Universiteit Leiden
The Netherlands

Molecular mechanisms of epithelial host defense in the airways

Vos, J.B.

Citation
Vos, J. B. (2007, January 11). *Molecular mechanisms of epithelial host defense in the airways*. Retrieved from https://hdl.handle.net/1887/9749

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/9749

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 2

EXPLORING HOST-PATHOGEN INTERACTIONS AT THE EPITHELIAL SURFACE: APPLICATION OF TRANSCRIPTOMICS IN LUNG BIOLOGY

Joost B. Vos [1], Nicole A. Datson [2], Klaus F. Rabe [1], Pieter S. Hiemstra [1]

[1] Department of Pulmonology, Leiden University Medical Center, Leiden, The Netherlands

[2] Department of Medical Pharmacology, Leiden Amsterdam Center for Drug Research, Leiden University Medical Center, Leiden, The Netherlands

American Journal of Physiology: Lung Cellular and Molecular Physiology, in press

ABSTRACT

The epithelial surface of the airways is the largest barrier-forming interface between the human body and the outside world. It is now well recognized that, at this strategic position, airway epithelial cells play an eminent role in host defense by recognizing and responding to microbial exposure. Conversely, inhaled microorganisms also respond to contact with epithelial cells. Our understanding of this cross-talk is limited, requiring sophisticated experimental approaches to analyze these complex interactions. High-throughput technologies such as DNA microarray analysis and SAGE have been developed to screen for gene expression levels at large scale within single experiments. Since their introduction, these hypothesis-generating technologies have been widely used in diverse areas such as oncology and brain research. Successful application of these genomics-based technologies has also revealed novel insights in host-pathogen interactions in both the host and pathogen. This review aims to provide an overview of the SAGE and microarray technology illustrated by their application in the analysis of host-pathogen interactions. In particular, the interactions between epithelial cells in the human lungs and clinically relevant microorganisms are the central focus of this review.

INTRODUCTION

Efficient host defense is essential for survival. The human body is equipped with an elaborate defense system comprised of the innate and adaptive immune system. The vast majority of threats are effectively eliminated by the immune system before they can cause injury [1,2]. By forming the primary physical barrier between the body's internal and external world, epithelial tissues are indispensable for host defense. In addition, epithelial cells contribute to the chemical barrier by producing a wide range of host defense effector molecules [1,3]. The effectiveness of epithelial host defense mechanisms is demonstrated by the rare incidence of severe infections at epithelial surfaces in the lungs of healthy individuals. The mechanisms underlying host defense and epithelial repair critically depend on the presence of functional proteins at appropriate quantities within a crucial time window. These proteins are encoded by genes whose transcription

Table 1: Overview of gene expression analysis technologies. A selection of gene expression profiling methods that have been introduced during the last three decades is listed. For each technology, the year of introduction, the type of profiling approach and profiling potential is provided.

Technology	Year of introduction	Approach	# genes analyzed	Advantages	Limitations
Northern blot	1977 [5]	Closed	1	Highly specific, detection of low abundant transcripts	Low throughput, less sensitive than RT-PCR
Subtractive hybridization	1983 [71]	Open	10-100	Positive selection of differentially expressed genes, gene discovery	Low throughput, Low detection limit
Polymerase Chain Reaction	1983 [72]	Closed	1	Fast, highly sensitive	Low throughput
Differential Display	1992 [73]	Open	>100s	Fast, gene discovery	High rate of false positives
Expressed Sequence Tag Sequencing	1993 [7]	Open	>100s-1000	Highly specific, gene discovery	Costly, time consuming, limited number of samples
Serial Analysis of Gene Expression	1995 [8]	Open	>1000s	Highly specific, gene discovery, efficient sequencing strategy	Limited sample number, technically demanding bioinformatics
DNA microarray	1995 [6]	Closed	>1000s	Ease of use, large number of samples	Bioinformatics
Massively Parallel Signature Sequencing	2000 [9]	Open	>1000s	Highly specific and accurate. Ability of gene discovery	Costly, limited number of samples, Bioinformatics

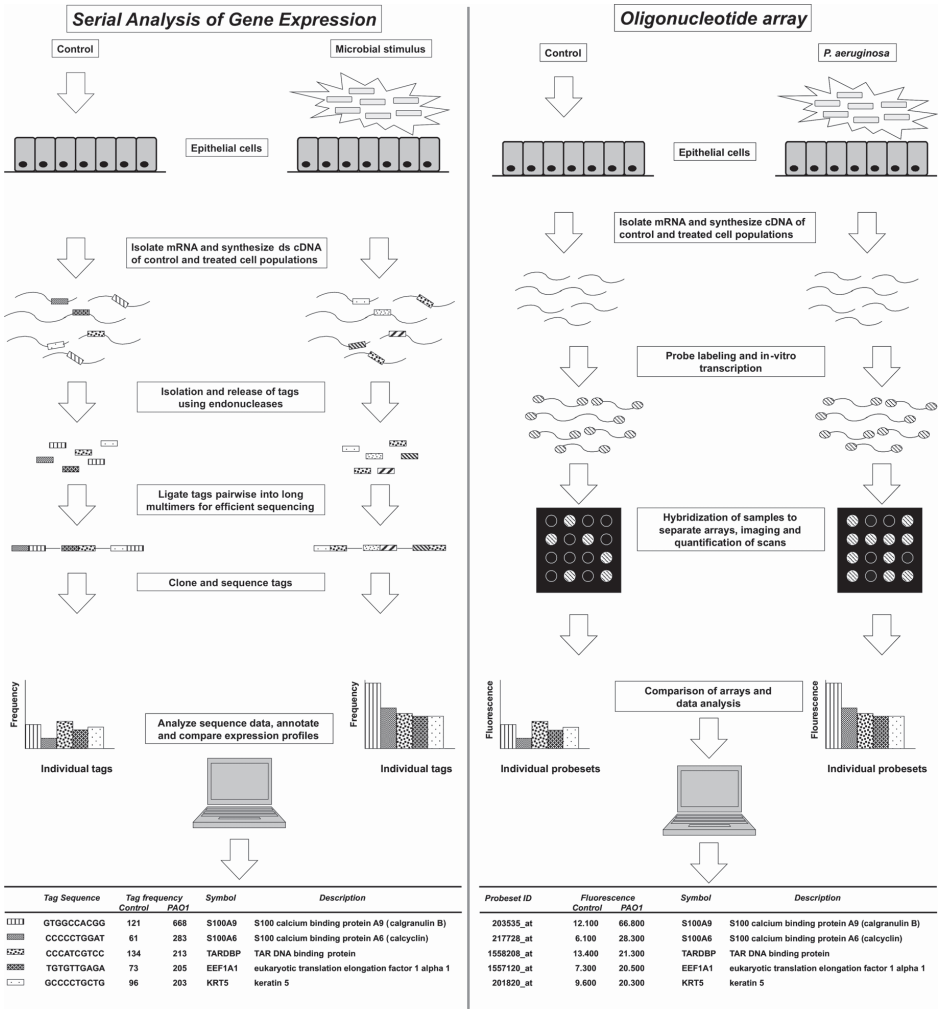
is tightly coordinated by complex programs of gene expression. Altered or defective gene regulation may not only increase the susceptibility of the host to infections, but may also be involved in the development of atopic disease [4].

Evaluating gene expression has greatly extended our understanding of behavior and function of cells and tissues under varying conditions. A diversity of technologies has been developed to assess levels of gene expression, ranging from analysis of single genes (*i.e.* Northern blot and PCR) to thousands of genes simultaneously (*i.e.* SAGE and microarray; ***Table 1***). Based on experimental approach, a distinction can be made between *"closed"* and *"open"* profiling methods. Closed approaches rely on hybridization of genes of interest to complementary nucleic acids and, therefore, genome knowledge is a prerequisite. Examples of closed approaches are Northern blot [5] and DNA microarrays [6]. In contrast, open approaches do not depend on genome knowledge since they are based on sequencing of poly(A)+ messenger RNA (mRNA) molecules expressed in the cell or tissue of interest. Examples of open approaches are Expressed Sequence Tag (EST) sequencing [7], SAGE [8] and Massively Parallel Signature Sequencing (MPSS) [9]. The identity of transcripts is determined by matching the experimental sequence to available genomic data.

The most widely applied high-throughput strategies are SAGE [8] and DNA microarray technology [6]. These hypothesis-generating technologies are highly suitable to study host-pathogen interactions and are being increasingly applied in hypothesis-driven pulmonary research [10-13]. So far, interactions between airway epithelial cells and respiratory pathogens such as *Staphylococcus aureus*, *Pseudomonas aeruginosa* and *Bordetella pertussis* have been studied with high throughput expression profiling techniques. The introduction of large-scale gene expression profiling in host-pathogen research has not only provided novel insights into how epithelial cells cope with potential hazards of infection, but has also demonstrated how microbes respond to the host. This review will focus on the application of the SAGE and microarray technologies in pulmonary research, in particular to investigate the molecular mechanisms underlying host-pathogen interactions in the lung.

SERIAL ANALYSIS OF GENE EXPRESSION

SAGE is an open high-throughput expression profiling technique that allows unbiased assessment of virtually all polyadenylated transcripts in a single sample [8,14]. The outline of the technology is presented in ***Figure 1***. Shortly, double stranded cDNA is synthesized from mRNA molecules that are biochemically purified by the poly(A)+ tail. In a series of two endonuclease reactions, representative short nucleotide sequences of 10-14 base pairs (bp) are isolated. In the first endonuclease step, the restriction enzyme *Nla*III is



Tag Sequence	Tag frequency Control	Tag frequency PAO1	Symbol	Description
GTGGCCACGG	121	668	S100A9	S100 calcium binding protein A9 (calgranulin B)
CCCCCTGGAT	61	283	S100A6	S100 calcium binding protein A6 (calcyclin)
CCCATCGTCC	134	213	TARDBP	TAR DNA binding protein
TGTGTTGAGA	73	205	EEF1A1	eukaryotic translation elongation factor 1 alpha 1
GCCCCTGCTG	96	203	KRT5	keratin 5

Probeset ID	Fluorescence Control	Fluorescence PAO1	Symbol	Description
203535_at	12.100	66.800	S100A9	S100 calcium binding protein A9 (calgranulin B)
217728_at	6.100	28.300	S100A6	S100 calcium binding protein A6 (calcyclin)
1558208_at	13.400	21.300	TARDBP	TAR DNA binding protein
1557120_at	7.300	20.500	EEF1A1	eukaryotic translation elongation factor 1 alpha 1
201820_at	9.600	20.300	KRT5	keratin 5

Figure 1: Outline of the SAGE and microarray technology. A flow-chart of the SAGE and microarray technologies is presented from top to bottom. In the first step, cultured cells are exposed to a pathogen. Isolated RNA is processed as such required for SAGE or microarray technology. Collected data reveals the expression profiles in the cells of interest under the studied conditions as illustrated in the lower table.

used to digest double-stranded cDNAs at every four base CATG sequence. Since each cDNA is immobilized at the poly(A)+ tail, only the most 3' fragments are captured for further processing. Before the second endonuclease step, a linker sequence containing the recognition site for the second restriction enzyme, *Bsm*F1, is attached to the generated four base CATG overhang. The type IIs endonuclease *Bsm*F1 specifically cleaves at 10-14 bp distance from its recognition site, thereby releasing the 10-14 bp fragments from the 3'-end immobilized cDNA fragments. Although relatively short (10-14 bp), these so-called SAGE tags contain sufficient genetic information to uniquely identify individual transcripts since they are derived from a defined position within each indi-

vidual transcript: directly adjacent to the most 3' recognition site of *Nla*III. Serial ligation of multiple SAGE tags into long multimers not only allows qualitative and quantitative determination of SAGE tags, but also drastically increases sequencing efficiency by limiting the number of required sequencing reactions. The identity of SAGE tags representing known genes is then established by comparing its nucleotide sequence to available gene sequences deposited in genetic databases. Since SAGE libraries are generated by random sampling of transcripts, expression profiles will become more accurate when analyzing larger numbers of tags. When sufficient SAGE tags are analyzed, the sensitivity of SAGE appears to be comparable to DNA microarrays in estimating expression levels of in particular medium to high abundant genes [15-17].

An attractive feature of SAGE is that the technology allows gene discovery. Despite the completion of sequencing the human genome, annotation of the genes within the genome is still very much an ongoing process. Our knowledge of the repertoire of human genes is still incomplete and novel genes are discovered on a frequent basis. The SAGE platform is excellent for conducting comparative expression studies because SAGE yields digital sequence-based data. Libraries can be compared with one another without complex mathematical normalization methods, even if these libraries were generated in different laboratories [18]. A drawback of SAGE is the laborious nature of the technology. On average, it takes a few weeks to generate and sequence a single library. Consequently, the number of samples that can be realistically studied is limited. Because of the laborious nature of SAGE, researchers often prefer the microarray technology.

Interestingly, recently the application of SAGE has expanded from expression analysis to include whole genome analysis. Combining the specificity of chromatin-immunoprecipitation (ChIP) with the sensitivity of SAGE allows to identify genome signature tags defining functional genomic elements and transcribed regions [19] such as transcription factor binding sites [20] and regions with hyperacetylated histone proteins [21]. The combined approach of ChIP and SAGE was first introduced as Serial Analysis of Chromatin Occupancy (SACO; [20]), and is also known as Genome-wide Mapping Technique (GMAT; [21]), Sequence Tag Analysis of Genomic Enrichment (STAGE; [22]) and Serial Analysis of Binding Elements (SABE; [23]). Essential to application of novel technologies is the availability of analysis software. Recently, the first web-based software tool has been introduced that facilitates the analysis of this data [24]. Although the combination of ChIP and microarray, called the "ChIP-on-chip" is available, this method has major drawbacks. Most importantly, a substantial number of promoter and other regulatory regions within in the human genome have not been well characterized and are therefore not represented on the microarray. In addition, binding sites for transcription factors may well be located outside the transcription initiation sites of genes [20]. Perhaps not surprisingly, the genome-wide and open, unbiased nature of SAGE offers clear advantages over the microarray, both in terms of efficiency and precision of identifying (unknown) regulatory regions within

the genome. Attractive applications of ChIP and SAGE can be foreseen in studying the regulation of the response of airway epithelial cells to pathogens. Transcription factors associated with the regulation of expression of pro-inflammatory genes involved in the response of airway epithelial cells to microorganisms include NF-κB and AP-1. The use of antibodies directed against these transcription factors for chromatin immunoprecipitation would allow the identification of regulatory sequences employed by these transcription factors in the response of airway epithelial cells to microorganisms. This information could provide more integrated insights on how the transcriptional changes in airway epithelial cells exposed to microbial stimuli are regulated.

DNA MICROARRAYS

The microarray technology is a closed high-throughput method that enables the measurement of a large, predetermined set of known genes or sequences [6]. In microarray technology, DNA molecules representing specific transcripts are fixed onto a solid support, ranging from oligonucleotides (25-70 mers) to complete cDNAs. Inherent to closed approaches like microarray, a finite collection of arrayed sequences can be analyzed. However, microarrays are available that contain approximately 45.000 probe sets covering all known human genes as well as thousands of undefined expressed sequence tags (ESTs).

To visualize gene expression on microarrays, samples are labeled with a fluorescent dye prior to hybridization and fluorescence intensity is quantified as a measure for gene expression in the original sample. The outline of the labeling and hybridization steps for the single-color microarray technology is depicted in ***Figure 1***. For the single-color microarray, each sample is hybridized to a separate microarray. Prior to hybridization, samples are biotin labeled and stained with a streptavidin-bound fluorophore (*i.e.* phycoerythrin) and visualized by confocal laser microscopy. The single-color approach is common to commercial oligonucleotide arrays. Advantages of oligonucleotide microarrays are that *(i)* repetitive sequences within the genome can be circumvented because of the use of short and uniquely designed probes; *(ii)* probes have more uniform hybridization efficiencies; and *(iii)* standardized protocols and equipment are used providing consistent and reproducible data generation [25]. Because of unique probe design and uniform hybridization efficiencies, oligonucleotide arrays have a larger dynamic range of detecting gene expression [26]. A disadvantage of oligonucleotide arrays is that the production procedure is costly and relatively inflexible.

An alternative approach for single color microarray is the two-color microarray. The two-color approach is mostly used for cDNA microarrays, but can also be applied to oligonucleotide arrays. For cDNA arrays, long and double-stranded cDNA probes are

fixed onto this type of microarray. For hybridizing cDNA microarrays, samples are each labeled with a distinct fluorescent dye (*i.e.* Cy3 and Cy5) and co-hybridized to the same microarray. Binding of transcripts from both samples is detected using confocal laser microscopy by scanning the chip for the two fluorescent channels separately. Expression levels of genes within the two co-hybridized samples can be directly compared. An advantage of this approach is that two samples can be applied at once to a single microarray. However, a disadvantage of the two-color approach is the need to perform dye-swap experiments and mathematical signal normalization strategies to control for varying emission intensities of the Cy3 and Cy5 dyes (reviewed in [25]).

Since a substantial number of academic organizations invested in their own arraying facilities, the cDNA microarray is frequently used. When equipment and probe collections are available, "in house-made" cDNA are more flexible and cost effective as compared to the commercially available oligonucleotide microarrays. However, cDNA microarrays have a smaller dynamic range because of a less efficient signal to noise ratio [26]. This less efficient signal to noise ratio is partly due to the fact that the arrayed cDNAs probes are lengthy and double-stranded, increasing the likelihood of non-specific and cross-hybridization to related sequences [27]. In addition, the density of arrayed cDNA probes is generally lower compared to oligonucleotide microarrays. Despite these differences, cDNA and oligonucleotide microarrays perform equally with respect to detection of abundantly expressed genes [27].

A major advantage of DNA microarrays is the commercialization of the most labor-intensive parts of the methodology: collecting sequences (synthesized oligonucleotides or cDNA clones) and array fabrication. Commercial manufacturers of microarrays provide extensive and well-documented annotation of probe sets which eases data mining and interpretation, whereas annotation of SAGE tags is not straightforward (see paragraph on annotation). Nowadays, ready-to-use microarrays are available for many different organisms which makes the use of this technology possible without the need of having microarray fabrication equipment and own cDNA clone collections. Collecting data by using pre-fabricated microarrays typically takes less than one week. Therefore this high-throughput profiling technology is often the preferred choice of many scientists. However, in contrast to SAGE comparing DNA microarray results between experiments and between laboratories is hampered by differences in the type of array used (single versus dual color; oligo vs. cDNA; homemade vs. commercial), the spotted probe sequences and the lack of standardized experimental procedures. Not all fluorescent labels perform equally and different probe sequences used by the various manufacturers representing the same gene may give rise to varying hybridization efficiencies. To facilitate comparative analysis on microarray data, the Microarray Gene Expression Data Group (MGED; http://www.mged.org) proposed a uniform annotation format for microarray experiments (MIAME) [28]. Although complying with the same standardized

annotation format, complex computational normalization methods are still required to conduct comparative research based on microarray experiments.

DATA MINING AND INTERPRETATION

The most challenging part of the analysis of SAGE and microarray experiments is to assign biological significance to the observed findings and to finally formulate new hypotheses. Although high-throughput expression profiling technologies are very efficient in producing sizeable quantities of data, simple and standardized computational methods for data analysis and mining are not easily accessible. The need for appropriate tools is becoming increasingly important in successful expression profiling experiments since data analysis is often hampered by the lack of knowledge on available methods and data mining resources. Recent efforts have greatly improved and facilitated the application of statistical and functional data analysis methods. However, investigators should be aware of both the experimental difficulties as well as issues related to statistical and functional data analysis since these factors can influence the final results.

ANNOTATION

Prior to data mining and interpretation, accurate annotation of experimental data is essential. For most commercially available microarrays, appropriate annotation of probe sets including links to reference databases such as Gene Ontology is generally supplied by the manufacturer. In contrast, accurate annotation of SAGE tags is a tedious task that is not easily accomplished. A short guide on how to annotate SAGE data is provided here since this is a specific SAGE issue. For tag identification, SAGE libraries can be matched to NCBI's reliable Unigene cluster to SAGE tag map (ftp://ftp.ncbi.nlm.nih.gov/pub/sage) [29] or to CGAP's SAGEgenie (http://cgap.nci.nih.gov/SAGE) [30]. For well-known genes, both mapping strategies yield the same results. However, when tags match to poorly-described transcripts, the NCBI and SAGEgenie maps yield different outcomes. Therefore, best tag identification is achieved by combining the two tag mapping strategies [31]. Besides poorly-described transcripts, SAGE libraries are contaminated with tags isolated from non-preferred restriction sites within transcripts. Typically ~30% of the unique sequenced tags are derived from non-preferred 3'-end positions (unpublished observations). Discarding these non-representative tags not only facilitates experimental validation of SAGE data [18] by decreasing the number of false positives but also eases subsequent data mining and interpretation. SAGEgenie allows pre-screening of SAGE libraries for the presence of known non-representative tags. The presence of single

nucleotide polymorphisms (SNP) may also give rise to alternative anchoring sites for the restriction enzyme to isolate SAGE tags [32], a feature addressed in the recent versions of SAGEgenie. Once SAGE libraries have been annotated and non-representative tags have been discarded, association to additional data sources (Genbank, Gene [formerly known as Locuslink], Online Mendelian Inheritance Man (OMIM), Gene Ontology or any other useful biological data source) can be established to facilitate the biological interpretation of the data.

MATHEMATICAL CLUSTERING

Mathematical clustering is a useful and powerful next step to structure expression data in an unbiased fashion. Clustering methods are based on the assumption that genes showing similar transcriptional behavior across samples might correspond to or be involved in the same biological process. Since clustered data is more easily accessible, it is a recommended analysis method for SAGE and microarray experiments. Clustering methods can be divided into supervised and unsupervised methods. Supervised or knowledge-assisted clustering algorithms are guided by existing biological knowledge about specific subsets of genes (*i.e.* signal transduction cascades, cell cycle, metabolic pathways) [33]. In contrast, unsupervised clustering methods allow complete unbiased structuring of gene expression data. To date, unsupervised clustering methods have been most widely applied to large-scale gene expression data. Commonly used unsupervised mathematical clustering methods are hierarchical clustering [34], *K*-means clustering [35], self-organizing maps [36] and principle component analysis [37]. A detailed description of these clustering algorithms is beyond the scope of this review. An accessible description of each of the commonly used algorithms is excellently reviewed elsewhere [38]. Software tools to perform mathematical clustering analysis include the commercially available Spotfire Descision Site Software (Spotfire, Göteborg, Sweden) and the freely available Cluster ([34]; http://rana.lbl.gov/EisenSoftware.htm), Genesis ([39]; http://genome.tugraz.at) and web-based "Classification of Expression Arrays" ([37] ;http://classify.stanford.edu).

Application of clustering methods has become a main step in the analysis of microarray data. In contrast, clustering methods are not frequently applied to SAGE data. Indeed, without normalization of SAGE tag counts into relative frequencies, these methods are not directly applicable to SAGE data. This might explain the limited application of these methods in SAGE studies. When using mathematical clustering, it is recommended to try different methods (*i.e.* *K*-means, hierarchical) since the different methods should yield similar results.

FUNCTIONAL MATHEMATICAL CLUSTERING

In silico subtraction methods can be used to identify preferential tag expression in the SAGE libraries of interest using the "Tissue Preferential Expression" (TPE) algorithm. Using this algorithm, expression specificity is established for each tag by comparing its occurrence and frequency to a panel of reference SAGE libraries derived from a number of whole tissues. This method has been successfully used to identify preferential expression of "no match" SAGE tags that possibly correspond to unknown genes that may serve as specific markers for disease [40,41].

TRANSLATING EXPRESSION DATA INTO FUNCTIONAL DATA

Large-scale expression profiling technologies are often utilized to characterize genetic hallmarks of disease (*i.e.* cancer) or to uncover potential novel groups of genes that participate in a certain biological process. Typically, a few tens to hundreds of genes are differentially expressed between experimental conditions. But what is the biological significance of these transcriptional changes? The interpretation of large-scale gene expression data is one of the most challenging fields in genomics. An important issue determining correct interpretation of large-scale gene expression data is the reliability of the generated data. Although microarray and SAGE are equally sensitive, both technologies are limited in the detection of instable transcripts and low abundant gene expression. Therefore, it is essential to determine an appropriate threshold that defines the reliable minimal detection limit of the large-scale gene expression profiling technology. To obtain more depth of analysis and improve the detection of low abundant transcripts using SAGE, one could increase the number of sequenced tags per library. Estimates of the number of tags required to cover the complete human transcriptome within a single cell population exceed 1.2 million tags [42]. Single molecule sequencing has enourmously expanded the number of SAGE tags that can be sequenced, both in practical and cost terms [43].

Due to technological limitations of large-scale expression profiling methods, the most representative results are yielded from uniform, single cell populations. A frequently encountered problem in expression profiling studies is the heterogeneity of the samples used. Gene expression variations in the cells of interest are often leveled due to the presence of other cell types. Particularly for complex heterogeneous samples, such as tissue biopsies of the airway epithelium, laser microdissection of the cells of interest is a valuable method to only select the cells of interest for subsequent profiling analysis [44].

To address the multiple testing problem and the associated detection of false positives that occur with large-scale expression profiling technologies, it is essential to verify microarray and SAGE data with other techniques such as quantitative real-time

polymerase chain reaction (qPCR) or in situ hybridization. SAGE, microarray and qPCR perform equally in estimating the expression levels of high abundant genes. For genes expressed at low levels, large-scale expression profiling methods and qPCR correctly estimate the directional change in expression. However, discrepancies in the magnitude of differential gene expression may exist between large-scale expression profiling methods and qPCR [31,45].

The main purpose of data interpretation is to predict and establish coherence within the observed findings, ultimately leading to new hypotheses. To accomplish these goals, the Gene Ontology Consortium developed a unified annotation for genes and their functions consisting of three independent hierarchical treelike structures (biological process, molecular function and cellular localization) called Gene Ontology (GO). Aside from information on well-known genes, the GO database also includes predicted information on novel and inferred genes. Matching genomics data to GO not only allows screening for known biological processes but also for previously unexplored biological processes in the model system of interest. Because graphical representations are more illustrative than lists of differentially expressed genes, initiatives have been undertaken also to visualize genomics data. The "Kyoto Encyclopedia of Genes and Genomes" (KEGG) and BioCarta are excellent examples of these initiatives. To facilitate the usage of the GO database with KEGG maps, Dahlquist *et al.* have developed the Gene Microarray Pathway Profiler (GenMAPP) [46] allowing automated mapping and visualization of genomics data. The recent advancements in the field of functional data analysis have been very successful in accelerating genomics data analysis.

Analyzed data should be interpreted with caution since the gene discovery rate still exceeds the speed by which researchers can experimentally assess gene function of newly identified genes. Consequently, the biological function(s) of a substantial proportion of genes in genetic databases have not yet been experimentally established. In addition, gene expression profiling reveals differences at the level of gene expression that are not necessarily reflected at the protein level. For instance, if proteins are constructed of different subunits that are encoded by multiple genes, follow-up research is warranted to establish the functional consequences of the transcriptional changes in one or more genes encoding the subunits. Nonetheless, genomics technology provides very efficient ways to screen for putative functional consequences in multimeric protein complexes encoded by multiple genes.

APPLICATION OF SAGE AND MICROARRAY TO STUDY PATHOGEN-EPITHELIAL CELLS INTERACTIONS

So far, we have discussed the potentials and limitations of commonly applied large-scale expression profiling methods including their data analysis strategies. In the next part of this review, the application of the SAGE and microarray analysis in host-pathogen research will be highlighted by discussing recent publications that applied these technologies. This part is divided into three sections: *(i)* application of SAGE and *(ii)* microarray analysis to study the epithelial host defense response in the human airways; and *(iii)* application of microarray analysis to investigate the processes that occur within the micro-organism upon contact with airway epithelial cells. Large-scale gene expression profiling technology can be applied to prokaryotic model systems, but there are limitations. For instance, SAGE can not be used since mRNAs in prokaryotes are not polyadenylated; a requirement for the SAGE technology. Microarray technology is the only platform that can be used to investigate transcriptional changes in prokaryotes at large scale. The first studies published have demonstrated that microarray analysis of the invading pathogen can be very instrumental in understanding the cross-talk between the host and the pathogen

APPLICATION OF THE SAGE TECHNOLOGY TO STUDY CHANGES IN EPITHELIAL GENE EXPRESSION AFTER MICROBIAL EXPOSURE

SAGE has been applied to investigate diverse areas of immunology and host defense including the processes of differentiation and development of T cells [47], B cells [48] and Natural Killer cells [49]. Furthermore, SAGE has been used to monitor the transcriptional changes upon HIV infection in T cell lines [50] as well as LPS activation in monocyte-derived dendritic cells [51]. The use of SAGE to specifically investigate host-pathogen interactions in the human airways is so far limited to our own study [18]. In our search for novel epithelial host defense molecules, bronchial epithelial cells were exposed for 6 hours to *P. aeruginosa* and a mixture of the pro-inflammatory cytokines IL-1β and TNFα [31]. SAGE revealed that the expression of keratins, proteinase inhibitors, S100 calcium-binding proteins and IL-1 family members was affected upon exposure to *P. aeruginosa* and the pro-inflammatory cytokines IL-1β and TNFα. The first three families of affected genes all contribute to cytoskeletal architecture. It was therefore suggested that bronchial epithelial cells specifically strengthen the primary physical barrier upon microbial exposure to cope with infection. However, the exact mode of action is unclear. The S100 calcium-binding proteins and the proteinase inhibitors Secretory Leukocyte Proteinase Inhibitor (SLPI) and Elafin (SKALP/PI3) were among the differentially expressed genes

with highest expression in both bronchial epithelial cells and keratinocytes. The protein complex formed by the S100 members A8 and A9 serves as antimicrobial agent, acts as chemoattractant for leukocytes and enhances the transendothelial migration of these cells [52]. Association of this protein complex with inflammatory disorders has been demonstrated in 1975 by Wilson *et al.* [53] who observed elevated serum levels of the S100A8/A9 complex in patients with cystic fibrosis (CF). The proteinase inhibitors SLPI and Elafin not only inhibit protease activity, but also exert antimicrobial activity [54]. The fourth group of molecules that was identified comprised cytokines. In particular, several members of the IL-1 family were identified. Despite its unknown function, the identification of the novel IL-1-family member IL-1F9 as cytokine contributing to the microbial response might be of interest. Our comparative genomic analysis revealed that this cytokine might fulfill specific immune signaling functions in epithelial cells since its expression is highly restricted to these cells [18].

To further delineate the early host defense response in bronchial epithelial cells, we performed a comparative genomic analysis of our model and of a culture model of epithelial inflammation in keratinocytes [18]. A genetic signature of host defense was identified, the expression of which was restricted to these cell types. In keratinocytes, the majority of these genes encode for proteins that are required for the assembly of the cornified envelope. This protein structure forms the physical barrier in skin. In analogy, this comparative genomic analysis revealed that bronchial epithelial cells may strengthen their physical barrier by the formation of an impermeable protein envelope. A number of the components of the protein envelope, including members of the S100 calcium-binding protein family and proteinase inhibitors, exert additional host defense functions. Therefore, the incorporation of these molecules in the protein envelope appears to serve multiple purposes.

In summary, the use of SAGE showed that bronchial epithelial cells respond to microbial exposure by strengthening the physical barrier and by releasing cytokines to alert the immune system.

APPLICATION OF DNA MICROARRAYS TO STUDY EPITHELIAL GENE EXPRESSION AFTER MICROBIAL EXPOSURE

The use of microarray technology has been more widespread than SAGE in studying host-pathogen interactions as judged by the number of publications. Studies on *P. aeruginosa* are of special interest because both the transcriptional changes in the host and the pathogen have been assessed using the microarray technology. Ichikawa *et al.* [11,55] were the first to demonstrate differential gene expression in the lung epithelial cell line A549 upon 1-3 hours exposure to *P. aeruginosa*. Although a relatively small number of

1,506 probes was assessed using a cDNA microarray, this study illustrated the usefulness of microarrays in this field of research. Differential gene expression of the transcription factor interferon regulatory factor 1 (IRF-1) was reported, which was independent of LPS or its signaling pathways. It was demonstrated that microbial adherence is crucial for the increased expression of IRF-1 since the non-adherent PAK-NP *P. aeruginosa* strain was unable to enhance IRF transcription. The PAK-NP strain carries a mutation in the *pilA* gene that encodes for the major subunit of the type IV pili that normally serves as adhesin to epithelial cells.

The interaction between cells of the bronchial epithelial cell line BEAS-2B and *Bordetella pertussis* has been studied using oligonucleotide arrays. This microorganism is the causative pathogen for whooping cough in humans [10]. Differential gene expression was assessed after 3 hours of exposure to *B. pertussis.* Transcriptional changes were observed in pro-inflammatory genes as characterized by the increased expression of cytokines and chemokines such as IL-8 (CXCL8), CCL2/MCP-1, CXCL1/Groα and CXCL2/Groβ. Enhanced expression of the chemokine IL-8 is considered one of the hallmarks of the epithelial response to *P. aeruginosa.* Also other studies showed regulation of the epithelial expression of these chemokines after microbial exposure, including our SAGE analysis.

Notably, epithelial cells require at least 3-6 hours of microbial exposure before the transcriptional response can be measured robustly [11,31,56]. This observation was also demonstrated in the cystic fibrosis (CF) bronchial epithelial cell line IB3-1 upon 3 hours of exposure to *P. aeruginosa* [57]. Compared to their mutation-corrected counterparts, lower expression of proteinase inhibitors and increased expression of cytokines such as IL-6 and IL-8 was observed in the CF cell line.

Using both oligonucleotide and cDNA microarrays, the transcriptional response of MM-39 submucosal tracheal gland epithelial cells to *S. aureus* was assessed [56]. *S. aureus* is one of the first pathogens to colonize the airways of cystic fibrosis patients. Especially the supernatants of this microorganism caused a robust transcriptional response in MM-39 cells, with increased gene expression of members of the JAK/STAT and NK-κB and AP-1 pathways after 3 hour exposure to *S. aureus* supernatants. These pathways lead to downstream transcription of the pro-inflammatory cytokines IL-1α, IL-1β and IL-6 and the chemokine IL-8.

Collectively, these microarray data suggest that soluble virulence factors released by microorganisms contribute to the induction of an epithelial host defense response. All discussed reports observed transcriptional regulation of members of these cytokine and chemokine families including their associated signaling pathways. The microarray studies revealed that especially the cytokine expression is enhanced in the early phase of microbial infection of airway epithelial cells.

Sensing microbial exposure by bronchial epithelial cells is in part mediated through pattern recognition receptors such as the Toll-like receptors (TLR2, TLR3, TLR4, TLR5 and

TLR9). In the studies discussed, the focus was on TLR2 and TLR5. The ligands for TLR2 appear to be amongst others lipoproteins, whereas flagellin is a major ligand for TLR5. Recent studies suggest that lipoprotein I of *P. aeruginosa* is a TLR2/4 ligand [58]. Antibodies against TLR2 can attenuate the action of lipoproteins. Enhanced expression of TLR2 and TLR5 or increased activity of the downstream signaling cascades in CF epithelial cells may possibly account for the exaggerated inflammatory response in these cells. Flagellin is a structural component of the bacterial flagellum that provides motility to the microorganism and appears to be essential in infection. The activating potential of flagellin, which is secreted by bacteria, on bronchial epithelial cells has been demonstrated in several investigations [59,60], and leads to increased expression of various mediators, including the antimicrobial peptide human β-defensin-2 (DEFB4; [60]). Despite the eminent function of specialized antimicrobial peptides in the epithelial innate immune system, they do not appear to be highly expressed in the early epithelial host defense response. The reason for this is unknown, but could possibly be explained in part by the location of the sensing TLR5 at the basolateral membrane of epithelial cells [61]. In time course experiments using live *P. aeruginosa*, presence of the microorganism at the basolateral membrane was not demonstrated earlier than 12 hours post infection. This implies that TLR5 activation occurs later in the infection process, providing an explanation for the delayed response of epithelial cells to produce human β-defensins. However, it needs to be noted that a recent study demonstrated also the apical localization of TLR5 on tracheal epithelial cells [62] and therefore localization of TLR5 may not be the sole explanation for the limited expression of human β-defensin-2 early after microbial exposure.

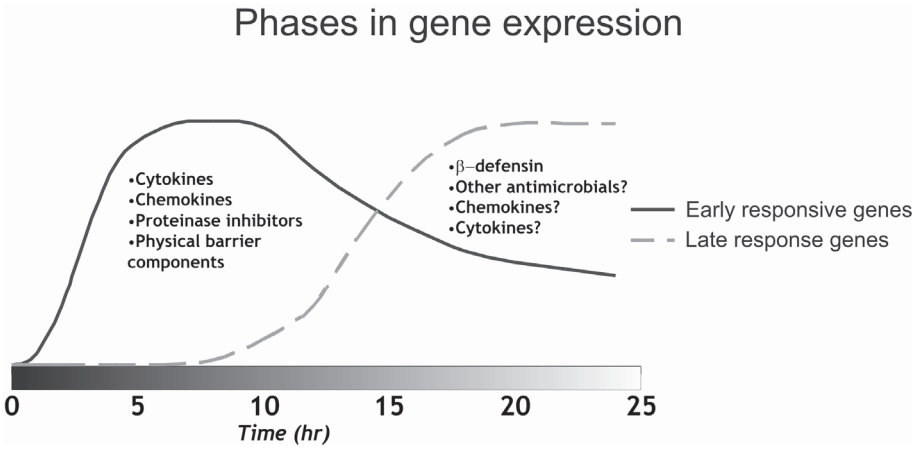


Figure 2: Phases in gene expression in bronchial epithelial cells upon microbial contact. The early transcriptional response in epithelial cells upon microbial contact is characterized by expression of cytokines, chemokines, and proteins that are involved in strengthening the physical barrier. Expression of specialized antimicrobial peptides may occur at later stages during infection, possibly in part because of the localization of pattern recognition receptors that lead to the transcription of these peptides. See text for details.

Hypothetically, the epithelial host defense response may therefore consist of two phases ***(Figure 2)***. During the first phase when microbes attach to the apical surface of epithelial cells, the physical barrier is strengthened and expression of signaling molecules such as cytokines and chemokines is increased. In the proceeding phase when microorganisms migrate to the basolateral epithelial membrane, production of specialized antimicrobial agents may occur.

APPLICATION OF DNA MICROARRAYS TO STUDY MICROBIAL GENE EXPRESSION AFTER EXPOSURE TO EPITHELIAL CELLS

In order to survive in the lung, pathogens have to deal with the defense mechanisms of the host such as the release of antimicrobial peptides and the oxidative burst utilized by phagocytes. How pathogens accomplish their survival under these circumstances is largely unknown. As discussed, the microarray technology is the only high-throughput method available to assess microbial gene expression at large scale. For *P. aeruginosa*, the first oligonucleotide array was launched shortly after the completion of sequencing the genome of the PAO1 laboratory strain [63]. The PAO1 microarray has been used to uncover regulatory networks in *P. aeruginosa* upon varying environmental conditions (reviewed in [64]). A limited number of investigations have specifically focused on the interactions of this microorganism with host epithelial cells. It is known that virulence factors are crucial for the ability of *P. aeruginosa* to infect the host. The conversion of the non-mucoid to the alginate-overproducing mucoid form seems to be essential for *P. aeruginosa* to cause chronic colonization of the airways in patients with cystic fibrosis (CF) [65]. Although alginate is immunologically inert, other virulence factors associated with the conversion to mucoidy may cause host tissue destruction. Bacterial virulence factors often act as two-edged swords: on the one hand they damage host tissue to promote bacterial adherence and survival, whereas on the other hand they activate the defense system of the host [60]. Frisk *et al.* [66] and Firoved *et al.* [67] were the first to profile gene expression in *P. aeruginosa* upon contact with human airway epithelial cells. Frisk *et al.* evaluated 4- and 12 hours interactions of primary normal human airway epithelial cells with the non-mucoid *P. aeruginosa* laboratory strain PAO1. During the course of infection, *P. aeruginosa* migrated from the apical membrane to the basolateral membrane. Global expression profiling revealed the activation of phosphate and repression of iron acquisition genes, indicating that *P. aeruginosa* may be able to acquire sufficient quantities of iron from host cells but not phosphate for growth. Alternatively, the repression of iron acquisition genes has also been associated with the antioxidant response of *P. aeruginosa* upon oxidative stress [68].

Using *in vitro* models for CF, Firoved *et al*. found that specifically the expression of bacterial lipoproteins is strongly induced in mucoid *P. aeruginosa* upon contact with human epithelial cells [67]. Because of their toxicity, these molecules are referred to as lipotoxins. Synthetic lipopeptides resembling the N-terminal parts of these mature lipoproteins caused the activation of NF-κB through TLR2, a pattern recognition receptor of the innate host defense system. Mature lipoproteins were able to induce IL-8 production in normal bronchial epithelial cells. This induction could be partially suppressed by antibodies against TLR2 [67]. In addition to lipoproteins, *P. aeruginosa* flagellin has been identified as the virulence factor that specifically induced the epithelial host defense response through TLR5 [60] and in combination with TLR2 [59]. Flagellin is a structural component of bacterial flagella necessary for normal growth and infection. By modulating the expression of flagellin, *P. aeruginosa* may be able to enhance colonization in for instance patients with CF [69]. Whereas these studies focused on the role of TLR2 and TLR5, other studies have shown that a range of TLRs are involved in recognizing and responding to bacterial exposure. These include TLR4, that recognizes lipopolysaccharide, and TLR9, that is involved in the recognition of bacterial DNA [70].

In summary, during the course of infection *P. aeruginosa* increases its expression of phosphate acquisition genes, lipotoxins and DNA repair genes and decreases the expression of iron acquisition genes. Increased expression of phosphate acquisition genes might play a role in safeguarding the nutritional requirements. Changes in expression of genes related to Iron metabolism and DNA repair are associated with the antioxidant response of *P. aeruginosa* to oxidative stress. The excessive induction of lipotoxins upon contact seems to be due to yet unknown properties of CF cells. These molecules are potent agonists of TLR2 and may account for the excessive induction of cytokine gene expression and the enhanced inflammatory response.

CONCLUSION

In this review we have provided an overview of commonly used large-scale gene expression profiling methods, and discussed how these have been used in host-pathogen research focused on the interaction between respiratory pathogens and lung epithelial cells. Since both the SAGE and microarray technology enable the profiling of thousands of genes at once, a frequently asked question is which technique is the best to choose. The answer is simple: neither of the techniques is *the best*. The selection of a high-throughput gene expression profiling technique highly depends on the research question and practical considerations such as sample number. The elegance of the SAGE technology is its ability of gene discovery and ease of comparative research. On the other hand, the

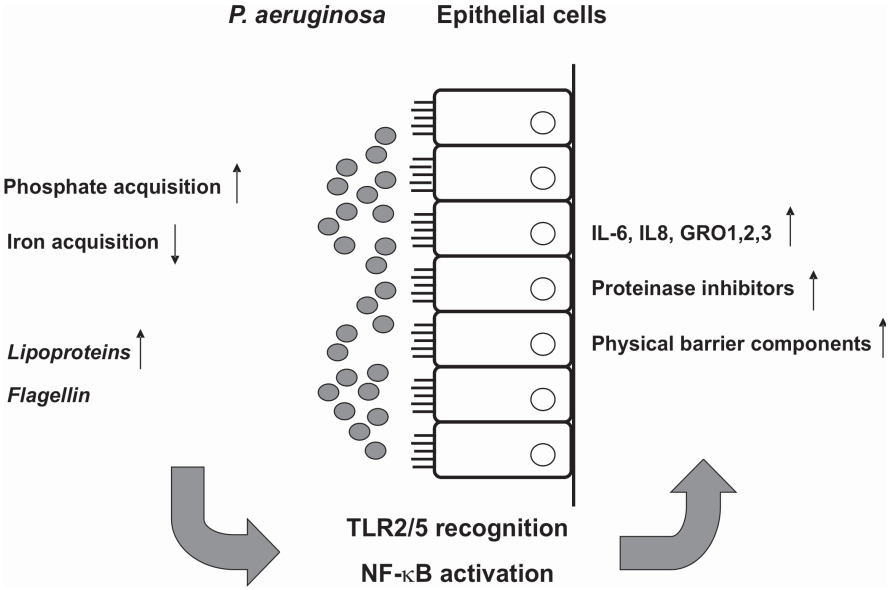


Figure 3: Transcriptional changes in P. aeruginosa and epithelial cells upon interaction: novel insights from gene expression profiling studies. The transcriptional changes in *P. aeruginosa* are characterized by increased expression of phosphate acquisition genes and lipoproteins and decreased expression of iron acquisition genes. Both lipotoxins and flagellin activate pattern recognition receptors present on epithelial cells. Activation of these receptors leads to increased expression of cytokines, chemokines, proteinase inhibitors and components necessary to strengthen the physical barrier.

benefits of the microarray analysis include the ability to profile prokaryotic gene expression and the less labor intensive nature of the technology as compared to SAGE.

One of the advantages of the application of large-scale gene expression profiling methods is that it allows the discovery of previously unknown associations at the level of gene expression. Application of these methods has increased our understanding of the dynamic interplay between the host and the pathogen during the course of infection ***(Figure 3)***. Particularly, large-scale gene expression profiling revealed insights in the sequential events that underlie the inflammatory epithelial host defense response. In addition, the combined investigations provided novel mechanistic insights into the interplay during the course of infection in both the host and the pathogen. At the initial stage of infection, bronchial epithelial cells predominantly respond by increasing their cytokine production to alert the immune system and by strengthening the physical barrier. The transcriptional changes in the pathogen are dominated by differential expression of genes involved in the response to oxidative stress. These include the repression of iron acquisition genes and increased expression of lipoproteins and phosphate acquisition genes. Lipoprotein expression by *P. aeruginosa* was particularly altered upon contact with CF epithelial cells.

Although large-scale gene expression profiling methods are highly efficient in data generation, the analysis of expression profiles remains challenging and time consum-

ing. Interpretation of large-scale gene expression data is the main challenge in the post-genomics era. Due to a lack of knowledge on data analysis methods and resources, researchers often tend to focus on those genes known to be involved in the biological process they are investigating. Continuing developments in the field of bioinformatics are expected to provide easy-to-use analysis methods that are helpful for non-bioinformaticians to analyze large-scale gene expression data. Notwithstanding the limitations of functional data interpretation, large-scale gene expression profiling has great promise to increase our understanding of and to uncover novel insights in the crosstalk between the host and the pathogen in health and disease.

REFERENCES

1. Bals R. & Hiemstra P.S. (2004) Innate immunity in the lung: how epithelial cells fight against respiratory pathogens. *Eur.Respir.J.* **23**, 327-333.
2. Travis S.M., Singh P.K., & Welsh M.J. (2001) Antimicrobial peptides and proteins in the innate defense of the airway surface. *Curr.Opin.Immunol.* **13**, 89-95.
3. Diamond G., Legarda D., & Ryan L.K. (2000) The innate immune response of the respiratory epithelium. *Immunol.Rev.* **173**, 27-38.
4. Cookson W. (2004) The immunogenetics of asthma and eczema: a new focus on the epithelium. *Nat.Rev.Immunol.* **4**, 978-988.
5. Alwine J.C., Kemp D.J., & Stark G.R. (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc.Natl.Acad. Sci.U.S.A* **74**, 5350-5354.
6. Schena M., Shalon D., Davis R.W., & Brown P.O. (1995) Quantitative Monitoring of Gene-Expression Patterns with A Complementary-Dna Microarray. *Science* **270**, 467-470.
7. Adams M.D., Soares M.B., Kerlavage A.R., Fields C., & Venter J.C. (1993) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat. Genet.* **4**, 373-380.
8. Velculescu V.E., Zhang L., Vogelstein B., & Kinzler K.W. (1995) Serial analysis of gene expression. *Science* **270**, 484-487.
9. Brenner S., Johnson M., Bridgham J., Golda G., Lloyd D.H., Johnson D., Luo S., McCurdy S., Foy M., Ewan M., Roth R., George D., Eletr S., Albrecht G., Vermaas E., Williams S.R., Moon K., Burcham T., Pallas M., DuBridge R.B., Kirchner J., Fearon K., Mao J., & Corcoran K. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat.Biotechnol.* **18**, 630-634.
10. Belcher C.E., Drenkow J., Kehoe B., Gingeras T.R., McNamara N., Lemjabbar H., Basbaum C., & Relman D.A. (2000) The transcriptional responses of respiratory epithelial cells to Bordetella pertussis reveal host defensive and pathogen counter-defensive strategies. *Proc.Natl.Acad.Sci.U.S.A* **97**, 13847-13852.
11. Ichikawa J.K., Norris A., Bangera M.G., Geiss G.K., 't Wout A.B., Bumgarner R.E., & Lory S. (2000) Interaction of pseudomonas aeruginosa with epithelial cells: identification of differentially regulated genes by expression microarray analysis of human cDNAs. *Proc.Natl.Acad.Sci.U.S.A* **97**, 9659-9664.
12. Kagnoff M.F. & Eckmann L. (2001) Analysis of host responses to microbial infection using gene expression profiling. *Curr.Opin.Microbiol.* **4**, 246-250.
13. Diehn M. & Relman D.A. (2001) Comparing functional genomic data sets: lessons from DNA microarray analyses of host-pathogen interactions. *Curr.Opin.Microbiol* **4**, 95-101.
14. Velculescu V.E., Zhang L., Zhou W., Vogelstein J., Basrai M.A., Bassett D.E., Jr., Hieter P., Vogelstein B., & Kinzler K.W. (1997) Characterization of the yeast transcriptome. *Cell* **88**, 243-251.
15. Evans S.J., Datson N.A., Kabbaj M., Thompson R.C., Vreugdenhil E., de Kloet E.R., Watson S.J., & Akil H. (2002) Evaluation of Affymetrix Gene Chip sensitivity in rat hippocampal tissue using SAGE analysis. Serial Analysis of Gene Expression. *Eur.J Neurosci.* **16**, 409-413.

16. Ishii M., Hashimoto S., Tsutsumi S., Wada Y., Matsushima K., Kodama T., & Aburatani H. (2000) Direct comparison of GeneChip and SAGE on the quantitative accuracy in transcript profiling analysis. *Genomics* **68**, 136-143.

17. Ibrahim A.F., Hedley P.E., Cardle L., Kruger W., Marshall D.F., Muehlbauer G.J., & Waugh R. (2005) A comparative analysis of transcript abundance using SAGE and Affymetrix arrays. *Funct.Integr. Genomics* **5**, 163-174.

18. Vos J.B., Datson N.A., Van Kampen A.H., Luyf A.C., Verhoosel R.M., Zeeuwen P.L., Olthuis D., Rabe K.F., Schalkwijk J., & Hiemstra P.S. (2006) A molecular signature of epithelial host defense: comparative gene expression analysis of cultured bronchial epithelial cells and keratinocytes. *BMC. Genomics* **7**, 9.

19. Harbers M. & Carninci P. (2005) Tag-based approaches for transcriptome research and genome annotation. *Nat.Methods* **2**, 495-502.

20. Impey S., McCorkle S.R., Cha-Molstad H., Dwyer J.M., Yochum G.S., Boss J.M., McWeeney S., Dunn J.J., Mandel G., & Goodman R.H. (2004) Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* **119**, 1041-1054.

21. Roh T.Y., Ngau W.C., Cui K., Landsman D., & Zhao K. (2004) High-resolution genome-wide mapping of histone modifications. *Nat.Biotechnol.* **22**, 1013-1016.

22. Kim J., Bhinge A.A., Morgan X.C., & Iyer V.R. (2005) Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. *Nat.Methods* **2**, 47-53.

23. Chen J. & Sadowski I. (2005) Identification of the mismatch repair genes PMS2 and MLH1 as p53 target genes by using serial analysis of binding elements. *Proc.Natl.Acad.Sci.U.S.A* **102**, 4813-4818.

24. Marinescu V.D., Kohane I.S., Kim T.K., Harmin D.A., Greenberg M.E., & Riva A. (2006) START: an automated tool for serial analysis of chromatin occupancy data. *Bioinformatics.* **22**, 999-1001.

25. Sievertzon M., Nilsson P., & Lundeberg J. (2006) Improving reliability and performance of DNA microarrays. *Expert.Rev.Mol.Diagn.* **6**, 481-492.

26. Barrett J.C. & Kawasaki E.S. (2003) Microarrays: the use of oligonucleotides and cDNA for the analysis of gene expression. *Drug Discov.Today* **8**, 134-141.

27. Petersen D., Chandramouli G.V., Geoghegan J., Hilburn J., Paarlberg J., Kim C.H., Munroe D., Gangi L., Han J., Puri R., Staudt L., Weinstein J., Barrett J.C., Green J., & Kawasaki E.S. (2005) Three microarray platforms: an analysis of their concordance in profiling gene expression. *BMC.Genomics* **6**, 63.

28. Brazma A., Hingamp P., Quackenbush J., Sherlock G., Spellman P., Stoeckert C., Aach J., Ansorge W., Ball C.A., Causton H.C., Gaasterland T., Glenisson P., Holstege F.C., Kim I.F., Markowitz V., Matese J.C., Parkinson H., Robinson A., Sarkans U., Schulze-Kremer S., Stewart J., Taylor R., Vilo J., & Vingron M. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat.Genet.* **29**, 365-371.

29. Lash A.E., Tolstoshev C.M., Wagner L., Schuler G.D., Strausberg R.L., Riggins G.J., & Altschul S.F. (2000) SAGEmap: a public gene expression resource. *Genome Res.* **10**, 1051-1060.

30. Boon K., Osorio E.C., Greenhut S.F., Schaefer C.F., Shoemaker J., Polyak K., Morin P.J., Buetow K.H., Strausberg R.L., De Souza S.J., & Riggins G.J. (2002) An anatomy of normal and malignant gene expression. *Proc.Natl.Acad.Sci.U.S.A* **99**, 11287-11292.

31. Vos J.B., van Sterkenburg M.A., Rabe K.F., Schalkwijk J., Hiemstra P.S., & Datson N.A. (2005) Transcriptional response of bronchial epithelial cells to Pseudomonas aeruginosa: identification of early mediators of host defense. *Physiol Genomics* **21**, 324-336.

32. Silva A.P., De Souza J.E., Galante P.A., Riggins G.J., De Souza S.J., & Camargo A.A. (2004) The impact of SNPs on the interpretation of SAGE and MPSS experimental data. *Nucleic Acids Res.* **32**, 6104-6110.

33. Okada Y., Sahara T., Mitsubayashi H., Ohgiya S., & Nagashima T. (2005) Knowledge-assisted recognition of cluster boundaries in gene expression data. *Artif.Intell.Med.* **35**, 171-183.

34. Eisen M.B., Spellman P.T., Brown P.O., & Botstein D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc.Natl.Acad.Sci.U.S.A* **95**, 14863-14868.

35. Tavazoie S., Hughes J.D., Campbell M.J., Cho R.J., & Church G.M. (1999) Systematic determination of genetic network architecture. *Nat.Genet.* **22**, 281-285.

36. Tamayo P., Slonim D., Mesirov J., Zhu Q., Kitareewan S., Dmitrovsky E., Lander E.S., & Golub T.R. (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc.Natl.Acad.Sci.U.S.A* **96**, 2907-2912.

37. Raychaudhuri S., Sutphin P.D., Chang J.T., & Altman R.B. (2001) Basic microarray analysis: grouping and feature reduction. *Trends Biotechnol.* **19**, 189-193.

38. Quackenbush J. (2001) Computational analysis of microarray data. *Nat.Rev.Genet.* **2**, 418-427.

39. Sturn A., Quackenbush J., & Trajanoski Z. (2002) Genesis: cluster analysis of microarray data. *Bioinformatics.* **18**, 207-208.

40. Moreno J.C., Pauws E., Van Kampen A.H., Jedlickova M., de Vijlder J.J., & Ris-Stalpers C. (2001) Cloning of tissue-specific genes using serial analysis of gene expression and a novel computational substraction approach. *Genomics* **75**, 70-76.

41. Pauws E., Veenboer G.J., Smit J.W., de Vijlder J.J., Morreau H., & Ris-Stalpers C. (2004) Genes differentially expressed in thyroid carcinoma identified by comparison of SAGE expression profiles. *FASEB J.* **18**, 560-561.

42. Trendelenburg G., Prass K., Priller J., Kapinya K., Polley A., Muselmann C., Ruscher K., Kannbley U., Schmitt A.O., Castell S., Wiegand F., Meisel A., Rosenthal A., & Dirnagl U. (2002) Serial analysis of gene expression identifies metallothionein-II as major neuroprotective gene in mouse focal cerebral ischemia. *J Neurosci.* **22**, 5879-5888.

43. Rogers Y.H. & Venter J.C. (2005) Genomics: massively parallel sequencing. *Nature* **437**, 326-327.

44. Datson N.A., Meijer L., Steenbergen P.J., Morsink M.C., van der L.S., Meijer O.C., & de Kloet E.R. (2004) Expression profiling in laser-microdissected hippocampal subregions in rat brain reveals large subregion-specific differences in expression. *Eur.J.Neurosci.* **20**, 2541-2554.

45. Datson N.A., van der P.J., de Kloet E.R., & Vreugdenhil E. (2001) Identification of corticosteroid-responsive genes in rat hippocampus using serial analysis of gene expression. *Eur.J.Neurosci.* **14**, 675-689.

46. Dahlquist K.D., Salomonis N., Vranizan K., Lawlor S.C., & Conklin B.R. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat.Genet.* **31**, 19-20.

47. Nagai S., Hashimoto S., Yamashita T., Toyoda N., Satoh T., Suzuki T., & Matsushima K. (2001) Comprehensive gene expression profile of human activated T(h)1- and T(h)2-polarized cells. *Int.Immunol.* **13**, 367-376.

48. Feldhahn N., Schwering I., Lee S., Wartenberg M., Klein F., Wang H., Zhou G., Wang S.M., Rowley J.D., Hescheler J., Kronke M., Rajewsky K., Kuppers R., & Muschen M. (2002) Silencing of B cell receptor signals in human naive B cells. *J Exp Med* **196**, 1291-1305.

49. Kang H.S., Kim E.M., Lee S., Yoon S.R., Kawamura T., Lee Y.C., Kim S., Myung P.K., Wang S.M., & Choi I. (2005) Stage-dependent gene expression profiles during natural killer cell development. *Genomics* **86**, 551-565.

50. Ryo A., Suzuki Y., Ichiyama K., Wakatsuki T., Kondoh N., Hada A., Yamamoto M., & Yamamoto N. (1999) Serial analysis of gene expression in HIV-1-infected T cell lines. *FEBS Lett.* **462**, 182-186.

51. Hashimoto S.I., Suzuki T., Nagai S., Yamashita T., Toyoda N., & Matsushima K. (2000) Identification of genes specifically expressed in human activated and mature dendritic cells through serial analysis of gene expression. *Blood* **96**, 2206-2214.

52. Nacken W., Roth J., Sorg C., & Kerkhoff C. (2003) S100A9/S100A8: Myeloid representatives of the S100 protein family as prominent players in innate immunity. *Microsc.Res.Tech.* **60**, 569-580.

53. Wilson G.B., Fudenberg H.H., & Jahn T.L. (1975) Studies on cystic fibrosis using isoelectric focusing. I. An assay for detection of cystic fibrosis homozygotes and heterozygote carriers from serum. *Pediatr.Res.* **9**, 635-640.

54. Williams S.E., Brown T.I., Roghanian A., & Sallenave J.M. (2006) SLPI and elafin: one glove, many fingers. *Clin.Sci.(Lond)* **110**, 21-35.

55. Lory S. & Ichikawa J.K. (2002) Pseudomonas-epithelial cell interactions dissected with DNA microarrays. *Chest* **121**, 36S-39S.

56. Moreilhon C., Gras D., Hologne C., Bajolet O., Cottrez F., Magnone V., Merten M., Groux H., Puchelle E., & Barbry P. (2005) Live Staphylococcus aureus and bacterial soluble factors induce different transcriptional responses in human airway cells. *Physiol Genomics* **20**, 244-255.

57. Virella-Lowell I., Herlihy J.D., Liu B., Lopez C., Cruz P., Muller C., Baker H.V., & Flotte T.R. (2004) Effects of CFTR, interleukin-10, and Pseudomonas aeruginosa on gene expression profiles in a CF bronchial epithelial cell Line. *Mol.Ther.* **10**, 562-573.

58. Revets H., Pynaert G., Grooten J., & De Baetselier P. (2005) Lipoprotein I, a TLR2/4 ligand modulates Th2-driven allergic immune responses. *J Immunol.* **174**, 1097-1103.

59. Adamo R., Sokol S., Soong G., Gomez M.I., & Prince A. (2004) Pseudomonas aeruginosa flagella activate airway epithelial cells through asialoGM1 and toll-like receptor 2 as well as toll-like receptor 5. *Am.J Respir.Cell Mol.Biol.* **30**, 627-634.

60. Cobb L.M., Mychaleckyj J.C., Wozniak D.J., & Lopez-Boado Y.S. (2004) Pseudomonas aeruginosa flagellin and alginate elicit very distinct gene expression patterns in airway epithelial cells: implications for cystic fibrosis disease. *J Immunol.* **173**, 5659-5670.

61. Gewirtz A.T., Navas T.A., Lyons S., Godowski P.J., & Madara J.L. (2001) Cutting edge: bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial pro-inflammatory gene expression. *J.Immunol.* **167**, 1882-1885.

62. Zhang Z., Louboutin J.P., Weiner D.J., Goldberg J.B., & Wilson J.M. (2005) Human airway epithelial cells sense Pseudomonas aeruginosa infection via recognition of flagellin by Toll-like receptor 5. *Infect.Immun.* **73**, 7151-7160.

63. Stover C.K., Pham X.Q., Erwin A.L., Mizoguchi S.D., Warrener P., Hickey M.J., Brinkman F.S., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R., Smith K., Spencer D., Wong G.K., Wu

Z., Paulsen I.T., Reizer J., Saier M.H., Hancock R.E., Lory S., & Olson M.V. (2000) Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen. *Nature* **406**, 959-964.

64. Goodman A.L. & Lory S. (2004) Analysis of regulatory networks in Pseudomonas aeruginosa by genomewide transcriptional profiling. *Curr.Opin.Microbiol* **7**, 39-44.

65. Bonfield T.L., Panuska J.R., Konstan M.W., Hilliard K.A., Hilliard J.B., Ghnaim H., & Berger M. (1995) Inflammatory cytokines in cystic fibrosis lungs. *Am.J Respir.Crit Care Med* **152**, 2111-2118.

66. Frisk A., Schurr J.R., Wang G., Bertucci D.C., Marrero L., Hwang S.H., Hassett D.J., & Schurr M.J. (2004) Transcriptome analysis of Pseudomonas aeruginosa after interaction with human airway epithelial cells. *Infect.Immun.* **72**, 5433-5438.

67. Firoved A.M., Ornatowski W., & Deretic V. (2004) Microarray analysis reveals induction of lipoprotein genes in mucoid Pseudomonas aeruginosa: implications for inflammation in cystic fibrosis. *Infect.Immun.* **72**, 5012-5018.

68. Chang W., Small D.A., Toghrol F., & Bentley W.E. (2005) Microarray analysis of Pseudomonas aeruginosa reveals induction of pyocin genes in response to hydrogen peroxide. *BMC.Genomics* **6**, 115.

69. Wolfgang M.C., Jyot J., Goodman A.L., Ramphal R., & Lory S. (2004) Pseudomonas aeruginosa regulates flagellin expression as part of a global response to airway fluid from cystic fibrosis patients. *Proc.Natl.Acad.Sci.U.S.A* **101**, 6664-6668.

70. Akira S., Uematsu S., & Takeuchi O. (2006) Pathogen recognition and innate immunity. *Cell* **124**, 783-801.

71. Sargent T.D. & Dawid I.B. (1983) Differential gene expression in the gastrula of Xenopus laevis. *Science* **222**, 135-139.

72. Mullis K., Faloona F., Scharf S., Saiki R., Horn G., & Erlich H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb.Symp.Quant.Biol.* **51 Pt 1**, 263-273.

73. Liang P. & Pardee A.B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967-971.