State-of-the-Art Lecture

Genomics and Hypertension
Concepts, Potentials, and Opportunities

Richard E. Pratt, Victor J. Dzau

Abstract—We are at the beginning of a biological revolution, spurred on by the Human Genome Project and associated studies. Within the next few years, expressed sequence tags (ESTs) representing all sequences expressed in humans will be determined and their genomic positions will be defined (STSs). The discovery of all the variants in the human genome that contribute to the genetic diversity of the human population will result in the construction of dense polymorphic maps. The rapid growth of the EST, STS, and single-nucleotide polymorphism (SNP) databases, coupled with impressive technological advances, will surely have a dramatic effect on biomedical research. In this review, we will examine the recent advances in genetics and genomics and place these within the context of medical research and patient care, with an emphasis on studies in the cardiovascular system. (Hypertension. 1999;33[part II]:238-247.)

Key Words: Human Genome Project ■ gene expression ■ genes ■ polymorphism, single-nucleotide ■ genetics ■ pharmacogenomics ■ proteomics

The elucidation of the genetic basis of disease has been a goal of medicine for many decades. With the advent of molecular genetic technologies, over the past 10 years, nearly 100 genes causing various genetic diseases have been identified by positional cloning. However, these diseases are for the most part monogenic in nature (examples include cystic fibrosis, Huntington’s disease, Alzheimer’s disease, and familial cardiomyopathy); studies of complex, polygenic diseases, such as hypertension and atherosclerosis have proven much more difficult. Genes responsible for rare monogenic forms of these diseases, such as Liddle’s syndrome, Barter’s syndrome, and glucocorticoid-remediable aldosteronism (for review, see reference 1), have been identified and may provide clues to genes contributing to more common forms of disease.2 Linkage and association studies using candidate genes have also provided insight into the causes of diseases such as hypertension and atherosclerosis. However, the number of genes that have proven consistently positive in multiple studies is small and the predictive power of these genetic variants is, unfortunately, limited.

Nevertheless, it is now possible in a few specific diseases to stratify the risk and prognosis of individual patients on the basis of the identification of specific genetic variants. An example of the promise of genomic medicine is the identification of specific mutations of the β-MHC (Arg403Gln) have a very poor prognosis with early mortality due to sudden cardiac death. Other variations within β-MHC as well as variations in myosin binding protein C and cardiac troponin T are of similar predictive value (Table 1) and were recently reviewed.4,5

The predictive power of these genetic variants with respect to familial hypertrophic cardiomyopathy is possible because these particular variants lead to monogenic mendelian disorders. Results suggest that these variants are responsible for a large proportion of the cases of familial hypertrophic cardiomyopathy; therefore, genotyping of patients with cardiac hypertrophy for these variants is highly informative. Unfortunately, in more complex, polygenic diseases the current state of the technology does not lend itself to such easy and simple analysis. What then is the future of genetic analysis with respect to these complex diseases? Will it be possible in the future to use genetic analysis to define patients at risk for disease (in much the same way that current risk factors such as diet, smoking, cholesterol levels and family history are examined)? Will genetic analysis allow the physician to accurately define the exact prognosis of an individual with a common disease such as hypertension and/or lead to the selection of specific therapy based on specific genetic variation?

Mapping the Human Genome
An important byproduct of the Human Genome Project has been the definition of human diversity at the genomic level. Current estimates of the degree of diversity range from 1 in 500 to 1 in a 1000,6 resulting in millions of variants in the

Received September 22, 1998; first decision October 14, 1998; revision accepted November 6, 1998.
From the Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, Mass.
Correspondence to Richard E. Pratt, PhD and Victor J. Dzau, MD, Laboratory of Genetic Physiology, Brigham and Women’s Hospital, 75 Francis St, Boston, MA 02115. E-mail rpratt@bustoff,bwh.harvard.edu
© 1999 American Heart Association, Inc.
Hypertension is available at http://www.hypertensionaha.org

238
human genome (whose size is estimated at $3 \times 10^9$). The accumulated effect of these variants collectively defines, at the genomic level, the diversity of the human population. An understanding of the genetic diversity and how this diversity contributes to variations in normal and abnormal physiology will have a powerful effect on medicine in the coming decades. Someday it will be technically feasible to genotype and analyze these variants on an individual basis. Currently, an estimated an order of magnitude greater than the estimates for the total number of human genes. This redundancy allows for sequence comparisons and the definition of consensus sequences. This approach, used by The Institute for Genomic Research (TIGR, www.tigr.org) has led to the identification of over 72,333 (as of the March 18, 1998, release) tentative human consensus (THC) sequences based on overlapping ESTs. These overlapping ESTs account for approximately 600,000 EST sequences out of 770,000 in the TIGR database; the remaining 160,000 EST sequences are singletons, represented only once in the database.

The TIGR approach leads to an overestimate, because 1 transcript can yield $>1$ nonoverlapping EST sequence. Another, more conservative approach has been proposed (UniGene, www.ncbi.nlm.nih.gov/Schuler/UniGene), involving comparison of the ESTs to known gene or cDNA sequences to define clusters of sequences. For genes not clustered by this approach, a second step, involving comparisons of the 3' untranslated regions of the cDNA, was undertaken on the basis that the 3' UTR had the greatest variation between similar but not identical genes. On the basis of this approach, 47,956 unique clusters (as of September 1, 1998) have been defined.

How complete is the list of unique expressed human genes? The number of genes in the human genome has been estimated to be between 50,000 and 150,000. The accumulation of unique sequences (defined by UniGene) has dramatically decreased, after a near-vertical rise in 1995, when sequences accumulated at a rate of 1500 ESTs per day. Sequencing of standard cDNA libraries would be expected to yield multiple sequences for the highly abundant clones, whereas the genes expressed at low copy numbers will be underrepresented. Moreover, many genes exhibit tissue or temporal specific patterns of expression and/or may be abundantly expressed only in certain disease states. Thus, efforts are now directed toward normalized or subtracted libraries from normal and diseased tissues and from fetal libraries. A hint to the completeness of the EST database is provided by comparing these databases to independently isolated genes. As an example, of the 91 genes identified by positional cloning, 83 (91%) are represented in the dbEST database (www.ncbi.nlm.nih.gov/dbEST/dbEST.html). Similarly, of the 94 genes cloned as oncogenes or tumor suppressors, 94% are represented in the EST database (www.ncbi.nlm.nih.gov/dbEST/CancerGene.html).

The definition of the expressed sequences in the human genome has been useful in terms of defining gene families, discovering novel proteins, and determining patterns of expression in different tissues and disease states. However, the full value of this database would not be realized if the position of these sequences on the genome was not also defined. Thus, to extract the full potential of these sequences, an effort to map these sequences onto the human genome was initiated. Primarily on the basis of radiation hybrid techniques for mapping, > 30,000 genes (combination of ESTs and known genes) have been mapped, (www.ncbi.nlm.nih.gov/genemap98).

The knowledge of the location of genes in the genome will greatly enhance the ability of investigators to identify disease genes. Genetic analysis results in the definition of genomic regions linked to disease. With the accumulation of mapped

### TABLE 1. Predictive Variants in Hypertrophic Cardiomyopathy

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutation</th>
<th>Prognosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-Myosin heavy chain</td>
<td>Gly256Glu</td>
<td>Benign</td>
</tr>
<tr>
<td></td>
<td>Arg403Trp</td>
<td>Benign</td>
</tr>
<tr>
<td></td>
<td>Phe513Cys</td>
<td>Benign</td>
</tr>
<tr>
<td></td>
<td>Val606Met</td>
<td>Benign</td>
</tr>
<tr>
<td></td>
<td>Leu900Val</td>
<td>Benign</td>
</tr>
<tr>
<td></td>
<td>Arg403Gln</td>
<td>Increased SCD</td>
</tr>
<tr>
<td></td>
<td>Arg403Leu</td>
<td>Increased SCD</td>
</tr>
<tr>
<td></td>
<td>Arg453Cys</td>
<td>Increased SCD</td>
</tr>
<tr>
<td></td>
<td>Gly716Arg</td>
<td>Increased SCD</td>
</tr>
<tr>
<td></td>
<td>Arg719Trp</td>
<td>Increased SCD</td>
</tr>
<tr>
<td>Cardiac troponin T</td>
<td>Arg92Gln</td>
<td>Mild LVH but increased SCD</td>
</tr>
<tr>
<td>Myosin binding protein-C</td>
<td>Multiple</td>
<td>Mild LVH but benign</td>
</tr>
<tr>
<td>α-Tropomyosin</td>
<td>Asp175Asn</td>
<td>Benign</td>
</tr>
</tbody>
</table>

The knowledge of the location of genes in the genome will greatly enhance the ability of investigators to identify disease genes. Genetic analysis results in the definition of genomic regions linked to disease. With the accumulation of mapped
genes, the likelihood that these regions will contain candidate genes is growing. Indeed, the likelihood of the disease gene of interest being listed among these candidate ESTs in a chromosomal region is rapidly escalating. However, it should be pointed out that currently the majority of these mapped ESTs encode proteins of unknown function. Perhaps <10% of the ESTs correspond to known genes, another 20% are orphans (sequences of homology to known genes), and the remaining 70% are unknown genes with no homology and with no known function.10 Nevertheless, the ability to identify genes (even unknown genes) residing within an interval linked to a particular disease, especially if the tissue patterns of expression of these genes are known or can be determined, will dramatically increase the power of genetic analysis and more rapidly yield candidate genes for further analysis.

The mapping of STSs and the infrastructure for rapid sequence analysis has enabled the determination of single nucleotide polymorphisms within a subset of these STSs. In a recently published report, Wang and colleagues demonstrated the potential of this approach. In this study, over 16,000 STSs from 7 individuals were sequenced using standard gel-based sequencing (for a small subset) and a newer technique, chip-based sequencing.11,12 More than 3000 SNPs were identified using these approaches and >2000 of these have been mapped on the human genome, at an average spacing of 2 cM (www.genome.wi.mit.edu/SNP/human/index.html).

Application of Genomic Sciences to Clinical Medicine

Prediction and Prognosis of Disease

The availability of such densely spaced (and easily assayed) polymorphisms should greatly facilitate the identification of genetic variants that contribute to the development of disease and may change the future design of genetic studies. For example, it has been calculated that for linkage studies, a collection of 1000 biallelic SNPs would have the same power as the commonly used 400 multiallelic SSLPs. Moreover, Risch and Merikangas have calculated that in complex polygenic diseases, the limited power of linkage analysis necessitates extremely large sample sizes. On the other hand, association studies with candidate genes are more powerful and require fewer subjects. In the past, association studies were usually performed with genetic variants found within candidate genes that were chosen on the basis of an understanding of physiology or pathophysiology; hence, variants within the genes encoding renin, angiotensinogen, NO synthase, etc., were tested for association with hypertension in case-control studies. Obviously, this approach necessitates an understanding of the gene under investigation; the protein is first discovered, the cDNA is cloned (followed by the cloning of the gene), and variants within the gene are determined. This approach is highly limited, because the vast majority of genes sequenced and mapped will encode proteins of unknown function. If one were to use the past paradigm, it will take many years for the function of these proteins to be determined, and the full power of the Human Genome Project will never be utilized. Fortunately, one need not be restricted to genes chosen on the basis of previous understanding of physiology; association between variants in anonymous genes can also be tested in case control studies. Thus, as the identification of SNPs within gene coding regions progresses, so too will the ability to perform simultaneous whole-genome association studies. One must be careful in the design of such studies, because there are certain caveats to association studies verses linkage analysis. Nevertheless, in the near future, we may see an increase in the development of SNPs both in candidate gene regions as well as throughout the genome. Indeed, request for applications from the National Institutes of Health have been issued with this precise goal. On the other hand, several recent reports have cautioned that these approaches may have limitations, and discussions of their uses may be oversimplified. Nevertheless, the use of these SNPs in the search for genes causal to the development of disease will probably increase dramatically.

The analysis of genetic variations in the prediction, diagnosis, and prognosis of disease has been used in the analysis of several monogenic diseases, such as retinoblastoma, cystic fibrosis, and breast cancer. These examples not only point out the potential of genetic testing but are also illustrative of the pitfalls of this approach. For example, variants of two genes, BRCA1 and BRCA2, account for approximately 70% of the cases of “large family” breast cancer (families with more than 4 affected individuals). Thus, it would appear that genotyping these 2 genes would be a health benefit. However, these genes are highly complex (22 and 27 exons, respectively spanning up to 70 kb). Moreover, causal mutations are spread throughout the genes, making the genotyping of these genes non-trivial using standard technologies. Similar problems exist in genes for cystic fibrosis and retinoblastoma as well as in many genes encoding other tumor suppressors. Thus, even for the monogenic diseases, the analysis of genetic variants can be a tedious and expensive proposition.

As stated in the introduction, the prognosis of familial hypertrophic cardiomyopathy can be determined with a high degree of accuracy through the genotyping of a handful of genes. However, this statement is an underrepresentation of the complexity of this disease. While mutants in β-myosin may account for 20% to 30% of the cases (with Arg403Gln being the most common mutation), β-myosin, like the genes described above, is large and complex (24 kb, 40 exons), with as many as 40 variants having been described, which cosegregate with familial hypertrophic cardiomyopathy. While variants in the other genes do not yet appear as numerous, multiple variants have also been described. Thus, as noted above, the analysis of genetic variants by current technology can be time-consuming and expensive.

As described above for SNP analysis, recent technical advances may increase the ability to assess genetic variants. Using DNA microarrays, Hacia and colleagues have demonstrated proof-of-concept using a single-chip assay for multiple polymorphic sites in BRCA1, exon 11. Thus, in a single assay, multiple variants can accurately be assayed. As described by Wang and colleagues, as many as 550 loci can be examined on a single chip. Thus, a chip containing all the known variants for virtually any gene or set of genes can be produced to assess quickly the risk for a particular disease.
Indeed, chips are commercially available on the web (www.affymetrix.com) to provide the sequence of the protease and reverse transcriptase genes for HIV, to sequence exons 2 to 11 of p53, and to assess 18 known mutations of the human CYP2D6 and CYP2C19 genes encoding the cytochrome P-450 enzymes.

Can these approaches be used for complex, polygenic diseases such as hypertension? In theory, if all the causal genes for hypertension were identified by linkage and association studies and the actual causal variants were determined, similar assessment of risk may be possible. Several different genes have been suggested to be involved but in few, if any, cases (other than the rare, monogenic forms) have the actual causal variant been identified. Clearly, considerable work remains in this area.

**Prediction of Therapeutic Outcomes: Pharmacogenetics/Pharmacogenomics**

Recent advances in the ability to rapidly genotype large numbers of genetic variants has spawned a new discipline, pharmacogenomics: the influence of genomic variation toward the individual patients response to therapy. It has long been known that individual variation exists in response to drug therapy due to several factors, such as drug uptake, activation, metabolism, and excretion. Moreover, variations in protein structure of drug targets can also greatly influence response. Adding to this complexity is the case of polygenic diseases, in which multiple causal genes are contributing to the development of disease. For example, one can envision a population of hypertensives in whom drug X will statistically exhibit differential potencies toward the adrenergic receptors. While both isomers are equally potent at the α1 receptor, only the S(-) isomer can antagonize the β-receptors. The different isomers also exhibit vastly different metabolic profiles, the S(-) isomer is metabolized by several enzymes while the metabolism of R(+) is dependent on CYP2D6 activity. Thus, subjects with genetic variants of CYP2D6 would exhibit different metabolic rates of the isomers and would exhibit different circulating levels of the 2 isomers, which would result in altered ratios of α1/β-blockade.

Clearly, drug metabolism is only 1 aspect. Several examples of genetic variants are shown in Table 2. The majority of studies have focused on 3 areas: hepatitis C, HIV, and Alzheimer’s disease. The case of Alzheimer’s disease is particularly interesting. Apolipoprotein E gene variants appear to be predisposing factors for the disease. One variant, ε4, is associated with a decreased response, especially in women, to tacrine, a cholinesterase inhibitor and the first approved therapy for Alzheimer’s. Moreover, a second experimental drug, S12024, worked better in Alzheimer patients with ε4 variant. All these observations require further investigation; nevertheless, these results are intriguing in light of the morbidity and mortality associated with these 3 diseases.

A recent study in atherosclerosis has also demonstrated the predictive power of pharmacogenomics. Cholesterol ester transfer protein (CETP) catalyzes the transfer of cholesterol ester from HDL to VLDL and LDL. A variant of CETP (presence of a Taq1B restriction site in the first intron, B1, absence called B2) has been described that results in a decreased plasma HDL, increased (VLDL+LDL)/HDL ratio and increased rate of progression of atherosclerosis. Interestingly, the B1 genotype is associated with a better lipid profile after dietary intervention and a dramatic response to pravastatin.

**Table 2: Examples of Pharmacogenomics: Examples in Which a Marker Gene (Not Necessarily the Target Gene) May Be Used to Predict Physiological or Drug Response**

<table>
<thead>
<tr>
<th>Disease</th>
<th>Response To</th>
<th>Predictive Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer’s</td>
<td>Tacrine, cholinesterase inhibitor</td>
<td>ApoE</td>
</tr>
<tr>
<td>Alzheimer’s</td>
<td>S12024</td>
<td>ApoE</td>
</tr>
<tr>
<td>Hypertension (rat)</td>
<td>Calcium channel blocker</td>
<td>Loci on chromosome 2 (calmodulin dependent protein kinase II Δ?)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>Diuretics</td>
<td>Adducin</td>
</tr>
<tr>
<td>Hypertension</td>
<td>Sodium restriction</td>
<td>Adducin</td>
</tr>
<tr>
<td>Asthma</td>
<td>Albuterol, β-receptor agonist</td>
<td>β-Receptor</td>
</tr>
<tr>
<td>AIDS</td>
<td>Protease inhibitors</td>
<td>Protease</td>
</tr>
<tr>
<td>AIDS</td>
<td>Reverse-transcriptase inhibitors</td>
<td>Reverse-transcriptase</td>
</tr>
<tr>
<td>Hepatitis C</td>
<td>Interferon-α</td>
<td>Multiple</td>
</tr>
</tbody>
</table>

for CYP2D6 and CYP2C19 in use. Because there have been several recent reviews addressing the importance of these proteins toward drug response, we will not expand on this subject except to present one interesting example (recently reviewed). Carvedilol is a mixed, nonselective β-adrenergic antagonist/β1-adrenergic antagonist approved for use in heart failure. Carvedilol exists as a mixture of S(-) and R(+) isomers that exhibit differential potencies toward the adrenergic receptors. While both isomers are equally potent at the α1 receptor, only the S(-) isomer can antagonize the β-receptors. The different isomers also exhibit vastly different metabolic profiles, the S(-) isomer is metabolized by several enzymes while the metabolism of R(+) is dependent on CYP2D6 activity. Thus, subjects with genetic variants of CYP2D6 would exhibit different metabolic rates of the isomers and would exhibit different circulating levels of the 2 isomers, which would result in altered ratios of α1/β-blockade.
Examples of pharmacogenomics can also be seen in hypertension. For example, in rat crosses, several loci have been shown to be linked to sodium sensitivity. One gene, \textit{ADD1}, encodes adducin, a heterodimeric cytoskeletal protein found in renal tubules that is thought to be involved in the regulation of ion transport. Variants in the gene encoding \textit{α}-adducin have been shown to be linked and associated to hypertension in both animal models and in some,\textsuperscript{36–38} but not all, human ethnic groups.\textsuperscript{39} Moreover, \textit{ADD1} variants are also linked to sodium sensitivity and response to diuretics.\textsuperscript{36–38} In rats, a locus on chromosome 2 has been identified that may mediate the antihypertensive (both diastolic and systolic) response to a dihydropyridine calcium antagonist, PY108-068 in a cross between Lyon hypertensive and Lyon normotensive. Of note, a candidate gene involved in calcium homeostasis, calmodulin-dependent protein kinase II, is located in the region.\textsuperscript{40}

On the other hand, there are other examples in hypertension that underscore the difficulties of these studies. The renin-angiotensin system is intimately involved in the regulation of blood pressure, and the regulation of this system is tightly coupled to sodium intake. Thus it is reasonable to hypothesize that alterations in this system may contribute to sodium-sensitive hypertension. Recently, 2 studies in the same ethnic group addressed this hypothesis. Curiously, in 1 study, the deletion/insertion polymorphism in the ACE gene was shown to be associated with salt sensitivity,\textsuperscript{41} whereas in the other it was not.\textsuperscript{42} Both studies also examined haptoglobin phenotype. Interestingly, the study that demonstrated an association with ACE showed no association with haptoglobin, whereas the study that showed no association with ACE was able to demonstrate association with haptoglobin. Similar contradictory results have been observed in treatment studies. In several different reports, subjects were analyzed for genetic variants in ACE and/or angiotensinogen.\textsuperscript{43–47} Patients were treated with ACE inhibitors; clinical end points were decreases in blood pressure or decreases in left ventricular hypertrophy (LVH). None of the studies demonstrated an association between ACE genetic variants and decreases in blood pressure. In 1 study, angiotensinogen variants were associated with decreases in pressure,\textsuperscript{45} whereas in another, the variants were not.\textsuperscript{44} Studies with LVH were no more consistent; 1 study demonstrated an association between \textit{ACE} gene variants and ACE inhibitor–dependent decrease in LVH,\textsuperscript{47} whereas in the other, no association was demonstrated.\textsuperscript{43} The protocols for the determination of genetic and molecular variants of ACE, angiotensinogen and haptoglobin, are straightforward, which emphasizes the need for careful documentation of the ascertainment criteria and for the phenotypic analysis in these studies.

The vast majority of pharmacogenomic studies to date have, understandably, focused on candidate genes: those chosen because they are either the target of the drug in question or are intimately involved in the pathway being targeted. However, the advances in ability to genotype using a dense panel of SNPs will dramatically alter pharmacogenomic studies. Indeed, the power of using a whole-genome search approach was demonstrated in the rat study cited above,\textsuperscript{40} examining the response to calcium antagonist. The same rationale concerning the use of a whole-genome candidate approach in case-control studies can be used with responders/nonresponders. Indeed, the number of new biotechnology companies established and industrial partnerships formed to study pharmacogenomics attests to the perceived potential of this approach.\textsuperscript{48,49}

The potential value of pharmacogenomics is multifold. With respect to drug development, the use of pharmacogenomics has the potential to salvage drugs that may not be efficacious on a population basis but might be in a specific subset of patients. An example of this mentioned above is S12024, the experimental drug used to treat Alzheimer’s disease that did not exhibit beneficial effects in a broad population but was effective in those patients with the APO e4 genotype.\textsuperscript{33} When focusing on hypertension, the value of pharmacogenomic studies may be, on the surface, more difficult to grasp. Indeed, it has been pointed out that a vast majority of patients will respond with a decrease in blood pressure to an ACE inhibitor, a diuretic, or a calcium channel blocker, either alone or in combination, and that the “individualization of medical therapy” for hypertension is done empirically. However, if we consider blood pressure to be an intermediate phenotype and the actual disease to be end-organ damage, the benefit of such an approach may become clear. For example, it may become possible to identify patients more susceptible to stroke, myocardial infarction, or renal disease and to treat these individuals more aggressively. Moreover, it may become possible to find subsets of patients who are more susceptible to end-organ damage and who respond differentially to different medication, not just with a decrease in pressure but also with a reduction in risk of end-organ damage. While a physician may be able to empirically define an appropriate treatment for blood pressure, defining an appropriate therapy to prevent end-organ damage is more difficult or impossible to achieve empirically.

**Expression Profiling as a Genomic Tool**

To fully realize the potentials of the EST and STS databases, it would be of great interest to be able to rapidly assess the patterns of expression for these genes, both at the RNA and protein levels (see Table 3). As an example, if a genomic mapping study revealed that a chromosomal region was linked to hypertension, a next step might be to determine whether candidate genes exist in this region. As more ESTs are discovered and mapped, the potential list of candidate genes in any particular region will grow, however, at present, the ESTs will be listed only by number with little other information available. Knowledge of the patterns of expression of those ESTs will be informative and useful for determining the course of future investigation.

The growth of the expression databases will have other benefits for research into the causes of cardiovascular diseases. The identification of genes differentially expressed in disease states may greatly aid the determination of the pathophysiology of the disease and may provide potential therapeutic targets or diagnostic markers. These concepts are not new and have been the paradigm for more than a decade.
TABLE 3. Expression Profiling Technologies

<table>
<thead>
<tr>
<th>Transcript Profiling</th>
<th>Protein Profiling</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Differential display</strong></td>
<td>Potential for novel transcript discovery, not quantitative</td>
</tr>
<tr>
<td><strong>Random sequencing of cDNA libraries</strong></td>
<td>Potential for novel transcript discovery, more quantitative than differential display, semiquantitative depending on quality and characteristics of library, strong bioinformatic support</td>
</tr>
<tr>
<td><strong>Serial analysis of gene expression (SAGE)</strong></td>
<td>Potential for novel transcript discovery, more quantitative than above technologies, simple, strong bioinformatic support</td>
</tr>
<tr>
<td><strong>Microarray technology</strong></td>
<td>Potential to examine all transcripts simultaneously, quantitative, reproducible, strong bioinformatic support, suitable for high throughput, limited by EST database and/or availability of cDNA clones</td>
</tr>
<tr>
<td><strong>2-Dimensional electrophoresis</strong></td>
<td>Potential for novel protein discovery, not quantitative, little current bioinformatic support, labor intensive</td>
</tr>
<tr>
<td><strong>Mass spectrometry</strong></td>
<td>Potential for novel protein discovery, quantitative, rapid, reproducible, little current bioinformatic support</td>
</tr>
</tbody>
</table>

For example, in cardiac hypertrophy and heart failure, a review of the literature reveals that >100 genes have been shown to be differentially regulated. This list comes from studies in multiple organisms (mouse, rat, human), using multiple models (pressure overload, volume overload, coronary artery ligation, viral infection, long-term hypertension) and at multiple points in the development of the disease (acute, compensated hypertrophy, overt failure). Unfortunately, while this is an impressive accumulation of data, it accounts for only 0.1% of the genome. To better understand which gene products are causally related to the disease state versus those that may play compensatory roles, a comparative study examining several models at different time points is desirable. However, before the advent of high throughput technologies, this would be too laborious and time intensive to accomplish for the aforementioned 100 or so sequences, much less for mention a more global study involving thousands of ESTs.

Several approaches have been used to examine transcriptional profiles of expression on a genomic scale. The development of the EST database has itself generated a definition of transcriptional profiles. Widespread sequencing of clones from a cDNA library has allowed the tabulation of genes expressed in different normal and pathologic tissues in a semiquantitative manner. A caveat to this approach is that as investigators in this area attempt to fill the gaps in the EST database by the use of subtracted and normalized cDNA libraries, the quantitative aspects to this approach suffer. Nevertheless, this approach has been used for several tissues, notably by Hwang and colleagues who have amassed a database of cardiac expressed mRNAs (www.tcgu.med.utoronto.ca/homepage.html).

A modification of this approach has been developed by Velculescu and colleagues, called serial analysis of gene expression (SAGE). In this approach, short (9 base pairs; the size, though small, allows for a 95% certainty that the sequence can be uniquely identified), concatamered cDNAs (tags) are constructed and cloned. Individual clones are picked and sequenced; each clone, because of the small size of the insert, contains tags from 20 to 40 distinct transcripts. Such approaches have been used extensively in cancer research and will undoubtedly be used in other areas as well.

With the development of DNA microarray technology, these global approaches to the identification of profiles of expression can be accomplished with improved ease. Several formats for array expression profiling are currently available and involve the arraying of either single synthetic oligonucleotides or PCR-generated cDNAs onto silicon, glass, or nitrocellulose (www.affymetrix.com, www.incyte.com, www.clontech.com). The arrays are probed with radiolabeled or fluorescently tagged cDNA or cRNA generated from mRNA isolated from the test samples and the signals quantified.

Several publications using DNA arrays to examine expression profiles have appeared and have demonstrated the power of the technology. Lockhart and coworkers, using oligomers bound to silicon, demonstrated that the sensitivity of the method is sufficient to detect approximately 0.1 mRNA molecule per cell and yielded a linear response over 3 to 4 orders of magnitude. DeRisi et al. using cDNAs bound to glass, examined changes in expression profiles as yeast underwent changes from aerobic to anaerobic respiration. Several different temporal and directional specific alterations in expression were observed, and, interestingly, these coordinated changes in expression were related to distinct biochemical pathways.

These approaches may prove useful in biological and physiological studies. With respect to studies of biology, the ability to examine all of the thousands of potential transcripts will add immeasurably to the understanding of physiological principles involved in normal regulation as well as in disease development. With this in mind, a systematic examination of normal patterns of gene expression in human tissues has begun at the Brigham and Women’s Hospital (Drs Steven Gullans and Richard Pratt, www.geneindex.org). Moreover, by comparison of normal tissue with pathological tissues, this approach may also lead to the identification of novel targets for drug development and the elucidation of pathways of disease. These approaches will, of course, require the input of bioinformaticists to provide the ability to track and analyze the massive amounts of data that will be generated in these
types of studies. Moreover, the ability to design the appropriate experiments utilizing the appropriate comparisons is vitally important. These approaches have the potential problem to be viewed as not hypothesis-driven but merely exercises in data generation. However, when viewed in genomic terms, these studies can aid in the development of hypotheses. Moreover, the analysis of clusters of transcripts can, in fact, be hypothesis-driven under appropriately designed experiments.

The ability to elucidate patterns of expression in various tissues and disease states will significantly enhance medical and biological studies. However, the elucidation of RNA patterns is only 1 aspect. The vast majority of RNA species function as templates for the production of protein with the obvious exception of transfer RNA, ribosomal RNA, and the rare examples such as H1956 and the small nucleolar RNA (snoRNA57) whose functions are expressed at the RNA level. Therefore, global-scale studies of the protein-coding potential of the genome is 1 of the next major frontiers.18 This new area, termed proteomics,58 is still in its infancy, because large-scale, high throughput technologies are just being developed.

Many of the technologies that will be used in proteomics are modifications of standard protein biochemical assays used for years. For example, large scale, 2-dimensional electrophoresis is being used for the resolution and quantitation of complex protein mixtures. At the largest scale, up to 10 000 proteins can be resolved59; however, even lower resolution can yield important information. For example, The National Cancer Institute Developmental Therapeutic Program (NCI-DTP) has screened 60 000 compounds for activity against 60 cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61 As part of this program, Myers and colleagues have examined the profiles of 150 proteins in the cancer cell lines.60,61

Databases such as those being developed under the NCI-DTP and elsewhere are invaluable resources but are in their infancy compared with the databases that have been developed for genomic studies. To reach its full potential, it will be desirable and necessary to link a genomic position (with its links to polymorphic regions in the protein) to an EST database (with its links to expression patterns) and to a predicted profile on 2-dimensional gel electrophoresis and/or mass spectrometry. To date, most of these databases are proprietary (eg, the Incyte/Oxford Glycosciences collaboration: www.incyte.com/products/lifeprot/index.html), but as more academic investigators enter these fields, the availability of these databases should increase.

Physiologic Genomics: an Emerging Field
What will be the usefulness of these breakthroughs to those investigators in the classical fields of physiology and cell biology? The most obvious benefit will be the discovery of targets for future study (see Table 4). In the past, when a candidate gene was shown to be associated with disease, the function of the candidate gene was already known. Indeed, the candidate gene was chosen on the basis of the known physiological relevance to the disease. However, of the approximately 75 000 to 100 000 potential genes in the human genome, only a small fraction (5% to 10%) are genes of known function. In the future, many novel genes with unknown activities and function will be discovered, and this will provide enormous opportunities for physiological research. This field of physiological genomics will cover a broad array of disciplines, technologies, and approaches (Table 4), including the classic disciplines (eg, biochemistry, cell biology, and whole animal physiology), as well as newer approaches such as transgenics, gene targeting, and congenic approaches; in vivo gene transfer; and studies of protein:protein and DNA:protein interactions.

In the case of a unique protein, how might studies proceed? One approach is to alter the expression of that protein and examine the consequences either in culture or in vivo. Well-known techniques such as antibody blockade of activity, gene transfer, and antisense approaches could be used.

| TABLE 4. Physiological Genomics: Approaches and Technologies |
|---------------------------------|---------------------------------|
| Classical biochemistry, molecular and cellular biology | Characteristics of novel protein, effect of expression in cell culture |
| Yeast 1-hybrid system | Examine protein:DNA interactions |
| Yeast 2-hybrid system | Examine protein:protein interactions, determine biochemical cascade for novel protein |
| Expression profiling | Determine tissue and temporal patterns of expression |
| Classic physiology | Define physiologic role using classic whole animal approaches |
| Transgenic | In vivo effect of expression, potential for tissue specific/regulated expression |
| Gene targeting | In vivo effect of gene deletion/alteration, potential for tissue specific/regulated targeting |
| Congenics | Change genetic background of variant |
| In vivo gene transfer | Rapid analysis of altered expression, potential for tissue specific/regulated effects |
| Clinical studies | Genotype/phenotype correlations |
Another approach is to define the proteins that interact with the unknown protein under study. By examining protein-protein interactions and defining the biochemical pathways in which the unknown protein is involved, one would obtain insight into the potential function of the novel protein and the appropriate plan for future studies of that protein. Recently, techniques such as the yeast 2-hybrid system have been developed to allow the detection of protein-protein interactions with anonymous proteins. This approach takes advantage of the fact that the yeast Gal4 transcription factor can be expressed as 2 separate proteins: a binding domain (BD) necessary for binding to the appropriate DNA element and an activation domain (AD) necessary to induce transcription. These 2 domains can be expressed as chimeric proteins with other heterologous proteins, and if these heterologous proteins can interact and form protein-protein complexes, the AD and BD are brought into proximity and will induce transcription via the Gal4 promoter. A selectable marker can be expressed to allow the selection of yeast containing 2 cloned vectors, each expressing 1 of a pair of proteins that can form protein-protein complexes. This approach can be used to show that 2 known proteins interact; i.e., to show that wild-type but not a mutant p53 can bind to SV40 T antigen or to identify unknown proteins that bind to p53. Using this technology, studies are underway to examine, on a genomic level, all potential interactions in the proteome. Proof of concept was provided with a study of bacteriophage T7, an E. coli phage encoding 55 proteins, which examined, all potential interactions involving these proteins. Recently, a similar approach for the yeast proteome, which contains approximately 6000 proteins (of which 60% have no known function) have been proposed.

The definition of genes involved in the development of disease by gene mapping or expression profiling will open considerable opportunity for the physiological assessment of those genes in vivo. Already, one can find transgenic and homologous recombinant approaches to examine the function of those genes in vivo. Currently, 2000 genes have been targeted and more are accumulating at a rate of 500 week. In addition, gene transfer approaches will allow the more rapid development of animal models to test the function of potential candidate genes. Again, the development of high throughput technologies to rapidly produce expression vectors and viral constructs will greatly aid these approaches. Moreover, development of specific and inducible promoters will further refine these studies.

Conclusions

We are now living through a time of remarkable and impressive changes in biology. The success of the Human Genome Project and associated studies has opened a new field for physiologists, cell biologists, and clinicians. New approaches and new paradigms are needed in hypertension research to take advantage of these advances. Our minds must be opened and remain open to the enormous possibilities in genomic medicine. The basic principles used in sequencing were being perfected some 25 years ago. Who at that time would predict that we would be on the verge of sequencing the entire human genome? Even 10 years ago, when the Human Genome Project was first proposed and funded, how many of us thought it was too early for such massive sums of money to be spent on this endeavor? Now, we are hearing of attempts to define all the variants in the human genome, to document, at the genetic level, the basis of human diversity. Just 15 years ago, many of us were struggling with Northern and Western blots attempting to measure 1 or 2 proteins or RNA transcripts; soon, we will be presented with the opportunity to examine the profile of every protein and potential transcript. The potential opportunities are vast, and as investigators in hypertension research, we must all accept the challenge and take this opportunity to integrate genomic research with physiological and clinical medicine.

Acknowledgments

The authors are supported by NIH grants HL-42663 and HL-59316. The authors wish to thank Drs Gary Gibbons, Matthew Pollman, and Steven Gullans for helpful comments during the writing of this review. VJD is the recipient of NIH MERIT award HL-58516.

References


