DDT Vol. 9, No. 4 February 2004

High-throughput molecular pathology in human tissues as a method for driving drug discovery

Julian Beesley, Christine Roush and Lauren Baker

To facilitate prioritization of potential drug targets, gene expression can be localized to individual cell types in normal and diseased tissues. Given the complexity of molecular physiology and pathology, the creation of large-scale molecular pathology databases collating data obtained from human tissues is a challenging marriage of old and new technologies, particularly when considering the many issues that preclude easy access to substantial quantities of human tissues. Molecular pathology databases are powerful tools and are essential for early-stage drug discovery, enabling informed decisions to be made with respect to scientific direction and follow-up research.

Julian Beesley*
Christine Roush
Lauren Baker
LifeSpan BioSciences
2401 4th Avenue
Suite 900
Seattle
WA 98121, USA
*e-mail: Julian.Beesley@lifespanbiosciences.com

▼ Historically, only a small number of gene families have proven to be excellent drug targets. This is reflected in the fact that approximately 50% of the best-selling drugs target a small number of gene families, which include the G-protein-coupled receptors (GPCRs), kinases, ion channels, nuclear hormone receptors, proteases, transporters, phosphatases and phosphodiesterases. However, the GPCR family is the most targeted gene family [1], with 39 of the top 100 marketed drugs acting directly or indirectly on GPCRs. The completion of the Human Genome Project has provided novel information regarding these gene families and it is expected that this information will yield a rich source of new potential drug targets. A list of GPCR orphans whose native ligands have been recently identified appears to justify the investment in this new research approach, which could potentially lead to new targets for a wide variety of disorders, including inflammation, asthma, cardiovascular disease and cancer [2].

Traditionally, potential drug targets have been selected from specific gene classes or with consideration to specific therapeutic areas. As the genome is analyzed further, the process of identifying novel targets will become less important; the emphasis will shift to the identification of the important characteristics of possible targets before competitors [3], because every pharmaceutical company will potentially have access to the same data. There will be immense pressure to mine this small number of gene families for new drug targets and effective use of the plethora of new technologies and available information will be crucial. To facilitate this process, there are a multitude of databases on the market that are based on a range of established and new technologies, and most include curation of various subsets of data from the public domain. Selection of the most useful databases is a daunting and expensive exercise, because proof of success is apparent only when a competitive edge in the drug discovery process is attained.

Identifying a gene target is just the beginning of the drug discovery process; a bigger challenge is to attribute a value to the protein product of the gene. At present, the problem facing the pharmaceutical industry is that there are too many leads at the level of the gene, and not at the level of the protein product.

Early target prioritization with molecular pathology

Molecular pathology is the study of the cellular basis of disease. Immunohistochemical analysis can link the expression of particular gene products to individual cell types in normal and diseased tissues. The cellular basis of many diseases has been analyzed and, in some cases, the biology of a given cell type

1359-6446/04/$ – see front matter PII: S1359-6446(03)02973-8

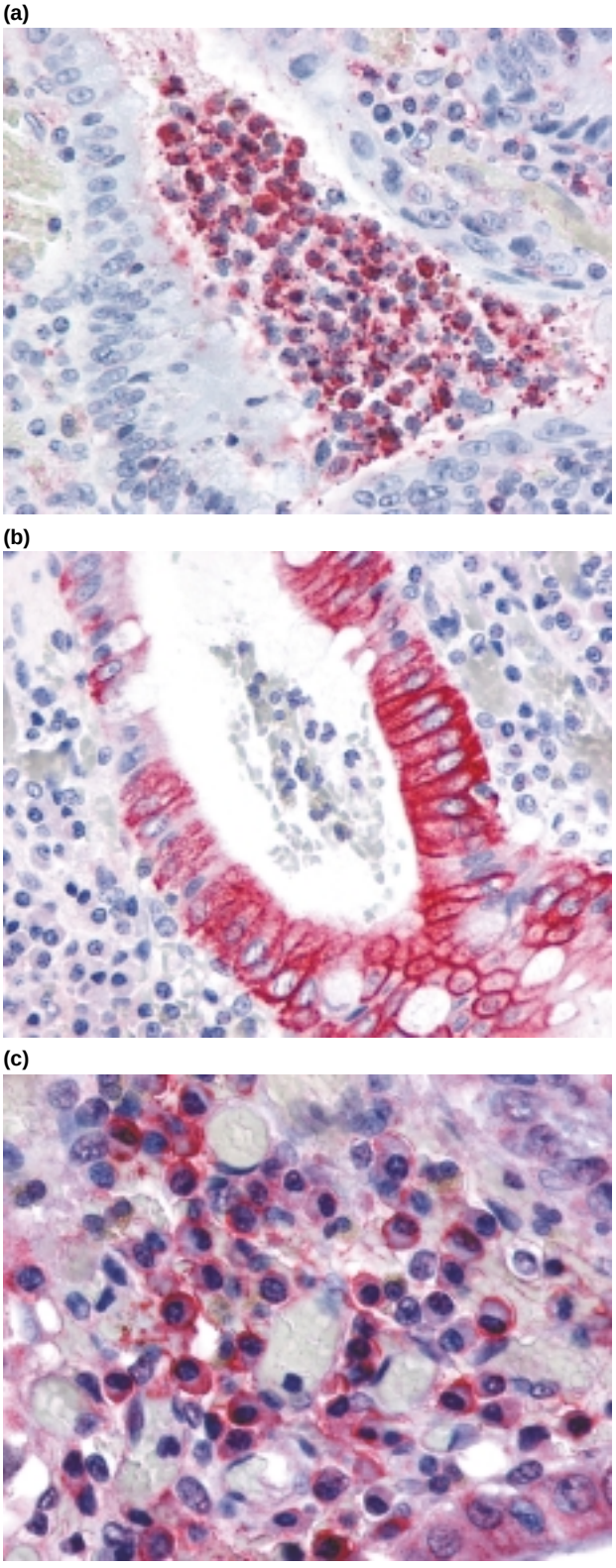

Figure 1. Immunohistochemical labeling of three proteins in colonic tissue from ulcerative colitis shows the value of protein localization in drug discovery. The three proteins are localized to individual human cell types by immunohistochemistry (IHC) (positive cells are stained red, purple/blue staining is hematoxylin counterstain to visualize tissue architecture). Labeling of protein 1 **(a)** is localized to neutrophils in the crypts, labeling of protein 2 **(b)** is localized to the epithelium and labeling of protein 3 **(c)** is localized to plasma cells in the lamina propria. Therefore, there are three different potential drug targets expressed in three different cell types. Assessment of the cell biology of these cells will facilitate the prioritization of these proteins as drug targets.

has been sufficiently characterized to provide an estimate of the 'drugability' of a given target within that cell type. Thus, the knowledge gained from molecular pathology facilitates the prioritization of gene products as potential drug targets. For example, molecular pathology approaches enabled the differential cellular expression of at least three proteins in ulcerative colitis to be elucidated; immunohistochemical staining for one protein was localized to the crypts of the villi, that of a second was localized to epithelial cells and that of a third was localized to plasma cells of the lamina propria (LifeSpan BioSciences proprietary data; Figure 1). Thus, in this example, the use of molecular pathologic and immunohistochemical techniques led to the identification of three distinct cell types, cell pathways and potential targets for investigation. In addition, knowledge of the distribution of a particular protein across an array of tissues can link expression of the gene encoding this protein to particular cell types, which in turn aids in the identification of new drug-use indications and potential toxicity problems, as well as providing information relevant to target validation. From this, it is possible to form an early hypothesis of the value of each potential target and, by taking into consideration the known cell biology, even suggest the drugability of each target. This information is highly applicable to the drug discovery process.

Molecular pathology is not a new field; therefore, given the complexity of molecular physiology and pathology, the creation of high-throughput molecular pathology databases to prioritize drug targets in human tissues requires the involvement of established and new technologies, including *in situ* protein localization techniques and bioinformatic analysis.

High-throughput localization of drug targets in human tissues

Immunohistochemistry and human tissues

Immunohistochemistry (IHC) and *in situ* hybridization (ISH) are two methods that are used to visualize gene expression in tissues and to localize gene expression within specific cell types. However, IHC is more applicable to large-scale localization than ISH, because reagent (e.g. antibody) cost and assay times are often significantly less for

IHC studies. IHC is a well-characterized technology, first described in the middle of the last century, and there is a considerable body of technical and scientific knowledge associated with it [4]. By contrast, ISH is costly and labor intensive, and it can be difficult to achieve high-quality, reproducible results with this technology. In addition, most drug targets have been proteins, not mRNAs. However, this might change in the future as the utility of approaches involving siRNA is determined [5].

Although animal tissues of good quality are readily available, their use is not always applicable to human pathology. Consequently, potential leads have failed at late stages in the drug discovery process because of the poor predictability of animal studies. The value of IHC studies is enhanced considerably if they are performed on human tissues.

There are five factors to consider in tissue localization studies for molecular pathology databases: acquisition of human tissues, target selection, antibody generation, high-throughput IHC and evaluation of immunohistochemical labeling.

Acquisition of human tissues

Human tissues are a valuable resource for the drug discovery process, however, certain issues preclude easy access to large quantities. To construct and operate a large-scale molecular pathology database efficiently, a consistent supply of high-quality tissues is required. It is widely recognized that the development and maintenance of a human tissue bank is a resource-intensive, expensive and long-term endeavor [6]. Furthermore, the collection of human tissues requires the fulfillment of legal, technical and privacy criteria [7,8].

Although the use of human tissues for transplantation is widely accepted, their use for research remains somewhat controversial despite the fact that this is a more appropriate alternative to the use of animal tissues, and that there is potentially a large supply of human tissues available from hospitals, much of which is routinely discarded. Several legal and ethical issues surround the collection and use of human tissues for molecular pathology, all of which become much more complex in studies that include genetic testing where the predisposition of a donor to develop a particular disease might be discovered. Other important issues include the legality of removing tissues from a body and what consent is required [9–11]. In addition, maintenance of patient confidentiality is paramount, and every facility that deals with human tissues must be rigorous in documenting their efforts to maintain the privacy of their patients. A biolaw and bioethics standpoint, which states that individuals possess no proprietary rights to their own bodies or body parts, has recently been challenged [12–14]. Several human tissue banks have been established, including Medical Solutions Biomaterials Resource (UK; http://www.biomaterialsresource.com), the Human Brain and Spinal Fluid Resource Center (USA; http://www.loni.ucla.edu/~nnrsb/NNRSB), Harvard Brain Tissue Resource Center (USA; http://www.brainbank.mclean.org) and the AMP Leukemia and Lymphoma Tissue Bank (Australia; http://www.ampgroup.com) to name a few. Many researchers strive to acquire their own stock of human tissues, in which case it is highly recommended that they are familiar and completely up to date with all published guidelines [6,15,16]. However, several biotechnology companies, including Ardais Corporation (http://www.ardais.com), Asterand (http://www.asterand.com), Clinomics Biosciences (http://www.clinomicslabs.net), Genomics Collaborative (http://www.genomicsinc.com) and LifeSpan BioSciences (http://www.lifespansbiosciences.com) have been established specifically to work with or supply human tissues. There is also a patient group (PXE International; http://www.pxe.org), established to support those suffering from *Pseudoxanthoma elasticum*, which has created its own human tissue bank and is making it available to researchers in an attempt to encourage research into this disease.

A major technical challenge in the accurate localization of potential drug targets is the complexity of the human body itself, which has ~100 organs, 1500 cell types and 10000 identified diseases. If a molecular pathology database is to be useful, a consistent supply of the majority of these tissues, or at least those related to the research areas of interest, is essential. Each tissue sample requires a defined pathologic analysis and the relevant associated patient information, for example, age, sex, diagnosis and post-mortem interval (where applicable). Some studies will also benefit from knowledge of the drug-treatment history and clinical outcome data. Each tissue sample requires validation of diagnosis, tissue architecture and target stability (particularly if high-throughput screening of many targets is to be performed). These constraints create a greater challenge as the size of the database increases and even more so if infectious material is being handled [17].

The acquisition of human tissues, unlike that of animal tissues, is not predictable and surgical procedures are often delayed or cancelled. Therefore, it is often necessary to wait many months for tissue samples, particularly in the case of tissues that are difficult to obtain (e.g. those from asthma sufferers), which are collected when available. The usefulness of the data obtained from tissue analysis is dependent upon specimen quality, which could be affected by conditions that contribute to the degradation of the specimen before it has undergone processing and analysis.

Not all human tissues are suitable for molecular pathologic analysis. For example, when post-mortem tissues are collected, it is often the agonal state of the tissues before death, not post-mortem delay, which defines tissue quality. The degree of degradation might vary from tissue to tissue, although it has been shown that lung tissue remains reasonably stable for up to 5 h after excision [18]. Because tissue quality is dependent on the duration of time before fixation, validation of tissue quality upon receipt is crucial.

Tissue validation involves ascertaining that a given tissue has the necessary pathology and that the specimen preparation was adequate. Hospital personnel initially obtain and diagnose specimens for medical reasons. Once a tissue has been received by a tissue bank, qualified pathologists also analyze the specimen. In this role, the pathologist has been termed a 'tissue refiner' and 'data miner' [19]. The first stage of the procedure performed by the pathologist is to examine sections from the tissue block to confirm the hospital diagnosis and to ensure that the diseased area has not already been excised during preparation for the hospital diagnosis. Different countries use different nomenclatures, and different research studies might require different stages of a disease process. For example, although a patient might not exhibit symptoms of diabetes, their pancreas could show some signs of early disease process. This might occur if 'normal' tissues are obtained from a patient who died as the result of another disease. To avoid any ambiguity, a thorough knowledge of the disease process must be understood at this stage. The second stage of tissue validation is to immunolabel sections of the tissue for an antigen that is known to be present in that particular tissue. If this immunohistochemical assessment fails, the tissue must be discarded. Tissue management is therefore an important process that requires the combined skills of pathologists and experts in tissue processing and IHC.

Target selection

Target selection could be narrowed down to gene class or classes of interest, or to therapeutic areas of study. For a database, it is necessary to select targets of interest for all potential users. If the database comprises a complete set of information (e.g. a complete gene family), no selection is required; however, if the database comprises a subset of information, for example, 70% of the members of a particular gene family, it is preferable to allow the stakeholders some choice, thereby ensuring high-priority selection of genes to the database.

Antibody generation

If a database, and therefore a high-throughput approach, is used, extreme care must be taken to review all aspects of the available technology, selecting one with the greatest overall reliability and efficiency. The quality of the database is founded on antibody specificity, and hence the selection criteria for antibodies are crucial. Many antibodies are commercially available, and these should be characterized with respect to performance in frozen- or formalin-fixed, paraffin-embedded human tissues and with respect to staining of positive and negative control tissues. This information is generally available from the literature. Where published expression data are lacking, northern blot studies can be used. Other assays used to characterize antibody specificity include immunolabeling and western blot studies on transfected cell lines and, when available, western blot studies on positive and negative control human tissues. Antibodies should be selected with the database goals in mind. The selection parameters change from the study of a few gene products to that of many gene products. Preservation of tissue architecture can be emphasized for single antibody studies and frozen- or paraffin-embedded sections are used with several antibody types (e.g. monoclonal, polyclonal). Preservation of tissue architecture is also a goal in molecular pathology database construction, and affinity-purified polyclonal antibodies generated against peptide antigens are typically used. However, paraffin-embedded sections are preferred because this is the method used to prepare the majority of hospital samples. In addition, it is easier to store paraffin-embedded blocks than frozen blocks of tissue for extended periods of time. If maximum sensitivity is required, frozen sections should be used to eliminate the loss of antigen during the fixation and paraffin-embedding procedures. Ultimately, the number of gene products to be analyzed will determine the time allowable for antibody optimization.

High-throughput IHC

High-throughput molecular pathology requires the use of a large number of tissues, which becomes difficult if the tissues are not readily available, as is the case with many human tissues. Tissue microarrays (TMAs), which are slides bearing several small tissue sections, are increasingly used to generate a large amount of information in a relatively short period of time [20]. Cylindrical tissue samples of up to 1000 or more tissues [21], but usually ranging from 50–100 tissues, are prepared from individual archival tissue blocks and are subsequently placed into a single recipient paraffin block. Sections can then be used for different types of analysis, including IHC and ISH. Multiple studies have shown that findings obtained through the use of TMAs are highly representative of those obtained with their donor tissues, despite the often small diameter used, which can be 0.6 mm [21]. Further research has shown that microarray

studies of breast cancer, when compared to conventional large sections from a single sample, identified ~95% of the information obtained from the larger sections for the oestrogen receptor, 75–81% for the progesterone receptor, and 70–74% of the information for p53 [22]. Analysis of three antibodies on four TMAs yielded equivalent or more significant association with tumor-specific survival than large-section analysis. A single sample from each tumor was sufficient to identify associations between molecular alterations and clinical outcomes. According to Torhorst *et al*. [22], tissue heterogeneity did not negatively influence the predictive power of the TMAs. The value of this technique was also shown using an array constructed from archived pathology tissue blocks [23]. A 20-year survival analysis was performed on a cohort of more than 600 patients with the use of only a few microliters of antibody in a single experiment. The use of paraffin-embedded tissues has limitations with regard to the analysis of RNA and certain proteins do not withstand the processing necessary to produce a block [24]. To address this issue, cryoarrays, which allow for the processing of multiple frozen-tissue specimens in a single array, could be used. Considerable effort is expended in the preparation and analysis of TMAs, and it is essential to adhere to the strict guidelines relating to the quality of the tissues and antibodies used.

Evaluation of immunolabeling - human or computer?

Accurate analysis of immunolabeling and cell identification data is crucial and requires skilled knowledge of pathology and technical knowledge of IHC. Slide analysis can be resource intensive and time consuming, but the cell type and intensity of immunolabeling should at least be recorded. The intensity of immunolabeling can be scored from 0 to 4 (0=negative, 1=blush, 2=faint, 3=moderate, 4=strong). Although these values are somewhat subjective, they do provide an indication of differences between specimens and cell types. The evaluation process can be shortened by the barcoding of slides and the use of instant imaging and voice recognition software to generate reports relatively quickly. However, with regard to pathology, the time-consuming aspect remains the analysis of the slides with the majority of the time considered routine screening. With the advent of sophisticated computer technology, accurate automated morphologic analysis of slides is now being developed.

LifeSpan BioSciences has built an automated image capture and analysis system that uses a fully motorized microscope to facilitate pathology analysis of slides. Slides are manually loaded into a cassette and fed automatically across a stage where the images are captured and cell types and key structures are identified. The immunohistochemical signals are analyzed for signal intensity and distribution throughout the nuclear and cytoplasmic compartments within the cell types and other structures of interest. This analysis provides the localization, frequency and intensity of labeling of the protein. The values for immunolabeling intensity are normalized to pathologist scores, which enables direct comparisons to be made between different cell types and/or experiments. This system is also being developed for toxicological examinations, where rat livers exhibiting lesions can be distinguished from normal tissues. With the advent of sophisticated computer technology, automated morphologic analysis of slides is now becoming a reality, along with the potential for considerable decreases in analysis time. Although computers can retrieve published data, estimate the degree of immunolabeling and even distinguish cell type, it remains the domain of human analysis to detect correlations and to develop hypotheses.

Bioinformatics

The term bioinformatics defines the field of study that merges biology with aspects of computer science [25]. This field was born out of necessity to handle the wealth of genomic data available, from genome sequences and functional genomics [26]. Bioinformatics seeks to collate and analyze this considerable body of information, which is widely available but which, without expert analysis, can be overwhelming. Ideally, the information should be collected, collated, analyzed and presented in a way that is comprehensible to scientists and project leaders and should be readily accessible and easily linked to related experimental results. Information obtained from the use of bioinformatic analysis has become an integral part of research and development in the biomedical sciences [27,28].

For the scientist, target selection often begins with the elucidation of a human gene sequence that encodes a potential drug target. Structured information available for this target in the public domain might include name, synonyms, and DNA, expressed sequence tag (EST) and protein sequences. Target selection decisions also involve the use of a wide range of largely unstructured sources often available in the public domain [3] to obtain information on where the target is expressed, the normal physiological role of the protein, the pathological role (if any) of the protein, disease relevance of the target, the potential of side-effects of drugs to the target, the patent outlook for the target and the technologies that would be appropriate for further study.

Sequence analysis

The need to compare newly acquired sequences to those already collected led to the development of sequence alignment programs to identify regions of sequence similarity,

such as FASTA [29] and BLAST [30]. BLAST was proven to be more efficient and is based on mathematical foundations to calculate statistics for high-scoring segment pairs and the principle of searching for common patterns shared by sequences. Recently developed tools such as PSI-Blast [31] and hidden Markov models [32] have increased the sensitivity of sequence similarity searches by as much as threefold [33]. These approaches can identify sequence homologs and can be used to infer function or to facilitate the study of evolution of protein families or domains. For example, a newly discovered gene product could be classified as a GPCR by virtue of the fact that it contains the 7-transmembrane domain signature of this family.

These same sequence analysis tools can be applied to the analysis of gene-based sequence-tagged sites [34] or ESTs [35]. Because the cDNA library tissue source for each EST is known, expression data for an uncharacterized sequence of interest can be amassed by using the BLAST program to compare all known ESTs for that sequence. To avoid overestimation or underestimation of the number of ESTs that match the uncharacterized sequence, care must be taken to normalize the EST 'hits' to the total number of ESTs sequenced from each tissue type. A tissue distribution pattern of EST matches can yield insights into gene function. For example, a gene with sequence similarity to a large number of ESTs from brain libraries, but to no other tissues, might play a role in neuronal function. In addition, disease association with cancer could be indicated by sequence similarity to ESTs isolated from carcinoma libraries but not those from normal tissues.

Clues to the normal physiological function of an uncharacterized protein can be uncovered by analysis with protein domain databases [36–38], which contain information on conserved structural motifs, including transmembrane domains, signal sequences and functional domains (e.g. kinase active-site domains, phosphorylation sites, leucine zippers and 3D folding domains). By comparison with characterized proteins, the identification of a signal sequence, a transmembrane domain and a kinase domain within the protein sequence indicates that the protein could be a receptor, and therefore a potential drug target. Unfortunately, the emphasis on molecular and cellular features and observations regarding post-translational modifications can be sparse and difficult to locate consistently in the many protein databases. Boguski and McIntosh [39] and Maglott [40] have reported on the difficulty in searching Swiss-Prot and LocusLink to obtain information on groups of proteins based on functional classification or well-known pathways, which would aid in understanding the physiological function of a particular protein.

Gene expression analysis

Data generated from molecular pathology regarding the localization of gene expression in individual cell types in normal and diseased human tissues also benefit from bioinformatic analysis. IHC data obtained from localization studies, followed by pathologic analysis and characterization across a wide array of normal and diseased tissues, can aid in the understanding of the physiological role of a protein in several ways. For example, the cellular localization of the protein can be determined at the light-microscopic level as membranous, cytoplasmic or nuclear. The overall distribution pattern could show that the protein is localized to a subset of neurons in the hippocampus or is a lymphatic marker present on inflammatory cells throughout the body, and thus can indicate potential disease associations. Alternatively, breast cancer association could also be identified if antibody staining is more intense in the neoplastic epithelial cells of breast carcinoma compared to that in the normal breast ductal epithelium.

Analysis of whole body IHC distribution can be used to cluster proteins of unknown function via hierarchical clustering analysis (HCA) [41,42]. HCA forces data points into a strict hierarchy of nested subsets. Each of the closest pairs is bundled and represented by a single point defined by their average, and it is this point that is compared to the next closest neighbor [43]. HCA could place an orphan-GPCR into the chemokine family, even though its ligand is unknown, thus aiding in the understanding of its physiological role.

Relevant literature reports regarding a gene product can be retrieved from Medline by the use of known synonyms. These reports can also be mined for expression data by analysis of medical subject headings (MeSH) or by scoring the available literature for expression data and summing the individual results. Although this information is useful, particularly in the prediction of potential positive control tissues, it might not provide a representative tissue distribution because research on a specific gene product is often biased toward particular tissues of interest.

Databases by their nature are large-scale composites of data from many sources. Although there are many databases available, not all contain the considerable body of knowledge that should ideally be found in one place and viewed in a logical manner. In-house experimental data should be available, along with published data, sequence data and other information. To enable a true comparison, specific information for one gene should be viewed alongside the same relevant information for other genes, and the sum knowledge for one gene should be comparable with the sum knowledge for other genes. The value of a database is directly dependent upon the quality of the data it contains.

A well-constructed database is a powerful tool in which all relevant literature, sequence and experimental data are present and easily accessible in a single format. Unless a database can provide information quickly and accurately, it is of little use.

Conclusions

Challenges to the pharmaceutical industry include the need to accelerate the drug discovery process against a backdrop of increased numbers of potential targets, and to produce drugs with increased safety and efficacy while at the same time reducing costs. The localization of gene products by molecular pathology aids the drug discovery process by highlighting potential targets, identifying new drug-use indications and helping to reduce late-stage attrition rates through the identification of potential toxicity problems. Expression data are useful from the early stage of drug discovery to clinical trials, and through the diagnostic process.

With the aid of a molecular pathology database, a disease is selected, and localization data are presented. Gene expression that occurs in specific cells in normal and diseased tissues is identified, and the relevance of this information, as well as prior use in drug discovery efforts, can be identified by searching the curated public data in the database. The database can be used effectively by the experienced scientist for hypothesis generation, as well as by the relative newcomer searching for information on genes of interest. The database approach is also cost effective. There is an initial substantial cost risk to the database constructor, but the risk can be managed by making the same datasets available to several customers. From the perspective of the customer, it is less risky to join a consortium than to produce either small amounts of data in-house and risk not understanding the overall picture, or to produce large amounts of data in-house and have to overcome potentially serious technical and logistic challenges as well as incurring considerable expense. The judicious selection of databases by pharmaceutical companies will supply essential knowledge for early stage drug discovery and will aid in the decision-making process with regard to follow-up research and scientific direction.

References

1 Menzaghi, F. *et al.* (2002) Constitutively activated G protein-coupled receptors: a novel approach to CNS drug discovery. *Curr. Drug Target CNS Neurol. Disord.* 1, 105–121
2 Dowell, S.J. (2001) Understanding GPCRs-from orphan receptors to novel drugs. *Drug Discov. Today* 6, 884–886
3 Sumner-Smith, M. (2001) Managing the drug target selection process. *Modern Drug Discov.* 4, 51–52
4 Beesley, J.E. (2000) Introduction. In *Immunohistochemistry and In Situ Hybridization in the Biomedical Sciences*, (Beesley, J.E., ed.), pp 1–5, Birkhauser, Basel
5 Arenz, C. and Schepers, U. (2003) RNA interference: from an ancient mechanism to a state of the art therapeutic application? *Naturwissenschaften* 90, 345–359
6 Hulette, C.M. (2003) Brain banking in the United States. *J. Neuropathol. Exp. Neurol.* 62, 715–722
7 De Moor, G.J. *et al.* (2003) Privacy enhancing techniques – the key to secure communication and management of clinical and genomic data. *Methods Inf. Med.* 42, 148–153
8 Berman, J.J. (2002) Confidentiality issues for medical data miners. *Artif. Intell. Med.* 26, 25–36
9 Cruz-Sanchez, F.F. *et al.* (1997) Ethical aspects to be considered in brain banking. *Ann. Ist. Super. Sanita* 33, 477–482
10 Oosterhuis, J.W. *et al.* (2003) Tumour banks: well-guarded treasures in the interest of patients. *Nat. Rev. Cancer* 3, 73–77
11 Kort, E.J. *et al.* (2003) A human tissue and data resource: an overview of opportunities, challenges, and development of a provider/researcher partnership model. *Comput. Methods Programs Biomed.* 70, 137–150
12 Beyleveld, D. and Brownsword, R. (2000) My body, my body parts, my property? *Health Care Anal.* 8, 87–99
13 Savulescu, J. (2002) No consent should be needed for using leftover body material for scientific purposes. Against. *BMJ* 325, 648–651
14 van Diest, P.J. (2002) No consent should be needed for using leftover body material for scientific purposes. For. *BMJ* 325, 648–651
15 Masui, T. *et al.* (2001) Japanese guidelines on research use of human materials and disease information. *Kokuritsu Iyakuhin Shokuhin Eisei Kenkyusho Hokoku* 119, 40–46
16 Sarris, M. *et al.* (2002) Banking for the future: an Australian experience in brain banking. *Pathology* 34, 225–229
17 Bell, J.E. and Ironside, J.W. (1997) Principles and practice of 'high risk' brain banking. *Neuropathol. Appl. Neurobiol.* 23, 281–288
18 Jewell, S.D. *et al.* (2002) Analysis of the molecular quality of human tissues: an experience from the Cooperative Human Tissue Network. *Am. J. Clin. Pathol.* 118, 733–741
19 Becich, M.J. (2000) The role of the pathologist as a tissue refiner and data miner: the impact of functional genomics on the modern pathology laboratory and the critical roles of pathology informatics and bioinformatics. *Mol. Diagn.* 5, 287–299
20 Packeisen, J. *et al.* (2003) Demystified …tissue microarray technology. *Mol. Pathol.* 56, 198–204
21 Bubendorf, L. *et al.* (2001) Tissue microarray (TMA) technology: miniaturized pathology archives for high-throughput *in situ* studies. *J. Pathol.* 195, 72–79
22 Torhorst, J. *et al.* (2001) Tissue microarrays for rapid linking of molecular changes to clinical endpoints. *Am. J. Pathol.* 159, 2249–2256
23 Rimm, D.L. *et al.* (2001) Tissue microarray: a new technology for amplification of tissue resources. *Cancer J.* 7, 24–31
24 Hoos, A. and Cordon-Cardo, C. (2001) Tissue microarray profiling of cancer specimens and cell lines: opportunities and limitations. *Lab. Invest.* 81, 1331–1338
25 Luscombe, N.M. *et al.* (2001) What is bioinformatics? A proposed definition and overview of the field. *Methods Inf. Med.* 40, 346–358
26 Reichhardt, T. (1999) It's sink or swim as a tidal wave of data approaches. *Nature* 399, 517–520
27 Molidor, R. *et al.* (2003) New trends in bioinformatics: from genome sequence to personalized medicine. *Exp. Gerontol.* 38, 1031–1036
28 Kanehisa, M. and Bork, P. (2003) Bioinformatics in the post-sequence era. *Nat. Genet.* 33(Suppl), 305–310
29 Lipman, P.J. and Pearson, W.R. (1985) Rapid and sensitive protein similarity searches. *Science* 227, 1435–1441
30 Altschul, S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
31 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
32 Krogh, A. *et al.* (1994) Hidden Markov models in computational biology. Applications to protein modeling. *J. Mol. Biol.* 235, 1501–1531

33 Park, J. *et al.* (1998) Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. *J. Mol. Biol.* 284, 1201–1210
34 Olson, M. *et al.* (1989) A common language for physical mapping of the human genome. *Science* 245, 1434–5
35 Adams, M.D. *et al.* (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252, 1651–1656
36 Falquet, L. *et al.* (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res.* 30, 235–238
37 Sonnhammer, E.L. *et al.* (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 28, 405–420
38 Haft, D.H. *et al.* (2001) TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res.* 29, 41–43
39 Boguski, M.S. and McIntosh, M.W. (2003) Biomedical informatics for proteomics. *Nature* 422, 233–237
40 Maglott, D.R. *et al.* (2000) NCBI's LocusLink and RefSeq. *Nucleic Acids Res.* 28, 126–128
41 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
42 Cho, R.J. *et al.* (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* 2, 65–73
43 Orr, M.S. and Scherf, U. (2002) Large-scale gene expression analysis in molecular target discovery. *Leukemia* 16, 473–477