

Leading article

Functional genomics in gastroenterology

Identification of the DNA structure as a double stranded helix consisting of two nucleotide chain molecules was a milestone in modern molecular biology. This discovery has led to a rapidly accumulating body of genomic information. The dramatic developments in genetic analysis during the past decade have placed genomics at the forefront of life science research. Large multinational projects (for example, the human genome project) have generated a host of sequence, mapping, and expression data. At the same time, new technologies for automated sequencing, data analysis, genotyping, expression, and protein analysis have been developed. An important result is identification of the genetic aetiology of many monogenic diseases and the enormous progress in the exploration of polygenic disorders.

This article reviews the possible applications and future developments of genomic technologies in disorders of the gastrointestinal tract.

Development of genomic technology

In recent years, significant progress in the elucidation of the pathophysiology and aetiology of gastrointestinal disorders has been made. Novel experimental therapies which specifically target single steps in disease relevant regulatory pathways have been designed and are in different stages of clinical development. Examples are inhibition of the gastric proton pump in peptic ulcer disease¹ and anti-tumour necrosis factor α therapy which has been developed for Crohn's disease and rheumatoid arthritis.^{2–4} This process was made possible by technological advances in cell biology, biochemistry, and microbiology and their application to disorders of the gastrointestinal tract. While this methodological approach has produced substantial advances in our understanding of gastrointestinal pathophysiology and increased available treatment options, there are currently several new challenges that call for the development of novel research techniques.

- (i) The pursuit of hypothesis driven functional research limits the scope of discovery to known genes and pathways and may therefore hamper the elucidation of novel genes and regulatory mechanisms.
- (ii) Exploration of signalling cascades and their role in disease pathophysiology has reached an enormous degree of complexity. Therefore, distinction between primary and secondary events is impossible in most cases. Even the exact mechanisms of a successful therapeutic application of targeted immune therapies (for example, therapy of Crohn's disease with tumour necrosis factor binding monoclonal antibodies) are not completely clear.^{5,6} Application of exploratory techniques

based on a hypothesis free approach using high throughput ("parallel") methodologies may help to establish a hierarchy of known and new players in signalling pathways and to identify key events in the mechanism of anti-inflammatory therapeutics.

- (iii) A genetic—or familial—component has been identified in many gastrointestinal disorders, including gastrointestinal malignancies, inflammatory bowel disease, susceptibility to *Helicobacter pylori* related pathology, pancreatitis, bile stone formation, and others.^{7,8}

Progress in the different fields outlined above may rely to different extents on the availability of genomic information and technology.

Methodological problems in the systematic exploration of disease pathophysiology

Taking a cell biological hypothesis to a specific test has for a long time been the only possible approach.^{9,10} The published experimental results from these studies are always biased towards a positive result with suppression of negative data. In addition, in vitro experimental systems are often highly sophisticated and results may not be fully applicable to the complex in vivo situation. Two recent examples may illustrate this situation: leukotrienes were regarded for a long time as the driving forces for chronic intestinal inflammation.^{11–13} Convincing and reproducible data were generated that showed upregulated leukotriene production which paralleled the activity of colonic inflammation in inflammatory bowel disease. It was also assumed that the anti-inflammatory action of glucocorticoids is mainly due to induction of lipocortin, a protein which blocks phospholipase A2 and therefore inhibits generation of leukotrienes.^{14,15} It took more than a decade until clinical trials with zileuton and other effective inhibitors of leukotriene synthesis showed that a dramatic reduction in leukotriene B4 did not result in a marked improvement in intestinal inflammation in ulcerative colitis.^{16,17} The role of leukotriene B4 in disease pathophysiology has since been re-evaluated. It now appears that the main anti-inflammatory action of glucocorticoids is by inhibition of activation of the transcription factor nuclear factor kappa B.^{18–21} Although inhibition of nuclear factor kappa B activation explains many of the pathophysiological and clinical findings in the analysis of glucocorticoid action, it is not clear if another, more important, mechanism of action may be discovered in future.

Abbreviations used in this paper: EST, expressed sequence tag; IMAGE, integrated molecular analysis of genomes and their expression; NCBI, National Center for Biotechnology Information; SAGE, serial analysis of gene expression.

The reason for using genomic technology

A close interaction with cell biological and immunological agendas is expected to be promoted through a systematic hypothesis free analysis of gene expression and protein interactions. This article therefore describes examples of genomic technologies that allow the highly parallel evaluation of large numbers (hundreds to several tens of thousands) of known and unknown genes (in the form of ESTs—expressed sequence tags). This methodology therefore offers the possibility of a hypothesis free approach and may identify new targets for further in depth exploration and therapeutic interventions.

The search for genetic susceptibility factors in gastrointestinal disorders requires the use of technological approaches that are conceptually very different from pathophysiology based research. The most attractive potential of genomic exploration is differentiation between primary and secondary events in disease pathophysiology, which is extremely difficult (with the likely exception of infectious disorders) using the analytical methods of cell biology and immunology. The high degree of complexity and redundancy of regulatory cascades as well as the obvious disadvantages of hypothesis driven research efforts appear to be the limiting factors in the discovery of novel pathways. The redundancy of physiological regulation is clearly demonstrated by a series of genetically altered, gene deficient (“knock out”) mice which all develop a similar (for example, colitis-like) phenotype.^{22, 23} Therefore, it appears unlikely that a primary cause of a complex disease can be identified by a “trial and error” algorithm which is used in most cases to dissect pathophysiology. The positional gene identification approach on the other hand—as demonstrated for example by identification of the haemochromatosis gene—has the potential to uncover entirely novel and unexpected molecules which will influence the pathophysiological understanding of the disease process.²⁴

Tools for a genomic exploration of pathophysiology

A critical idea in the progress of the human genome project was the introduction of systematic sequencing of cDNA libraries.²⁵ The public databases (<http://www.ncbi.nlm.nih.gov>) currently contain over 1.6 million of these fragments, which have been named expressed sequence tags (ESTs) (see glossary in table 1). Only 10–15% of all human genes have yet been defined. The availability of EST databases facilitates the identification of cDNA fragments ultimately resulting in the cloning of potentially novel genes. However, EST libraries also carry a large redundancy with many overlapping or partially identical fragments. Therefore, techniques have been developed to group ESTs into “UniGene” clusters (table 1) (representing only sequences which are unique in the library). These techniques are based on computational (in silico) methods using sequence homologies (<http://www.ncbi.nlm.nih.gov>, <http://www.tigr.org>) or on oligonucleotide fingerprinting analyses (see below).

The availability of mapping approaches that do not need polymorphic markers (that is, radiation hybrid mapping) has allowed a genomic localisation of the ESTs and the assembly of genome wide transcript maps.²⁷ An example of the EST representation of a known gene (STAT6) and its position on the radiation hybrid gene map is detailed in figs 1 and 2. Some limitations of the EST based approach are instantly evident from fig 1: (i) the 5' ends of genes are underrepresented in these databases and (ii) the complete coding mRNA sequence can rarely be derived from the EST data alone.

EXPRESSION ANALYSIS—TECHNOLOGY

The aim of genomic expression analysis techniques is measurement of gene expression of (ideally) all human genes in a parallel and automated fashion. Hybridisation based techniques require in general certain scientific and technological steps.

- (i) Generation of cDNA libraries representing the genes of interest.
- (ii) Clustering of the libraries to identify a non-redundant or “UniGene” set of transcripts.
- (iii) Representation of the identified transcripts on a hybridisation carrier, for instance through spotting of

Table 1 Glossary of terms used

EST	“Expressed sequence tag”. A piece of cDNA sequence generated from a randomly chosen cDNA clone that can later be retrieved by its address. Important tool that gives a handle to the sequence of unknown genes.
IMAGE	“Integrated molecular analysis of genomes and their expression”. The IMAGE Consortium was initiated in 1993 on a collaborative basis. The consortium shares high quality arrayed cDNA libraries and place sequence, map, and expression data on the clones in these arrays into the public domain. All of the clones are available free of any royalties and may be used by anyone agreeing with the guidelines of the consortium.
Linkage	Two genetic markers (or a trait and a marker) are <i>genetically linked</i> if they are transmitted together within families.
Linkage disequilibrium	More common co-occurrence of alleles at distinct genetic markers than would be expected on the basis of the individual allele frequencies alone.
UniGene	Non-redundant subset of the EST sequence databases that ideally represents each gene in an organism exactly once.
NCBI	“National Center for Biotechnology Information”, based at the NIH. It maintains a central database repository for the sequence and mapping data generated in the human genome project (www.ncbi.nlm.nih.gov).
SAGE	“Serial analysis of gene expression”, a representational cloning method that can be used to quantify gene expression by analysis of unique concatenated sequence fragments from the genes of a tissue/cell.

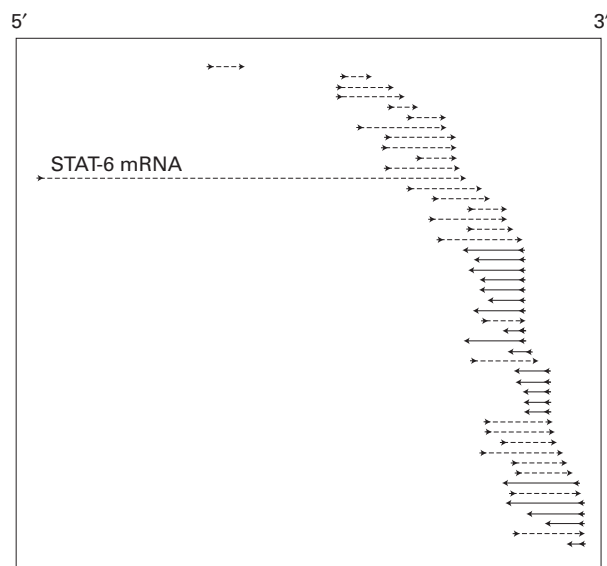


Figure 1 Illustration of expressed sequence tag (EST) coverage of a gene. The figure shows the STAT6 mRNA sequence which is involved in signalling of the interleukin 4 receptor. The sequence has been aligned with ESTs clustered in the corresponding NCBI UniGene cluster (Hs.181015). Sequences are shown as arrows with the respective sequence orientation (arrow in 5' to 3' direction). It is evident that the 5' end of the gene is underrepresented. Without the a priori knowledge of the full length mRNA, the cDNA fragments would have been clustered differently.

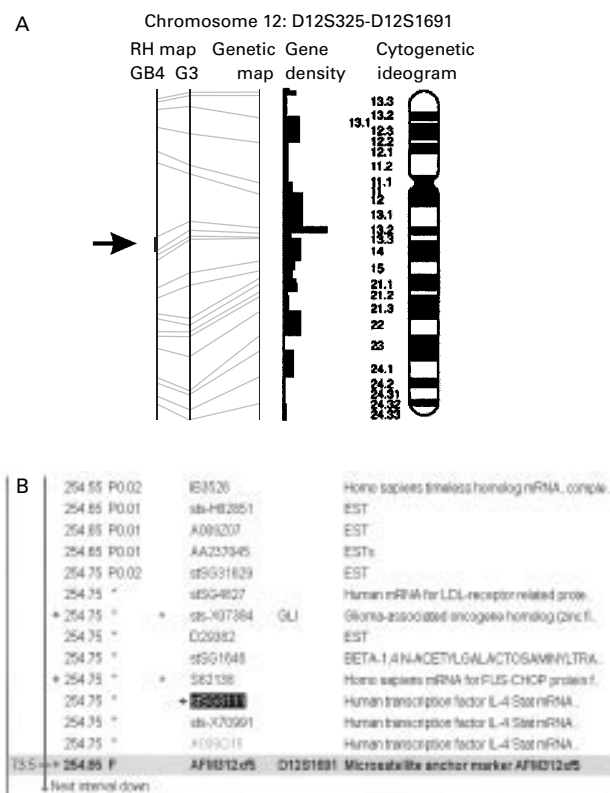


Figure 2 Example of the localisation of expressed sequence tags (ESTs) on a radiation hybrid (RH) map. The screenshots are derived from the NCBI website (www.ncbi.nlm.nih.gov). The interval in which the UniGene cluster from figure 1 localises is indicated by an arrow. Chromosome 12 is also the location of one of the major susceptibility regions in inflammatory bowel disease²⁶ which spans the pericentromeric region.

representative polymerase chain reaction products on membranes or glass slides^{28–30} or the use of representative oligonucleotides on a carrier.³¹

- (iv) Determination of the relative abundance (that is, expression) of the corresponding mRNAs (for instance, from diseased tissues) by complex cDNA hybridisations.
- (v) Image and data analysis.

An alternative to hybridisation based methods is representational cloning approaches such as serial analysis of gene expression (SAGE) (see table 1).³² These methods rely on identification of oligomers characteristic for each transcript that are identified by sequencing random linear assemblies of characteristic sequence elements. Particularly for identification of rare transcripts, this method requires large amounts of sequences to be generated for each experiment. A central repository for SAGE sequence data has been established at the National Center for Biotechnology Information (NCBI: <http://www.ncbi.nlm.nih.gov/SAGE>). It is not yet clear what importance this technique will attain relative to the hybridisation based expression analysis techniques.^{33–34}

In this report we will focus on hybridisation based methods for expression analysis.

GENERATION OF LIBRARIES

For generation of an appropriate cDNA library, high quality mRNA has to be prepared from appropriate tissues. The libraries in the EST sequencing projects have been derived from mainly non-gastrointestinal organs and tissues (for example, brain, liver, lymphatic cells, muscle). However, genes relevant for gastrointestinal pathology may not be adequately represented. Until now only a few libraries

have been generated from gastrointestinal organs. Therefore, collaborative projects to generate intestine and liver specific libraries have been started. In the first step, reverse transcribed cDNA molecules from tissue have to be randomly fragmented and inserted into the appropriate bacterial vectors. The size for a tissue specific cDNA library is typically a few hundred thousand clones. In order to reproducibly use the individual clones from a library, single colonies can be chosen in an automated way from agar plates and transferred into 384 well plates. The randomly distributed colonies are identified using image analysis software. The software selects clones to be picked on the basis of user specified criteria, including the roundness of colonies, their closeness to neighbouring clones, and others. Clone positions are identified, and coordinates are transferred to a robotic manipulation system which uses an automated arm to “pick” colonies. A single stainless steel tip is fired into the colony which sticks to the pin and is subsequently delivered into a microtitre plate. Current systems are capable of picking and inoculating approximately 3000 clones per hour into 384 well microtitre plates prefilled with growth media. The microtitre format of the libraries allows long term storage, analysis, and subsequent standardised retrieval of individual clones on the basis of the microtitre plate address.

CLUSTERING OF LIBRARIES

The libraries derived as outlined above will contain many redundant clones (that is, fragments of cDNA representing different overlapping or identical parts of the same transcript; example shown in fig 1). As a prerequisite for efficient use in later experiments, this redundancy has to be eliminated. This can be achieved by direct sequencing of cDNA clones, by oligonucleotide fingerprinting, cross hybridisations of the library on itself, or a combination of these methods. The datasets generated by all methods can be used to produce a non-redundant set of cDNA clones from the complete cDNA expression library (for example, a “UniGene” representation of the library). At the same time, information on the relative abundance of each transcript is generated because highly expressed transcripts will be represented in the library by a larger number of clones. The information generated by oligonucleotide fingerprinting techniques allows the matching of clones and to the theoretical (“in silico”) fingerprints of database sequences. Compared with EST sequencing, oligonucleotide fingerprinting can be conducted at 10-fold lower costs per clone. Therefore, much larger libraries can be analysed, which will result in the identification and accurate quantification of the expression level of a higher number of rarely expressed genes.³⁵ In addition, the more complete sequence coverage increases the chance to detect internal sequence variants (mutations and alternative splice forms), and simplifies the identification of related sequences. Since known sequences can be directly identified, based on the match of the observed to the predicted fingerprint, and clones from clusters representing unknown genes can be partially sequenced, the combination of fingerprinting and EST sequencing currently represents the most effective approach to identifying new genes. Analysis of both EST sequencing and oligonucleotide fingerprinting data requires complex statistics and results depend on the quality of the library and the experimental data. Therefore, “non-redundant” clusters of EST sequences (as shown in an example from the NCBI UniGene set in fig 1) may in reality represent multiple genes or may be parts of transcripts represented in other clusters also.

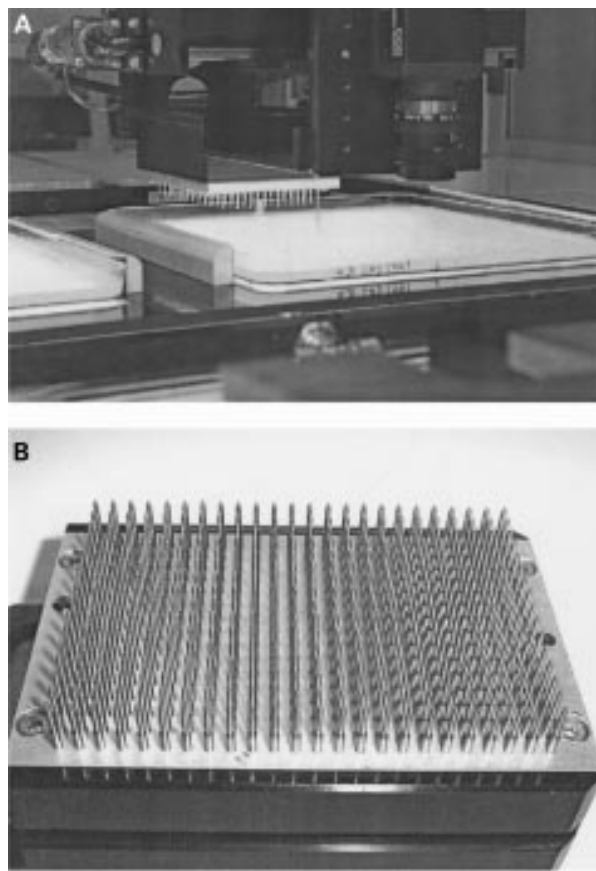


Figure 3 Robotic head used for automated picking of colonies from agar plates. (A). A camera and an image analysis system allow identification of colonies. A 384 pin head with 384 individually spring loaded pins is used as a stamp tool for both spotting and colony picking (B). The tip diameter depends on the spotting application and varies between 150 and 450 μm . The spotting head is moved on a Servo controlled three axis linear drive system with an accuracy of 5 μm . A complete run handles up to 72 microtitre plates and includes reading the barcode identifier, lifting the microtitre plate lid, transferring 384 clones in parallel onto nylon membranes or glass slides, sterilising the pins in sodium hydroxide/ethanol, and drying them with a hot air stream. With the system described here it is possible to immobilise up to 147 456 clones on a single 22 cm \times 22 cm nylon surface. Microarrays with 50 μm feature size can be produced with identical technology. Using sophisticated pin designs, 225 spots can be placed on a 1 mm \times 1 mm surface area. Thereby a typical microscope glass slide would allow over 250 000 single spots. This would easily facilitate immobilisation of the complete set of 20 000–100 000 genes generally thought to be typical for a eukaryotic organism for screening on a single slide.

SPOTTING ON HYBRIDISATION CARRIERS

To generate an expression array, DNA from all representative clones forming the non-redundant UniGene set has to be transferred to a filter or glass slide. Prior to spotting, the re-arrayed library is amplified by polymerase chain reaction for spotting rather than directly spotting the clones on the filters. This eliminates variations in the DNA amount which could be caused by different growth kinetics of the clones. DNA is spotted automatically in a full robotic set up from microtitre plates on nylon membranes or glass slides. An example of a spotting robot is demonstrated in fig 3. An alternative to attaching full length polymerase chain reaction products to a solid support is an oligonucleotide representation of each clone which can be generated and attached to a silicium based support medium.³¹ The latter approach is used for some commercial libraries including the ones manufactured by Affymetrix.

COMPLEX HYBRIDISATIONS

For expression analysis in diseased tissues or cell cultures, mRNA is reversely transcribed to cDNA which is then

Grid key

8	0	10	5	7	12
12	11	3	31	15	2
4	14	7	6	13	3
0	10	2	0	8	1
5	13	11	15	0	14
4	9	1	31	6	9

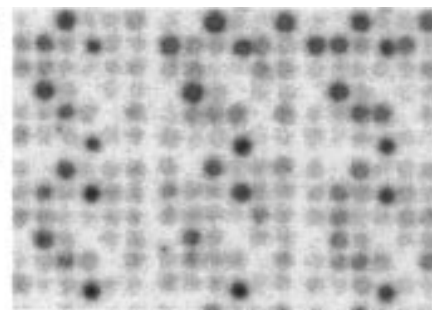


Figure 4 Example of a microarray hybridisation result. The figure shows a small part of a microarray hybridised with mRNA from normal colonic. The structure is part of a 35 000 clone microarray which carries the UniGene representation of the IMAGE library. *A. thaliana* has been used as a control and clones have been spotted in duplicate. Already the hybridisation with normal colonic lamina propria derived cDNA shows the high numbers of positively detected signals.

radioactively or fluorescently labelled and hybridised onto the array. Typical experiments require between 10 and 50 μg of total tissue RNA per experiment. This amount of RNA can for instance be obtained from 2–5 typical colonic biopsies. The results of the hybridisation are read by either direct exposure of x ray films, phosphor imaging, or fluorescent scanning.

These experiments involve hybridising reverse transcribed mRNA from a particular tissue, cell, probe, patient sample, etc., onto the UniGene array. The signals for each gene are then analysed to determine the relative abundance in the samples of mRNA derived cDNAs corresponding to each gene. Internal controls need to include back hybridisations with short oligomers, appearance of housekeeping genes in the array, and presence of control clones (for example, *Arabidopsis thaliana*) in the array. Controls also need to include the admixture of control RNA (for example, from *A. thaliana*) to the tissue RNA probe.

Figure 4 shows a typical example from a complex hybridisation. The magnified view is from an array of 35 000 UniGene clusters from the IMAGE library.

BIOINFORMATICS FOR EXPRESSION ANALYSIS

Images obtained by these methods are highly complex and require the use of sophisticated analysis software, which is available from different academic and commercial sources (for example, from GPC-AG under www.gpc-ag.com). Image analysis still requires substantial human interaction until expression data for each clone are obtained. The statistical analysis and evaluation of expression data from the signals obtained from thousands of clones (that is, the definition and differentiation of clusters) is the subject of intensive statistical research.³⁶ In particular, the statistical problems of false positive results (that is, due to multiple testing) are not yet adequately resolved.

The term cluster analysis (first used by Tryon³⁷) actually encompasses a number of different classification algorithms. The question is how to organise observed data into meaningful structures—that is, to develop taxonomies. A powerful exploratory data analysis tool is most likely the two way joining procedure³⁸ which is useful in circumstances when one expects that both cases and variables will simultaneously contribute to the uncovering of meaningful patterns of clusters. An example is the identification of clusters of patients that are similar with regard to particular clusters of similar measures of expression of certain genes. The difficulty with interpreting these results may arise from the fact that the similarities between different clusters may pertain to (or be caused by) somewhat different subsets of variables. Thus the resulting structure (clus-

ters) is by nature not homogeneous. This may seem confusing at first and, compared with other clustering methods including joining (tree clustering and K-means clustering), two way joining is probably the least commonly used. However, two way joining may be the most powerful among these exploratory data analysis tools.

APPLICATION OF EXPRESSION ANALYSIS

Expression analysis can be applied to any regulatory phenomenon both in cell lines and cell cultures. Initial experiments have included the response of defined cells to certain stimuli such as serum on fibroblasts³⁹ or interferon on a fibrosarcoma cell line.⁴⁰ The complexity of the systems tackled has increased to aging processes in mouse muscle⁴¹ or to T cell activation in mice.⁴² Gastrointestinal tissues have been investigated in comparison with cancer (normal colon versus colon cancer).³⁶ A pilot study of differential expression in inflammatory bowel disease tissue has been published as an abstract.⁴³ We anticipate that a broader and systematic use will be made of these techniques in the coming years for exploration of gastrointestinal diseases. The potential applications include differential characterisation of expression patterns describing disease pathophysiology but also exploration of mechanisms of drug action. In longitudinal short interval studies, repetitive sampling, which is possible with a limited invasion in the gastrointestinal tract, can be used for a longitudinal characterisation of expression patterns during the course of disease.

SYSTEMATIC INVESTIGATION OF PROTEIN INTERACTION WITH PROTEIN ARRAYS

A logical extension from expression analysis into cell function is complete analysis of all proteins, the proteome. Because of the enormous number of proteins which are potentially involved in the cause or pathophysiology of complex disorders, an automated approach, with very high throughput, is required for protein analysis. Until recently, it was not possible to analyse proteins using the same high density arrayed approach as for cDNA based expression studies. However, novel tools have been developed to generate protein arrays. Recombinant proteins are clonally expressed from arrayed cDNA libraries allowing direct connection from DNA sequence information on individual clones to protein products, and back again, on a whole genome level. This approach, which has been pioneered by Cahill and coworkers, makes translated gene products directly amenable to high throughput experimentation, generates a direct link between protein, expression, and sequence data, and facilitates additional generation of gene product-antibody catalogues.²⁸

Proteomic research is highly dependent on the quality and extent of databases which are currently being further extended (examples include the SWISS-PROT, TrEMBL, and PROSITE databases). Database access is bundled through systems such as the ExPASy (Expert Protein Analysis System, <http://www.expasy.ch>). ExPASy proteomic tools use annotations of the SWISS-PROT entries which reflect post-translational modifications as well as sequence variants which are important for prediction. The SWISS-MODEL server and its front end, the Swiss-PdbViewer, are a protein modelling tool also suitable for non-expert use. Currently, more than 65 000 model structures have been modelled from over 200 000 known protein sequences.

Future applications in our laboratories will include identification of new autoantigens by screening of sera for novel antibody specificities and identification of new transcription factors using labelled random sequences from inflammation gene promoter regions.

Pharmacogenomics

Many pharmacological interventions produce response rates of only approximately 30–50%.^{2 44 45} Clinical characteristics are often not sufficient to predict a favourable drug response. One can therefore hypothesise that genetic predisposition may play a role in this situation. This genetic claim is, however, very difficult to substantiate in the absence of a molecular lead finding. The main difficulty is that heritability cannot be established with the typical methods of genetic epidemiology (familial clustering or concordance in monozygotic twins) as is possible for other polygenic traits. Systematic linkage approaches are not possible because the relevant families cannot be recruited. On the other hand, the signalling pathways of many drugs are often known in detail or can be explored by expression analysis. Thus corresponding genes can be specifically evaluated. In principle, the drug responder trait may be linked to one of the disease genes or, alternatively, may represent genes which are completely distinct from the actual disease genes. Because the therapeutic principle of most drugs is pathophysiology and not aetiology based, the latter scenario is more likely. A recent example of this approach is the association of variants in the ALOX5 promoter and the response to an investigational leukotriene synthetase inhibitor in asthma.⁴⁵ A systematic investigation of a drug responder requires examination of all genes encoding relevant signalling molecules. At this point, expression analysis technology, as described above, may also be used to identify additional molecules involved in the mechanisms of drug response.

Genomic technologies to examine the genetic aetiology of gastrointestinal disorders

Genetic predisposition is suspected to play an important role in the aetiology of many gastrointestinal disorders. The spectrum of genetic determination spans from clearly inherited diseases such as haemochromatosis (classic Mendelian trait leading to high penetrance) to polygenic and highly complex disorders. A polygenic predisposition is suspected for the development of complications in infectious disorders such as hepatitis and *H pylori* induced gastroduodenal inflammation. Many clinically important gastrointestinal diseases including inflammatory bowel disease, coeliac disease, and diabetes mellitus have to be placed in the middle of this spectrum.^{46–48} It is assumed that genetic predisposition conveys the basic susceptibility but that epigenetic (for example, lifestyle or environmental factors) are needed to a different extent as trigger factors for the actual manifestation of disease. The epidemiology of these “complex” disorders is characterised by familial clustering and increased concordance of the phenotype in monozygotic twins. Once heritability of a disease is established it will allow the use of a systematic genetic analysis in the search for the molecular variants that convey the predisposition. Systematic genome wide linkage analyses have resulted in the definition of genomic susceptibility regions in a number of these disorders which likely contain one or multiple disease causing mutations.⁴⁹ The technology for these large scale experiments has been available for less than a decade. The next experimental step—identification of specific risk mutations in genes within the large linkage intervals—is much more complicated and has not yet found a resounding and widely applicable solution. Current approaches include systematic candidate gene analysis⁵⁰ and linkage disequilibrium studies in susceptibility intervals.⁵¹ Several requirements for successful execution of these experiments are only now becoming available through the progress of the human genome project. These include: (i) information on *all* genes in a genomic region which will be available with the complete genomic sequence of the human genome; (ii) the ability to perform

large scale mutation detection experiments is expanded due to recent technical advancements in the design of automated sequencing machines (automated loading, capillary arrays instead of manually cast gels, HPLC based heteroduplex analysis); (iii) efficient methods for high throughput single nucleotide polymorphism typing methods such as Taqman or mass spectroscopy (MALDI-MS, matrix assisted laser desorption/ionisation-mass spectroscopy); and (iv) novel bioinformatics and statistical analysis methods that may allow more powerful gene focused analyses than the conventional association designs.

Researchers have established replicated linkage regions and association leads for a number of these disorders which may prove instrumental in successful positional gene identification.⁴⁹

Conclusions

The technologies outlined above in our view represent tools that will have a major impact on gastroenterology based research in the future. However, these methods are still under development and have shortcomings that need to be resolved. This applies to all of the genomic technologies discussed. Major unclear issues include: procedures for standardisation and data analysis from the highly complex expression analysis experiments (applying to both SAGE or hybridisation techniques); conformational properties and degree of representation of the native protein structures on the protein arrays; and the appropriate experimental algorithm for the identification of actual risk genes for complex disorders. As of now, much of the work involved in finding these genes is of a highly exploratory nature.

Despite these problems, genomic technologies hold substantial potential for novel and important discoveries in gastrointestinal disorders. At the same time, genomic tools will not replace biochemical, cell biological, and immunological methods. For example, unknown transcripts which are identified through expression screening will have to be characterised. Full length mRNA, subcellular localisation, regulation, and precise interaction with other molecules will have to be determined. Ultimately, one wants to identify and validate single critical molecules that regulate large parts of signalling cascades and can therefore serve as successful drug targets.

The degree to which the apparent complexity of the pathophysiological regulation mechanisms can be reduced is not clear at this point. Researchers may also be left with systems of multiple genes and gene variants which are too complex to identify or predict the important drug targets or prognostic markers. Statistical and bioinformatics tools and methods to handle the enormous complexity will need to be developed. Recognised patterns are prone to false positivity because of their complexity and the multiple testing involved. Therefore, the consequent replication and verification of any results in independent and well characterised secondary cohorts and experimental set ups (for pathophysiology) is imperative. We anticipate that genomic technologies will prompt a closer cooperation of clinical and molecular expertise. At the end of the day only the combination of a reliable and prospective clinical characterisation which parallels the level of molecular sophistication will allow the robust definition and clinical use of a molecular epidemiology of complex gastrointestinal disorders.

S SCHREIBER
J HAMPE

*I Medizinische Klinik, Christian-Albrechts-Universität,
Kiel, Germany*

H EICKHOFF
H LEHRACH

*Max-Planck-Institut für Molekulare Genetik,
Berlin, Germany*

Correspondence to: Dr Stefan Schreiber, I Medizinische Klinik,
Schittenhelmstr. 12, D-24105 Kiel, Germany. s.schreiber@mucoosa.de

This work was supported by grants from Deutsche Forschungsgemeinschaft (SCHR512/1-3, SCHR512/5-1, SFB 415), BMBF (Competence Network-IBD), and MFG.

- Lauritsen K, Rune SJ, Bytzer P, et al. Effect of omeprazole and cimetidine on duodenal ulcer. A double-blind comparative trial. *N Engl J Med* 1985;312:958-61.
- Targan SR, Hanauer SB, van Deventer SJ, et al. A short-term study of chimeric monoclonal antibody cA2 to tumor necrosis factor alpha for Crohn's disease. Crohn's Disease cA2 Study Group. *N Engl J Med* 1997;337:1029-35.
- Stack WA, Mann SD, Roy AJ, et al. Randomised controlled trial of CDP571 antibody to tumour necrosis factor-alpha in Crohn's disease. *Lancet* 1997;349:521-4.
- Baert FJ, Rutgeerts PR. Anti-TNF strategies in Crohn's disease: mechanisms, clinical effects, indications. *Int J Colorectal Dis* 1999;14:47-51.
- D'Haens GR, Aerden I, van Hogeand R, et al. Duration of response following cessation of infliximab therapy for active or fistulizing Crohn's disease. *Gastroenterology* 1999;116:G3029.
- Bauditz J, Haemling J, Ortner M, et al. Treatment with tumour necrosis factor inhibitor oxpentifylline does not improve corticosteroid dependent chronic active Crohn's disease. *Gut* 1997;40:470-4.
- Schreiber S, Hampe J. Genetics in gastrointestinal disease: how much can we learn from molecular analysis? *Int J Colorectal Dis* 1999;14:10-2.
- Hugot JP, Zouali H, Lesage S, et al. Etiology of the inflammatory bowel diseases. *Int J Colorectal Dis* 1999;14:2-9.
- Sjogvist U, Ost A, Lofberg R. Increased expression of proliferative Ki-67 nuclear antigen is correlated with dysplastic colorectal epithelium in ulcerative colitis. *Int J Colorectal Dis* 1999;14:107-13.
- Stallmach A, van Look M, Scheffele F, et al. IgG, albumin, and sCD44 in whole-gut lavage fluid are useful clinical markers for assessing the presence and activity of pouchitis. *Int J Colorectal Dis* 1999;14:35-40.
- Casellas F, Guarner F, Antolin M, et al. Abnormal leukotriene C4 released by unaffected jejunal mucosa in patients with inactive Crohn's disease. *Gut* 1994;35:517-22.
- Rachmilewitz D, Stampler JS, Karmeli F, et al. Peroxynitrite-induced rat colitis—a new model of colonic inflammation. *Gastroenterology* 1993;105:1681-8.
- Noronha-Blob L, Lowe VC, Muhlhauser RO, et al. NPC 15669, an inhibitor of neutrophil recruitment, is efficacious in acetic acid-induced colitis in rats. *Gastroenterology* 1993;104:1021-9.
- Lauritsen K, Laursen LS, Bukhave K, et al. In vivo effects of orally administered prednisolone on prostaglandin and leucotriene production in ulcerative colitis. *Gut* 1987;28:1095-9.
- Wallner BP, Mattaliano RJ, Hession C, et al. Cloning and expression of human lipocortin, a phospholipase A2 inhibitor with potential anti-inflammatory activity. *Nature* 1986;320:77-81.
- Hawkey CJ, Dube LM, Rountree LV, et al. A trial of zileuton versus mesalazine or placebo in the maintenance of remission of ulcerative colitis. The European Zileuton Study Group For Ulcerative Colitis. *Gastroenterology* 1997;112:718-24.
- Roberts WG, Simon TG, Beslin RG, et al. Leukotrienes in ulcerative colitis: results of a multi-center trial of a leukotriene synthesis inhibitor, MK591. *Gastroenterology* 1997;112:725-32.
- Auphan N, DiDonato JA, Rosette C, et al. Immunosuppression by glucocorticoids: inhibition of NF-kappa B activity through induction of I kappa B synthesis. *Science* 1995;270:286-90.
- Scheinman RI, Cogswell PC, Lofquist AK, et al. Role of transcriptional activation of I kappa B alpha in mediation of immunosuppression by glucocorticoids. *Science* 1995;270:283-6.
- Schottelius AJ, Baldwin AS Jr. A role for transcription factor NF-kappa B in intestinal inflammation. *Int J Colorectal Dis* 1999;14:18-28.
- Schreiber S, Nikolaus S, Hampe J. Activation of nuclear factor kappa B in inflammatory bowel disease. *Gut* 1998;42:477-84.
- Kuhn R, Lohler J, Rennick D, et al. Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* 1993;75:263-74.
- Eynon EE, Flavell RA. Walking through the forest of transgenic models of human disease. *Immunol Rev* 1999;169:5-10.
- Feder JN, Gnirke A, Thomas W, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* 1996;13:399-408.
- Adams MD, Dubnick M, Kerlavage AR, et al. Sequence identification of 2,375 human brain genes. *Nature* 1992;355:632-4.
- Satsangi J, Parkes M, Louis E, et al. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* 1996;14:199-202.
- Schuler GD, Boguski MS, Stewart EA, et al. A gene map of the human genome. *Science* 1996;274:540-6.
- Busso K, Cahill D, Nietfeld W, et al. A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Res* 1998;26:5007-8.
- Maier E, Crollius HR, Lehrach H. Hybridisation techniques on gridded high density DNA and in situ colony filters based on fluorescence detection. *Nucleic Acids Res* 1994;22:3423-4.
- Zweiger G. Knowledge discovery in gene-expression-microarray data: mining the information output of the genome. *Trends Biotechnol* 1999;17:429-36.
- McGall G, Labadie J, Brock P, et al. Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc Natl Acad Sci USA* 1996;93:13555-60.
- Velculescu VE, Zhang L, Vogelstein B, et al. Serial analysis of gene expression. *Science* 1995;270:484-7.
- Spinella DG, Bernardino AK, Redding AC, et al. Tandem arrayed ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles. *Nucleic Acids Res* 1999;27:e22.
- Hashimoto S, Suzuki T, Dong HY, et al. Serial analysis of gene expression in human monocytes and macrophages. *Blood* 1999;94:837-44.

- 35 Meier-Ewert S, Lange J, Gerst H, *et al.* Comparative gene expression profiling by oligonucleotide fingerprinting. *Nucleic Acids Res* 1998;26:2216–23.
- 36 Alon U, Barkai N, Notterman DA, *et al.* Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* 1999;96:6745–50.
- 37 Tyron RC. *Cluster analysis*. Ann Arbor, MI: Edwards Brothers, 1939.
- 38 Hartigan JA. *Clustering algorithms*. New York, NY: Wiley, 1975.
- 39 Iyer VR, Eisen MB, Ross DT, *et al.* The transcriptional program in the response of human fibroblasts to serum. *Science* 1999;283:83–7.
- 40 Der SD, Zhou A, Williams BR, *et al.* Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc Natl Acad Sci USA* 1998;95:15623–8.
- 41 Lee CK, Klopp RG, Weindruch R, *et al.* Gene expression profile of aging and its retardation by caloric restriction. *Science* 1999;285:1390–3.
- 42 Teague TK, Hildeman D, Kedl RM, *et al.* Activation changes the spectrum but not the diversity of genes expressed by T cells. *Proc Natl Acad Sci USA* 1999;96:12691–6.
- 43 Dieckgraefe B, Stenson W, Swanson P, *et al.* Characterization of mucosal gene expression in inflammatory bowel disease by direct hybridization to massively parallel oligonucleotide arrays. *Gastroenterology* 1998;114:A964.
- 44 Shiffman ML, Hofmann CM, Contos MJ, *et al.* A randomized, controlled trial of maintenance interferon therapy for patients with chronic hepatitis C virus and persistent viremia (in process citation). *Gastroenterology* 1999;117:1164–72.
- 45 Drazen JM, Yandava CN, Dube L, *et al.* Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment. *Nat Genet* 1999;22:168–70.
- 46 Orholm M, Munkholm P, Langholz E, *et al.* Familial occurrence of inflammatory bowel disease. *N Engl J Med* 1991;324:84–8.
- 47 Dominici P, Bellentani S, Di Biase AR, *et al.* Familial clustering of Helicobacter pylori infection: population based study. *BMJ* 1999;319:537–40.
- 48 Martin BC, Warram JH, Rosner B, *et al.* Familial clustering of insulin sensitivity. *Diabetes* 1992;41:850–4.
- 49 Becker KG, Simon RM, Bailey-Wilson JE, *et al.* Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. *Proc Natl Acad Sci USA* 1998;95:9979–84.
- 50 Hampe J, Hermann B, Bridger S, *et al.* The interferon-gamma gene as a positional and functional candidate gene for inflammatory bowel disease. *Int J Colorectal Dis* 1998;13:260–3.
- 51 Spielman RS, McGinnis RE, Ewens WJ. The transmission/disequilibrium test detects cosegregation and linkage. *Am J Hum Genet* 1994;54:559–60.

Web extra

A new website has been developed to help gastroenterologists calculate CDAI (Crohn's Disease Activity Index) and PCDAI (Paediatric Crohn's Disease Activity Index).

The user registers his/her patient under a name/code and registers the sex, age, and weight of the patient, and is then asked to fill in raw data. The website will then calculate CDAI (and for patients below 19 years, PCDAI). The site enables the user to store indices over time, and the disease activity of patients will also be presented as a diagram. The data are stored on the Örebro Medical Centre Hospital for one year. You can visit the website at:

http://www.orebroll.se/rso/barn/crohn/Home_Eng.asp

Users must first register; this is done by entering a user name and password, which the user chooses (user names and passwords are not distributed by Dr Ludvigsson).

For further information: **Dr Jonas Ludvigsson,**
Scientific Coordinator,
Department of Paediatrics,
Örebro Medical Centre Hospital,
701 85 Örebro,
Sweden

jonasludvigsson@yahoo.com
www.geocites.com/jonasludvigsson