of valvular fibrosis prompting the removal of Phen-fen from the marketplace. This adverse effect occurred in a minority subset of otherwise healthy obese women, suggesting that one of the predisposing factors is most likely a polymorphism in an unidentified gene. By identifying the causative polymorphism(s), the pharmaceutical industry would then gain additional information to effectively triage individual obese patients, should novel, more effective fenfluramine-like compounds be developed. One or more of the obvious candidate genes (e.g. serotonin-receptor subunits, serotonin-reuptake transporters, genes involved in fenfluramine metabolism) can be screened *a priori* for functionally active polymorphisms displaying a statistically significant correlation among the affected population. An alternative strategy must also be available to expand the selection of marker genes to include a reasonable subset of rational candidates that can be screened in a cost-effective manner. Our laboratories have implemented an expression pharmacogenomics strategy coupled to identifying additional pharmacogenetic candidates for analysis in unanticipated drug toxicities, including the aforementioned Phen-fen example. The purpose is to identify the human orthologs of dexfenfluramine/fenfluramine-responsive rodent genes that might have the potential to serve as additional candidate genes for the pharmacogenetic etiology of fenfluramine-induced idiosyncratic valvulopathy.

Conclusions

Pharmacogenomics is a rapidly growing and evolving field that has developed beyond the search for genetic links associated with individual variability with respect to drug pharmacokinetics and pharmacodynamics. Advancing technology has made it possible to redefine pharmacogenomics as DGE profiling and to integrate it into drug discovery

and development programs. However, progress still needs to be made with DGE technology to account for interassay variance and to categorize the degree of false-positive and -negatives, topics that are not well documented (and possibly not truly known). Each technology platform has its advantages and disadvantages. Microarrays are faster but lack transcriptome coverage and with the continual updating that is necessary with microarrays, they will require subsequent experimental replication and data analysis to incorporate new sequence information. Open architecture methodologies can be viewed as more time consuming initially. However, the advantage of comprehensive transcriptome coverage overcomes this deficit and makes subsequent re-evaluation of samples from a given experiment unnecessary. With the ability to comprehensively survey entire genomes, there is little question that the future of drug discovery and development will be greatly enhanced by the coordinate application of pharmacogenomic strategies to identify predictive markers of drug efficacy and toxicity.

Acknowledgements

The authors would like to thank Kellye K. Daniels, Robert Gerwien, Cynthia Green, David Lewin and Pascal Bouffard for their advice and critical comments in the preparation of this manuscript.

REFERENCES

1 Farr, S. *et al.* (1999) Concise review: gene expression applied to toxicology. *Toxicol. Sci.* 50, 1–9

2 Bailey, D.S. *et al.* (1998) Pharmacogenomics – it's not just pharmacogenetics. *Curr. Opin. Biotechnol.* 9, 595–601

3 Schena, M. *et al.* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470

4 Lockhart, D.J. *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* 14, 1675–1680

5 Scherf, U. *et al.* (2000) A gene expression database for the molecular pharmacology of cancer. *Nat. Genet.* 24, 236–244

6 Lee, P.S. *et al.* (2000) Genomic analysis. *Curr. Opin. Biotechnol.* 11, 171–175

7 Shi, M.M. *et al.* (1999) Technologies for detecting genetic polymorphisms in pharmacogenomics. *Mol. Diagn.* 4, 343–351

8 Roberts, C.A. *et al.* (2000) Real-time RT-PCR fluorescent detection of tomato spotted wilt virus. *J. Virol. Methods* 88, 1–8

9 Heid, C.A. *et al.* (1996) Real-time quantitative PCR. *Genome Res.* 6, 986–994

10 Liang, P. *et al.* (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257, 967–971

11 Shimkets, R.A. *et al.* (1999) Gene expression analysis by transcript profiling coupled to a gene database query. *Nat. Biotechnol.* 17, 798–803

12 Sutcliffe, J.G. *et al.* (2000) TOGA: an automated parsing technology for analysing expression of nearly all genes. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1976–1981

13 Green, C.D. *et al.* Open systems: panoramic views of gene expression. *J. Immunologic. Meth.* (in press)

14 Velculescu, V.E. *et al.* (1995) Serial analysis of gene expression. *Science* 270, 484–487

15 Hubank, M. *et al.* (1994) Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res.* 22, 5640–5648

16 Harrington, C.A. *et al.* (2000) Monitoring gene expression using DNA microarrays. *Curr. Opin. Microbiol.* 3, 285–291

17 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

18 Wen, X. *et al.* (1998) Large-scale temporal gene expression mapping of central nervous system development. *Proc. Natl. Acad. Sci. U. S. A.* 95, 334–339

19 Lipshutz, R.J. *et al.* (1999) High-density synthetic nucleotide arrays. *Nat. Genet.* (Suppl.) 21, 20–24

20 Bowtell, D.D.L. (1999) Options available-from start to finish-for obtaining expression data by microarray. *Suppl. Nat. Genet.* 21, 25–32

21 McGall, G. *et al.* (1996) Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13555–13560

22 Simpson, J.W. *et al.* (2000) A transmission imaging spectrograph and microfabricated channel system for DNA analysis. *Electrophoresis* 21, 135–149

23 Renz, M. *et al.* (2000) Quantitative expression analysis of genes regulated by both obesity and leptin reveals a regulatory loop between leptin and pituitary-derived ACTH. *J. Biol. Chem.* 275, 10429–10436

24 Ohneda, O. *et al.* (2000) WECHE: a novel hematopoietic regulatory factor. *Immunity* 12, 141–150

25 Bruce, W. *et al.* (2000) Expression profiling of the maize flavonoid pathway genes controlled by estradiol-inducible transcription factors CRC and P. *Plant Cell* 12, 65–79

26 Kahn, J. *et al.* (2000) Gene expression profiling in an *in vitro* model of angiogenesis. *Am. J. Pathol.* 156, 1887–1900

27 Goodwin, B. *et al.* (2000) A regulatory cascade of the nuclear receptors FXR, SHP-1, and LRH-1 represses bile acid biosynthesis. *Mol. Cell* 6, 517–526

28 Rothberg, B.E.G. *et al.* The characterization of PPARα ligand drug action in an *in vivo* model by comprehensive differential gene expression profiling. *Funct. Integrat. Genomics* (in press)

29 Duggan, D.J. *et al.* (1999) Expression profiling using cDNA microarrays. *Nat. Genet.* (Suppl.) 21, 10–14

30 Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29

31 Alon, U. *et al.* (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. U. S. A.* 96, 6745–6750

32 Lyer, V.R. *et al.* (1999) The transcriptional program in the response of human fibroblasts to serum. *Science* 283, 83–87

33 Lee, M-L.T. *et al.* (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc. Natl. Acad. Sci. U. S. A.* 97, 9834–9839

34 Heller, R.A. *et al.* (1997) Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl. Acad. Sci. U. S. A.* 94, 2150–2155

35 Halaas, J.L. *et al.* (1995) Weight-reducing effects of the plasma protein encodd by the obese gene. *Science* 269, 543–546

36 Elmquist, J.K. *et al.* (1998) Unraveling the central nervous system pathways underlying responses to leptin. *Nat. Neurosci.* 1, 445–450

37 Marti, A. *et al.* (1999) Leptin: physiological actions. *J. Physiol. Biochem.* 55, 43–49

38 Houseknecht, K.L. *et al.* (1998) The biology of leptin: a review. *J. Anim. Sci.* 76, 1405–1420

39 Desvergne, B. and Wahli, W. (1999) Peroxisome proliferator-activated receptors: nuclear control of metabolism. *Endocr. Rev.* 20, 649–688

40 Watkins, P.B. and Whitcomb, R.W. (1998) Hepatic dysfunction associated with troglitazone. *New Engl. J. Med.* 338, 916–917

41 FDA (1999) New labeling and use changes for rezulin. *FDA talk paper*. June 16

42 Miller, J.L. (2000) Troglitazone withdrawn from market. *Am. J. Health Syst. Pharm.* 57, 834

43 Levine, T.B. *et al.* (2000) Effect of Mibefradil, a T-type calcium channel blocker, on morbidity and mortality in moderate to severe congestive heart failure: the MACH-1 study. Mortality Assessment in Congestive Heart failure trial. *Circulation* 101, 758–764

44 Mahon, N. and McKenna, W.J. (1998) Calcium channel blockers in cardiac failure. *Prog. Cardiovasc. Dis.* 41, 191–206

45 Evans, W.E. and Relling, M.V. (1999) Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 286, 487–491

46 Bertelsen, A.H. *et al.* (1998) High-throughput gene expression analysis using SAGE. *Drug Discov. Today* 3, 152–159