Transcriptomics and proteomics of human skin

Bastiaan J. H. Jansen and Joost Schalkwijk

Date received (in revised form): 24th September 2002

Abstract

The epidermis protects the organism against physical, chemical and biological challenges, and it acts as a signalling interface between the environment and the body. In order to perform these functions, the epidermal keratinocytes express a wide range of genes, several of which have been characterised previously. Recently, significant progress has been made in the large-scale analysis of keratinocyte gene expression, enabling a more profound insight into keratinocyte biology and human skin diseases. Transcriptome analysis — serial analysis of gene expression (SAGE) and microarrays — and proteome analysis have been performed on intact human epidermis and on keratinocytes cultured in model systems that mimic normal and diseased human epidermis. Here, we review the current state of large-scale gene expression analysis of human skin, with an emphasis on SAGE and complementary DNA microarrays. The merits and limitations of various approaches (transcriptomics versus proteomics) are discussed and the practical issues such as sample preparation from skin biopsies, and the use of in vitro models are briefly addressed.

INTRODUCTION

The human skin: structure and function

Human skin consists of an epidermis and an underlying dermis or connective tissue. The epidermis forms the outermost skin layer and is a terminally differentiating, stratified squamous epithelium (Figure 1). As the human epidermis encompasses the whole body, it serves as a signalling interface between the organism and the environment. The main cell type in this epithelium is the keratinocyte. The inner, basal layer (stratum basale) of the epidermis

Figure 1: Structure of the human epidermis. For explanation, see text
The human skin protects us against environmental challenges

The main cell type of the human epidermis is the keratinocyte

Various in vitro keratinocyte culture models have been developed

contains keratinocytes that proliferate and supply the upper, spinous layer (stratum spinosum) with cells that undergo a terminal differentiation programme through the granular layer (stratum granulosum). The outward migration of differentiating cells ultimately leads to a covalently cross-linked layer of dead, flattened cells, giving rise to a strong barrier called the stratum corneum or cornified layer. This process of differentiation in the epidermis requires a carefully choreographed, complex programme of gene expression. One of the most important functions of the human epidermis is to protect us from physical dangers, such as dehydration, ultraviolet (UV) radiation and from harmful chemical substances. An equally important feature of the epidermis is that it also protects us against immunological challenges, as evidenced by the presence of specialised immune or Langerhans cells in the epidermis and the ability of keratinocytes to express and respond to cytokines, chemokines and growth factors in both a paracrine and autocrine manner. Furthermore, it is becoming increasingly clear that keratinocytes invest a large effort in the expression of genes that are involved in host defence against microbes, such as defensins, cystatins, LL-37, secretory leukoproteinase inhibitor (SLPI) and skin-derived anti-leukoproteinase (SKALP/elafin).

Keratinocyte and epidermal culture models

In order to study keratinocytes more comprehensively, culture models have been developed in order to facilitate investigations of normal keratinocyte biology, the response of keratinocytes to cytokines, chemokines and growth factors, as well as to enable the large-scale isolation and analysis of genes expressed at the messenger RNA (mRNA) and protein levels. Various different models have been developed that are associated with different phenotypes and thus with differences in gene expression. Rheinwald and Green first described a submerged culture model that allows growth of cells on a feeder layer of lethally γ-irradiated 3T3 cells, in conditioned medium containing serum and growth factors. Several groups then developed systems that do not require a feeder layer or conditioned medium, but instead allow the growth of keratinocytes directly on a culture dish, reducing problems such as overgrowth by fibroblasts. Serum-free culture systems have been developed in which keratinocytes can be grown to high density, and can be kept in a proliferative state before being switched to a medium without growth factors to induce early differentiation or to a medium containing defined factors that induce other, specific biological or pathological responses in keratinocytes. In addition to submerged culture models, which allow early differentiation but not the formation of a complete epidermis with a cornified layer, air-exposed models have been developed and further refined. In these systems, keratinocytes are grown on either dead de-epidermised dermis or a matrix of collagen containing active fibroblasts, and these systems mimic, at least histologically, the epidermis. In vitro differentiation can be assessed by immunohistochemical analysis of differentiation markers, which have been extensively studied. Expression of cytokeratins 5 (CK5) and 14 (CK14) is usually associated with proliferative keratinocytes in the basal layer of the epidermis, whereas the expression of CK1 and CK10 is associated with early differentiation. Other markers include transglutaminase and involucrin, and these are associated with late differentiation. In vitro systems, although different at heart, have proven to be very useful in the research of human skin diseases, as differences in gene expression or the expression of stress-markers such as SKALP/elafin can be easily monitored in cultured cells. Until recently, these analyses were done on a relatively small scale, as technologies that allow the global analysis of gene expression were either
non-existent or in development, and thus out of reach for the average laboratory. Today, various methods are available that allow the wide-scale analysis of gene expression, potentially enabling researchers to get more insight into the genetic networks responsible for epidermal and keratinocyte phenotypes. In the following section, the state of large-scale gene expression on the mRNA level (transcriptomics) and protein (proteomics) level in human epidermis and keratinocytes will be reviewed.

**GENE EXPRESSION PROFILING TECHNOLOGIES: THE UNDERLYING DIFFERENCES**

Conventional expression profiling technologies

Although various technologies are available to study gene expression at the mRNA and protein levels, there are fundamental differences between them. The oldest method for analysing RNA expression levels is Northern blot analysis, involving the transfer of electrophoretically separated RNA molecules from an agarose gel to nitrocellulose membranes. Specific RNA molecules can be detected by hybridisation with $^{32}$P-labelled DNA or RNA probes followed by autoradiography. This procedure allows the detection of specific RNA molecules with high sensitivity and low background, but, although it allows the comparison of expression levels of a limited number of genes, the technique is of little value to those who want to investigate gene expression on a genome-wide scale.

Another method that has been used successfully for gene expression profiling is the RNase protection assay, the basis of which is a solution hybridisation of a single-stranded, discrete-sized anti-sense probe(s) to an RNA sample. Small-volume solution hybridisation is far more efficient than traditional membrane-based hybridisation. To quantitate mRNA levels using RNase protection assays, the intensities of probe fragments protected by the sample RNA are compared with the intensities generated from either an endogenous internal control (relative quantitation) or known amounts of sense-strand RNA. Although this technique is more sensitive than Northern blotting, it is equally hampered by a low throughput, and only a small number of genes can be analysed in parallel.

A widely used technique is subtractive hybridisation, numerous variations of which have been reported; in general, these involve the hybridisation of one population of complementary DNAs (cDNAs; the tester) to an excess of mRNA or cDNA from another population (driver), followed by a separation of the unhybridised fraction (target; upregulated cDNAs) from the hybridised common cDNAs. Earlier protocols required a lot of input material, which is highly inconvenient and laborious, and sometimes impossible to obtain from small samples. Therefore, various variants have emerged that take advantage of the polymerase chain reaction (PCR) to enable the use of a lower amount of input material, to bolster cDNA yields of differentially expressed genes or to take advantage of suppressive PCR in suppression subtractive hybridisation (SSH) to lower the number of false-positives and to enrich rare cDNAs 1,000–5,000-fold. These modern versions are collectively being referred to as representational difference analyses (RDA). On the upside, this technology potentially allows the identification of many differentially regulated genes at once, even those that are expressed at very low levels, and, as such, allows true differential gene expression analysis; the downside is that it is not suitable for quantitative measurements of gene expression.

One of the most popular and widely used techniques to investigate differential gene expression is probably differential display-PCR (DD-PCR; also referred to as mRNA differential display), since it is
based on easy to apply, affordable molecular techniques, and relies on PCR. Variations of the methodology exist, the most notable being RNA arbitrarily primed PCR (RAP-PCR), which was developed independently from DD-PCR and which is based on the use of arbitrary primers in both reverse transcription and PCR. Differential display is based on a series of steps: the isolation of undegraded cellular RNA; reverse transcription of the RNA with an ‘anchored’ oligo-dN1–2dT12–18 to produce a subset of single-stranded cDNAs; the PCR amplification with the same and/or another oligo to enrich a subset of cDNAs; and, ultimately, display on a polyacrylamide gel, where differences between two different pools of RNA can be visualised, and from which specific differentially displayed fragments can be excised and sequenced. The methodology has proven to be a very powerful tool to isolate new genes and investigate differential gene expression, but it is typically hampered by the fact that it also generates a lot of false-positives.

Quantitative PCR is becoming more and more the preferred choice to prove differences in expression of known genes, but, in contrast to DD-PCR, it does not allow the identification of unknown genes, however, it is a more sensitive approach than Northern blotting and the RNase protection assay, and one can establish differences in expression levels of genes of which only a few mRNAs are present per cell. It is therefore expected that this technique will replace Northern blotting and the RNase protection assay in the future. The relative amounts of PCR product (relative to an internal control) can be compared. Currently, fluorescence-based kinetic (real-time) PCR allows the monitoring of amplification during a PCR, and more accurately reflects the relative expression levels of target genes.

In general, drawbacks of the conventional methods described above are that throughput is low and that gene expression analysis is semi-quantitative at best. A comprehensive overview of the advantages and drawbacks of these conventional technologies is given in Table 1.

**Genome-wide expression profiling technologies**

One way to establish an unbiased catalogue of expressed genes at the mRNA level is by simply sequencing cDNAs or expressed sequence tags (ESTs) derived from the tissue of interest. This approach has only proven to be useful in the qualitative analysis of gene expression, as it is very laborious. SAGE allows the serial analysis of short cDNA sequences or ‘tags’ derived from a defined position within a cDNA, and allows both qualitative and quantitative analysis. SAGE is based on the following principles: (1) a short sequence tag (10–14 base pairs) contains sufficient information to identify uniquely a transcript provided that the tag is obtained from a unique position within each transcript; (2) sequence tags can be linked together to form long serial molecules that can be cloned and sequenced; and (3) quantitation of the number of times a particular tag is observed provides the expression level of the corresponding transcript (Figure 2). SAGE has been used successfully to construct a transcriptome of yeast and a partial transcriptome of humans. It has also proven to be a useful tool in genomics, as a human transcriptome map has been constructed in an effort to identify clusters of genes on chromosomes of unusually high or low transcriptional activity in cancer. Regions of highly expressed genes (called regions of increased gene expression, or RIDGEs) have been identified in breast tissue and breast cancer tissue on chromosome 11. Furthermore, SAGE has been used successfully to identify differentially regulated genes in colon cancer cell lines and tumours, pancreatic carcinoma and p53-induced apoptosis, and has led to the elucidation of the role of several genes involved in the cell cycle and malignant transformation.

**Genome-wide expression profiling technologies allow the simultaneous analysis of thousands of genes**
Table 1: An overview of various gene expression profiling technologies

<table>
<thead>
<tr>
<th>Expression analysis technology</th>
<th>Amount of material required (ng)</th>
<th>Sensitivity(^\text{a})</th>
<th>Transcripts/genes covered (%)</th>
<th>Ability to discover new genes</th>
<th>Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern blotting</td>
<td>0.1–1 (mRNA)</td>
<td>High</td>
<td>&lt;1</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>RNase protection</td>
<td>1–100 (total RNA)</td>
<td>High</td>
<td>&lt;1</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Subtractive hybridisation</td>
<td>20–25 (mRNA)</td>
<td>Low</td>
<td>&lt;1</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>RDA/Subtractive hybridization</td>
<td>0.02 (mRNA)</td>
<td>High</td>
<td>&lt;10</td>
<td>Yes</td>
<td>Medium</td>
</tr>
<tr>
<td>Random EST sequencing</td>
<td>5–10 (mRNA)</td>
<td>Low</td>
<td>&lt;10</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>DD-PCR</td>
<td>0.5 (mRNA)</td>
<td>Medium</td>
<td>&lt;1</td>
<td>Yes</td>
<td>Medium</td>
</tr>
<tr>
<td>RAP-PCR</td>
<td>0.012 (total RNA)</td>
<td>Medium</td>
<td>&lt;1</td>
<td>Yes</td>
<td>Medium</td>
</tr>
<tr>
<td>Quantitative PCR</td>
<td>0.1 (total RNA)</td>
<td>High</td>
<td>&lt;1</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>SAGE (cDNA)</td>
<td>Varies with size of libraries,</td>
<td>Varies with size of libraries, from 30 to &gt;90</td>
<td>Yes</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>SAGE (MicroSAGE)</td>
<td>5–10 (mRNA: conventional SAGE)</td>
<td>Varies with size of libraries, from 30 to &gt;90</td>
<td>Yes</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>cDNA microarray analysis</td>
<td>&gt;10 (total RNA)</td>
<td>Medium</td>
<td>10–50, depending on number of oligos/cDNAs on array</td>
<td>No</td>
<td>High</td>
</tr>
<tr>
<td>Proteome analysis (including 2D-PAGE</td>
<td>&gt;0.001 (protein)</td>
<td>Low</td>
<td>&lt;1</td>
<td>Yes</td>
<td>Medium</td>
</tr>
<tr>
<td>microsequencing and/or MALDI-TOF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^\text{a}\) Sensitivity with regard to detection of rare transcripts/genes
The use of cDNA and oligonucleotide arrays, either filters or microslides (‘chips’), is rapidly becoming the de facto standard in large-scale differential gene expression analysis. This technology received a boost from the fact that the sequencing of the human genome has been finalised. It is estimated that the human genome encodes about 30,000 genes, most, if not all, of which have been cloned, either as a complete cDNA or as an EST. These cDNAs and ESTs as well as cDNA libraries containing unknown genes and oligonucleotides can be gridded onto filter arrays or microslides, which in turn can be hybridised with fluorescent cDNAs derived from different pools of RNAs, thus allowing the comparison and quantitative analysis of gene expression. Hybridisation is performed in a small volume in order to improve sensitivity. Relative fluorescence intensity is then measured and compared between the samples. When using oligonucleotides, the density of unique sequences per array can be increased dramatically, and technology has been developed that allows the nucleotide-for-nucleotide synthesis of oligos on a microarray by means of light-directed, spatially addressable parallel chemical synthesis. More importantly, with the development of array-based screening technologies, it is now possible to examine the expression levels of thousands of genes at a time as arrays, either as filters or microslides, can accommodate literally tens of thousands of cDNAs or oligonucleotides. Although at this point in time SAGE is the only technology that in theory allows the generation of a complete gene expression catalogue for a given sample, it will be possible to achieve the same with
microarrays as all genes will have been identified in the near future. A clear advantage of the use of microarrays is that they do not require extensive sequencing, as opposed to SAGE, and they, therefore, represent a truly high throughput method.

When carrying out screens for differential gene expression on the level of mRNA, one should be aware that mRNA expression levels do not necessarily reflect those of the corresponding proteins. When protein expression levels were investigated in *Saccharomyces cerevisiae* growing in mid-log phase, and compared with SAGE data that had been obtained in the construction of the yeast transcriptome, it became apparent that the protein expression levels of some genes either did not vary while the expression level of their corresponding mRNAs varied more than 30-fold or, in some other cases, the protein expression levels varied more than 20-fold while the expression levels of the corresponding mRNAs were invariant. It may therefore be more appropriate to determine the expression profiles at the protein level as a way to obtain a broad view of gene expression. Only recently have high throughput protocols become available, and the field of proteomics is rapidly developing. Large-scale two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS) and various chromatography technologies now allow the simultaneous analysis of many proteins, and it is expected that these will only improve over time.

The term ‘transcriptomics’ is widely used to describe the research that deals with the establishment and analysis of large datasets on the level of mRNA expression in a well-defined tissue or cell type. Initial attempts to analyse mRNA expression in keratinocytes on a wide scale involved random sequencing of cDNA libraries, revealing the expression of at least 607 genes in human cultured keratinocytes, half of which did not have a match in the GenBank databases at that time. By current standards, this study did not address many genes, but could at that time still be considered a major effort. An attempt to analyse a large amount of expressed mRNAs in cultured keratinocytes and epidermis was undertaken by the authors’ group, by means of SAGE. SAGE was applied to keratinocytes in a culture model that allows normal or aberrant differentiation, and the response of these differentiating keratinocytes to tumour necrosis factor-alpha (TNF-α), an important inflammatory stimulus, was investigated. The seminal work of Celis et al. involved the use of large-scale 2D-PAGE and microsequencing in order to gather insightful information regarding expression patterns, signalling pathways and components of normal and diseased epidermis. In these studies, over 800 keratinocyte-derived proteins were identified. Allowing for redundancy and isoforms, these protein products appear to correspond to approximately 150 genes as determined by analysis of the UniGene database (see also ref. 66). Using this approach, they were able to identify the various forms in which gene products occur in the cell, but far more peptides need to be isolated and sequenced before a definitive catalogue of expressed proteins in keratinocytes can be made. Nevertheless, the achievement of the Celis group is remarkable, and their work has led to the identification of many keratinocyte-specific genes. Katz and Taichman have also produced a catalogue, albeit small, of proteins excreted by keratinocytes into the culture medium.

The first effort of large-scale gene expression analysis of keratinocytes was performed at the protein level.
investigated by comparing the gene expression profiles of resting and TNF-α stimulated keratinocytes. More than 25,000 tags were sequenced, possibly representing as much as 12,000 genes, and it became apparent that cultured keratinocytes put an enormous effort into the expression of genes that are involved in barrier formation and the cytoskeleton. A more surprising finding was that many genes, which are (putatively) involved in anti-microbial host defence, are expressed at high levels in keratinocytes, and that some of these are regulated by TNF-α. These findings were well in line with the idea that, given the environment where these cells normally occur in the body, keratinocytes play an important role in the protection of the organism. Furthermore, only one per cent of expressed genes appeared to be differentially regulated. This is well in line with other reports, where small changes led to huge differences in phenotype.

Although culture models have been very useful in studying cutaneous biology, the relevance of applying SAGE to normal and diseased epidermis is obvious, as it leads to immediate information that is of direct relevance to the disease. In an attempt to analyse gene expression in human epidermis, we have compared the expression profiles of a premalignant tumour, actinic keratosis, with unaffected epidermis and healthy epidermis. Therefore, we first had to isolate epidermis from skin biopsies, which usually also contain (confounding) dermal material. To that end, we used a protocol which allows the rapid separation of epidermis and dermis at 4°C by dispase, a proteolytic enzyme. This procedure allows recovery of intact mRNA and minimises degradation and induction of mRNA synthesis. The SAGE protocol was modified so that it can accommodate in principle less than 100 ng of mRNA, as opposed to >2.5 µg of mRNA for the conventional protocol. In this modified MicroSAGE protocol, SAGE is essentially carried out in one streptavidin-coated tube per linker, thus eliminating phenol extraction and precipitation steps, and thereby increasing yields and speed. The three resulting libraries and the two libraries derived from cultured keratinocytes were analysed by means of two-way cluster analysis after data reduction by extracting the most informative tags. The 75,000 sequenced tags of all libraries combined represented more than 15,000 genes, and cluster analysis revealed that there were clear partitions and correlations between libraries. Furthermore, an automated literature search with clusters of genes using PubGene showed that a subgroup of these genes is known to be co-expressed in other tissues, and is part of an epidermal differentiation gene cluster on chromosome 1q21. PubGene was developed by automated extraction of explicit and implicit biomedical knowledge from publicly available gene and text databases to create a gene-to-gene co-citation network for 13,712 named human genes by the automated analysis of titles and abstracts in over 10 million MEDLINE records. The analysis of the partial transcriptome of human epidermis revealed that human epidermis expresses a large number of genes that were previously unknown to be expressed in human epidermis. Furthermore, it was found that human epidermis, in contrast to cultured keratinocytes, expresses genes that are associated with epidermal differentiation at a lower level, but housekeeping genes at a higher level. The partial transcriptome harbours a wealth of information to identify genes involved in skin function, and candidate genes for genetic skin disorders.

Others have analysed gene expression in keratinocytes by means of filter and cDNA array hybridisation. Although Trenkle et al., one of the first groups to report the use of high-density cDNA arrays to analyse keratinocyte gene expression, described changes in gene expression of HaCaT keratinocytes in response to epidermal growth factor, they primarily described a refinement of the
way to generate probes for array analysis. Most of the studies using high-density microarrays centre around carcinogenesis, both de novo and by transforming human papillomaviruses (HPVs) and their associated oncogenes, and gene expression induction by UV irradiation. Nees et al.\textsuperscript{76} investigated the expression profiles of approximately 2,000 genes in cervical, differentiating keratinocytes after transformation with HPV-16 E6 and E7 oncogenes, and found that the expression of approximately 80 genes was altered reproducibly. Cluster analysis revealed that interferon-responsive genes were downregulated by E6 and even more when E6 and E7 were co-expressed, whereas genes involved in the nuclear factor kappa-beta (NF-\textkappa B) activating pathway and genes regulated in cell cycle progression and DNA synthesis were upregulated. Earlier, it had been shown that infection of keratinocytes with another subtype of HPV, HPV-31, also downregulated interferon-responsive genes, and suppression of these genes may contribute to immune evasion.\textsuperscript{77} NF-\textkappa B is known to induce cellular survival pathways that may make tumour cells resistant to apoptotic stimuli, and a deregulated cell cycle is characteristic of tumour cells. The results show the power of cDNA microarray analysis as cDNAs from genes involved in different, well-defined pathways can be analysed simultaneously. Furthermore, independent analyses of different laboratories show clear correlations not only in gene expression, but also between different subtypes of HPV.

Exposure of human epidermis to UVB radiation (290–320 nm) is known to induce pre-malignant and malignant skin tumours such as actinic keratosis, squamous cell carcinoma and basal cell carcinoma. Various groups have analysed the effects of UVB radiation on human keratinocytes, using either filter arrays containing over 50,000 cDNAs\textsuperscript{78} or oligonucleotide arrays covering more than 6,000 genes.\textsuperscript{79,80} In general, the response of cultured keratinocytes to UV irradiation involves the switch of the phenotype of proliferation to a phenotype of cell cycle arrest and enhanced differentiation. Li et al.\textsuperscript{80} investigated gene expression in a submerged, serum-free culture model without feeder layer at different time intervals after UV irradiation, and could thereby clearly separate waves of gene expression. In the first wave within two hours after UV exposure, genes that induce an activated phenotype and cell cycle arrest are upregulated, and the expression of pro-apoptotic genes is on the rise; however, there is also an induction of anti-apoptotic genes, and the balance between pro- and anti-apoptotic genes seems to shift to the latter as otherwise normal cells survive. In the second wave, approximately four to six hours after exposure, keratinocytes start to produce chemokines, cytokines and growth factors, alerting the surrounding cells of the UV damage. In the third wave, 16–24 hours after UV exposure, cells start their accelerated differentiation, thereby removing a carcinogenic threat. Also, mitochondrial genes that are involved in energy metabolism and nuclear genes that are involved in DNA repair seem to be generally upregulated. Sesto et al.\textsuperscript{79} used a culture model in which keratinocytes were grown on a feeder layer, and exposed their cells to three different doses ranging from 10 mJ/cm\textsuperscript{2} to 40 mJ/cm\textsuperscript{2}, while analysing gene expression four and 24 hours after irradiation. They were able to identify clusters of genes that were only regulated with high doses of UV radiation, and they also found clusters of genes that were co-regulated, ie that, for example, were all downregulated at four hours after stimulation, but were upregulated at 24 hours. Only one cluster of 24 genes showed a dose-dependent sustained upregulation after UVB exposure. Interestingly, the data obtained with microarray analysis suggest that a second event, such as a mutation in a tumour suppressor gene, has to take place before a cell commits to malignant transformation, because irradiated healthy
cells favour survival and differentiation, or rather defence, instead of transformation. Furthermore, despite differences in the culture models used, a general concept of gene expression induced by UV radiation can be extracted from the data of the various laboratories.

In vivo gene expression analysis with oligonucleotide arrays has also been applied to skin biopsies from patients suffering from psoriasis, a chronic autoimmune skin disease characterised by a leukocyte infiltrate and keratinocyte hyperproliferation.\textsuperscript{81,82} It is believed that both genetic and environmental factors are involved in this disease. In addition to a major contribution from the human leukocyte antigen (HLA) class I region, psoriasis susceptibility loci have been mapped to a number of chromosomal regions including 1q21, 3q21, 4qter, 14q31-q32, 17q24-q25, 19p13.3 and 20p. Some of these overlap with loci implicated in other autoimmune and inflammatory diseases. Global gene expression analysis on more than 12,000 genes clearly separated affected skin from unaffected and healthy biopsies after hierarchical clustering. Of 32 clusters of differentially expressed genes with similar expression patterns in the same samples, six contained a total of 177 transcripts that were differentially expressed among psoriatic and normal skin, and some of these were also differentially regulated in unaffected skin from patients. Although clear genetic partitions and correlations could be demonstrated, there was still considerable heterogeneity between patient samples, indicative of the multifactorial nature of psoriasis. Array analysis is ideally suited for the investigation of therapeutic strategies and the effect of drugs and irritants on gene expression. Genome-wide expression analysis has been used to identify 159 genes that make up the molecular signature of psoriasis, and the effects of drugs that antagonise NF-κB and calcineurin were investigated. Only the expression of a subset of these genes appeared to be changed after treatment and many, but not all, patients responded favourably. As pointed out before, the multifactorial nature of psoriasis and genetic diversity among patients makes the interpretation of the data difficult. Pharmacogenetics is, therefore, also focused on defining the genetic variability among patients with regard to disease and drug responses, and involves both high throughput genotyping and microarray analysis.\textsuperscript{83}

Microarray analysis has also been used to investigate the effect of gene therapy versus protein therapy for the genetic skin disease junctional epidermolysis bullosa (JEB), which is caused by mutations in the laminin-5β3 gene. The analysis of gene expression profiles of cultured JEB cells that had either been treated with wild-type protein or a genetic construct expressing the wild-type protein showed that therapies that are aimed at establishing the same effect may induce fundamentally different responses.\textsuperscript{84} Clontech Atlas arrays have been used to investigate the response of 3,600 genes in the EpiDerm skin culture model to the detergent and skin irritant sodium lauryl sulphate (SLS). From the generated data, different waves of gene expression could be identified that are associated with ion transport (within the first hour), wound healing (after one to three hours) and metabolism and protein translocation (after four to 24 hours). This study can be used to develop more robust and relevant in vitro methods for predicting the effect of drug treatments, therapy and skin irritancy.

**FUTURE CONSIDERATIONS**

The technology of the future: the microarray has the edge

When analysing expression patterns on the protein level, various isoforms and modifications of proteins can be found, and these may be physiologically very relevant; however, proteins differ significantly in stability and modifications, which sometimes hampers isolation and analysis. Furthermore, highly abundant
proteins may blur the role of proteins that are expressed to a lesser extent but hold considerable importance to the observed phenotype. Isolation procedures may further shift the balance to proteins that benefit from the extraction procedures used. Still, using various protocols to isolate proteins may, at least in part, overcome these problems. An advantage of the proteomic approach is, clearly, that it may uncover various modifications of the same gene product, which is clearly impossible to obtain from mRNA expression data.

When analysing gene expression on the RNA level, other considerations come into the picture. As noted earlier, array analysis usually involves genes that have already been characterised and does therefore not allow the discovery of new genes that may be crucial to an observed phenotype. Clearly, SAGE has an advantage here, as it not only allows the identification of new genes but, since it is not bound by pre-defined standards, in principle allows the analysis of literally tens of thousands of transcripts and thereby genes. There are problems though: SAGE relies heavily on single-pass sequences, and it is estimated that approximately five per cent of all sequenced tags arise due to sequencing errors, and are therefore irrelevant. Another drawback is that many mRNAs contain common repeats in their 3’ untranslated regions (3’ UTRs) that are dispersed throughout the human genome and result in non-uniqueness for at least a part of the identified tags. Furthermore, differential splicing or polyadenylation may influence the result significantly, as important tags may be mis-assigned or ignored. SAGE is also laborious as it requires an extensive sequence effort. When designing a microarray, one can simply ignore these problems, although one has to take into account the fact that strong homology of paralogous genes may falsely lead to under- or overestimation of the role of such genes. Although at this point in time SAGE still allows for the detection of unknown genes, the rapid pace at which the human genome is being unravelled will make the microarray the preferred tool to investigate differential gene expression. It is to be expected that the sensitivity of microarrays analysis for rare transcripts will only increase with improved protocols.

An often overlooked but more general problem in transcriptomics is sampling error. Biopsy material may contain traces of surrounding tissue or infiltrate, and should never be considered to be pure. This can be overcome by the dissection of tissue by means of laser capture, which allows the isolation of even single cells from the tissue of interest, this, however, poses another problem: from only a few cells, one can only isolate very little RNA. In these cases, it might be more convenient to use RNA amplification technologies such as the Switch Mechanism At the 5’ end of RNA Templates (SMART) and variants thereof or T7 RNA polymerase-based technologies in order to boost the input before starting the analysis. There is also evidence that genetic heterogeneity in the population may significantly influence gene expression and, thus, expression analysis. Furthermore, there is evidence that input material may influence the result, as two-way cluster analysis on microarray expression data from colon cancer and leukaemia could separate those that were obtained with isolated total RNA from those that were obtained with purified mRNA. Differences in array design (complete cDNA versus oligonucleotides) may also influence data, as oligonucleotides can be designed such that hybridisation can be carried out over a short temperature range. The use of arrays may also bypass possible problems with common repeats — from which SAGE generally suffers — as these repeats can be left out.

The quality of microarray analysis will only improve

The proteomic approach may uncover variations in the same gene product that cannot be found with the transcriptomic approach.

Regardless of profiling technology, sampling errors affect gene expression profiles.

SAGE allows the discovery of new genes, but is very laborious and is hampered by non-uniqueness and differential mRNA processing.

The complexity of data analysis

Genomic technologies generate new types of complex data, and the tools to handle and isolate appropriate subsets are rapidly being developed. Bioinformatics
and statistical analyses are of tremendous importance in the analysis of the wealth of generated expression data. Raw sequences generated in the SAGE procedure can be analysed with the SAGE2000 software package, but, when comparing multiple libraries, cluster analysis is preferred, provided that datasets are compressed such that only the most informative, expressed genes are clustered. Two-way or hierarchical clustering methods allow not only the clustering of co-regulated genes but also the discovery of correlations or relatedness between different datasets, which in turn may uncover previously unnoticed correlations and partitions. There are studies that have uncovered problems with hierarchical clustering: it may suffer from a lack of robustness and non-uniqueness, and from inversion problems that complicate the interpretation of the hierarchy. Another interesting way is the use of self-organising maps (SOMs; an unsupervised neural network algorithm). SOMs allow the easy visualisation of complex data and are robust in the face of minor experimental variation, which is not necessarily the case with hierarchical clustering methods. Several software packages have been applied that use the SOM algorithm and have successfully been applied in gene expression analysis.

Bioinformatics is now used to integrate medical knowledge with gene expression data. The Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information is a public repository of gene expression data, and should be considered as a central data hub where gene expression data can be deposited and retrieved for local analysis. The public availability of these data allows the comparison of results of different laboratories, provided that samples have been obtained and analysed in the same way, ie have been obtained on the same platform. Currently, this poses a problem as different laboratories generate different data as a result of different samplings methods and ways of generating gene expression profiles. The use of more robust statistical algorithms may overcome these problems in the long run, and SOM is clearly a step in the right direction. In fact, such algorithms may prove valuable in trying to find clear correlations between samples that do not share the same platform, ie to find genes of which the expression is a prerequisite for the phenotype investigated, regardless of platform and sampling method. Hopefully, the comparison of various gene expression patterns from normal and diseased tissue will lead to the identification of (groups of) transcripts that are truly linked to disease and response to therapy, and may aid in our understanding of disease and the development of therapeutic strategies.

References

Transcriptomics and proteomics of human skin


Transcriptomics and proteomics of human skin


66. URL: http://proteomics.cancer.dk/


