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 3

THE TRANSCRIPTIONAL RESPONSE OF BRONCHIAL EPITHELIAL CELLS TO *PSEUDOMONAS AERUGINOSA*: IDENTIFICATION OF EARLY MEDIATORS OF HOST DEFENSE

Joost B. Vos [1], Marianne A. van Sterkenburg [1], Klaus F. Rabe [1], Joost Schalkwijk [2], Pieter S. Hiemstra [1], Nicole A. Datson [3]

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

[2] *Laboratory of Skin Biology and Experimental Dermatology, Nijmegen Center for Molecular Life Sciences, Radboud University Medical Center Nijmegen, The Netherlands*

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

Physiological Genomics, 2005 May 11;21(3):324-36.

ABSTRACT

The airway epithelium responds to microbial exposure by altering the expression of a variety of genes to increase innate host defense. In this study we aimed to delineate the early transcriptional response in human primary bronchial epithelial cells exposed for 6 hours to a mixture of IL-1β and TNFα or heat-inactivated *Pseudomonas aeruginosa*. Since the molecular mechanisms of epithelial innate host defense tissues are not fully understood, the open-ended expression profiling technique SAGE was applied to construct gene expression profiles covering 30.000 genes. A number of 292 genes was found to be differentially expressed. Expression of seven genes was confirmed by quantitative real-time PCR. Among differentially expressed genes, four classes or families were identified: keratins, proteinase inhibitors, S100 calcium-binding proteins and IL-1 family members. Marked transcriptional changes were observed for keratins that form a key component of the cytoskeleton in epithelial cells. The expression of the antimicrobial proteinase inhibitors Secretory Leukocyte Proteinase Inhibitor and elafin was elevated after microbial or cytokine exposure. Interestingly, expression of numerous S100 family members was observed and eight members, including S100A8 and S100A9, were among the most differentially expressed genes. Differential expression was also observed for the interleukin-1 family members IL-1β, IL-1 receptor antagonist, and IL-1F9, a newly discovered IL-1 family member. Clustering of differentially expressed genes into biological processes revealed that the early inflammatory response in airway epithelial cells to IL-1β/TNFα and *P. aeruginosa* is characterized by expression of genes involved in epithelial barrier formation and host defense.

INTRODUCTION

Every day humans breathe thousands of liters of ambient air containing numerous potentially harmful pathogens such as bacteria, fungi, parasites or viruses. Despite the exposure to these pathogens, severe lung infections are rare. The innate immune system plays a pivotal role in safeguarding the airways from inhaled substances. The airway epithelium is continuously exposed to respiratory pathogens and plays a central role in innate immunity [1,2]. Three main mechanisms are utilized by the epithelium to protect the host from infection. First, epithelial cells form a major part of the physical barrier against entry of pathogens. Second, the airway epithelium actively contributes to innate immunity by the secretion of defense substances such as antimicrobial polypeptides and proteinase inhibitors and by mucociliary clearance through coordinated secretion of mucus and ciliary activity. Third, epithelial cells produce mediators including cytokines and chemokines that serve to attract, mature and activate cells of the innate and adaptive immune system. Through this, epithelial cells relay signals of danger from the outside world to the inner body.

The epithelial tissue of the airways is frequently exposed to inhaled respiratory pathogens. These pathogens may colonize the airways when first line defense is hampered, a phenomenon observed in a variety of chronic inflammatory lung disorders including cystic fibrosis and chronic bronchitis. The innate immune function of the airway epithelium is known to be increased after exposure to microorganisms. Microorganisms may directly affect epithelial gene expression or may do so indirectly by stimulating macrophages to release pro-inflammatory cytokines such as IL-1β and TNFα which subsequently activate epithelial cells [3,4]. Direct cell activation by microorganisms and microbial products is mediated at least in part by pattern recognition receptors expressed on host cells such as the Toll-like receptors (TLR) [5,6]. It is well recognized that a broad spectrum of effector molecules secreted by epithelial cells is involved in innate immunity [1,7,8]. Over the past decade, a large number of antimicrobial peptides and proteinase inhibitors with antimicrobial activity has been identified. The expression of many of these effector molecules including β-defensins, secretory leukocyte proteinase inhibitor (SLPI) and elafin is induced upon microbial exposure. However, it is largely unknown which molecules are essential in mounting the innate immune response nor have the kinetics of these processes been studied.

The aim of the present study was to delineate the early response of human bronchial epithelial cells to direct microbial stimulation with heat-killed *Pseudomonas aeruginosa* or to indirect stimulation via exposure to the macrophage-derived cytokines IL-1β and TNFα. Therefore, the effect of these stimuli on bronchial epithelial cells was assessed at the level of gene transcription using SAGE. SAGE is a highly sensitive and reliable method to generate comprehensive expression profiles. Since SAGE is not limited by a predefined

set of genes, both known and unknown genes can be studied [9,10]. The sensitivity and reliability of SAGE has been confirmed by a number of independent techniques such as northern blot [11], real-time PCR [12], DNA macroarray [13] and DNA microarray technology [14]. Using SAGE, at least four families of genes were identified to be affected in epithelial cells after exposure to the pro-inflammatory cytokines IL-1β and TNFα and to *P. aeruginosa*. Keratins, proteinase inhibitors, members of the IL-1 family and S100 calcium-binding proteins were prominently present among the most differentially expressed genes. Together with the other differentially expressed genes, these four families of genes may contribute to the onset of the innate immune response.

MATERIALS AND METHODS

Bronchial epithelial cells

Subcultures of human bronchial epithelial cells were derived from bronchial tissue specimens obtained from patients who underwent a thoracotomy with lobectomy for lung cancer at the Leiden University Medical Center (LUMC, Leiden, The Netherlands) as described previously [15]. Tissue specimens were found to be normal as determined macroscopically by a pathologist and microscopically at the time of dissection of the epithelial cells. Cells were grown to near-confluence in 25-cm^2 culture flasks pre-coated with a matrix of vitrogen (30 µg/ml; Celtrix Laboratories, Palo Alto, CA), fibronectin (10 µg/ml; isolated from human plasma) and bovine serum albumin (BSA, 10 µg/ml; Boehringer Mannheim, Mannheim, Germany) in serum-free keratinocyte-SFM medium (KSFM, GibcoBRL/Life Technologies, Breda, The Netherlands) supplemented with 0.2 ng/ml epidermal growth factor (EGF; GibcoBRL/Life Technologies), 25 µg/ml bovine pituitary extract (BPE; GibcoBRL/Life Technologies), 1 mM isoproterenol (Sigma Chemicals, St. Louis, MO), 20 U/ml penicillin (Bio Whittaker, Walkersville, MD) and 20 µg/ml streptomycin (Bio Whittaker). After the cells had reached near-confluence, cells were incubated for 36 hours in high calcium medium to allow differentiation as described previously [15]. High calcium medium was composed of KSFM (GibcoBRL/Life Technologies) supplemented with 1 mM $CaCl_2$, 5 nM retinoic acid (Sigma Chemicals), 0.2 ng/ml EGF, 25 µg/ml BPE, 20 U/ml penicillin and 20 µg/ml streptomycin.

Pseudomonas aeruginosa (PA01)

The non-mucoid *P. aeruginosa* strain PA01 (BAA-47, American Type Culture Collection, Rockville, MD, USA) was grown overnight in Brain-Heart Infusion (BHI) medium to stationary phase at an agitation rate of 275 rotations per minute at 37°C. After three washes with PBS, bacteria were resuspended in PBS containing 50% glycerol at a concentration of 10^9 colony forming units (CFU) per ml as determined by optical density. Samples of

the bacterial suspension were plated onto blood agar plates to verify the measured CFU of the suspension. The bacteria were heat-inactivated for 45 minutes at 95°C in a water bath and stored in aliquots at –80°C until further use.

Stimulation experiments

For the construction of the SAGE libraries, subcultures of primary bronchial epithelial cells derived from seven different donors were cultured in 25-cm^2 flasks and stimulated for 6 hours with either a mixture of the pro-inflammatory cytokines IL-1β (20 ng/ml; PeproTech, Rocky Hill, NJ) and TNFα (20 ng/ml; PeproTech) or with heat-inactivated *P. aeruginosa* (10^7 cfu/ml) or with high calcium medium alone. Three 25-cm^2 culture flasks containing approximately $5*10^6$ cells were used for each donor and each challenge. After challenge, RNA was isolated for SAGE and subsequent quantitative real-time PCR validation.

RNA isolation

Total RNA from the cell cultures was extracted using TRIzol reagent (GibcoBRL/Life technologies) according to the manufacturer's instructions. Enriched mRNA was obtained using the Oligotex mRNA Mini Kit (Qiagen Benelux BV, Venlo, The Netherlands) according to the manufacturer's instructions. After purification equal amounts of mRNA derived from the 7 donors were pooled to a final amount of 5 µg mRNA per challenge for the generation of SAGE libraries. The remainder of the purified mRNA samples was stored individually per donor at -80°C until further use.

Construction and analysis of Serial Analysis of Gene Expression (SAGE) libraries

In SAGE, two standard endonuclease reactions are subsequently performed to isolate short nucleotide sequences called tags. These tags are derived from a defined position at the 3'-end of each transcript flanked by the most 3'-end restriction site of *Nla*III. The second endonuclease specifically cleaves at a given distance of 14 bp from the *Nla*III restriction site, resulting in the release of the 10-14 bp tags. Isolated tags are paired tail-to-tail, ligated and cloned for automated sequencing. After sequencing, tags are easily recognized and extracted from the raw sequence data since each tag is flanked at one side by the *Nla*III restriction site.

SAGE was performed as described by Datson *et al.* [16] based on the original protocol developed by Velculescu *et al.* [9]. For the generation of each library, 5µg of mRNA was used. Double-stranded biotinylated cDNA was synthesized using Superscript Choice system (GibcoBRL/Life Technologies) and biotin-dT primers. In the first endonuclease reaction, double-stranded biotinylated cDNA was cleaved by *Nla*III (New England Biolabs, Beverly, MA), divided into two pools and bound to streptavidin-coated magnetic beads (Dynal Biotech, Oslo, Norway) for the attachment of linkers containing an

endonuclease restriction site for *Bsm*F1 and a PCR primer annealing site. After ligation of the linkers, the bound cDNA was cleaved in the second endonuclease reaction using the type IIS endonuclease BsmFI, thereby releasing the 10-14 bp tags. Released tags were blunt-ended, paired tail-to-tail and ligated. After PCR amplification of the ligation product, linker-sequences were released by cleavage with *Nla*III. The cleavage product containing the ditags was ligated into multimers and cloned into the SphI site of pZero 1.0 (Invitrogen/Life Technologies) and electroporated in ElectroMAX DH10B cells (GibcoBRL/Life Technologies). Colonies were transferred into 96-well plates containing Luria Broth/7.5% glycerol/Zeocin (50 µg/ml) medium, grown overnight at 37°C and stored at -80°C until further use. PCR products were sequenced on ABI377 and ABI3700 automated sequencers (Applied Biosystems, Foster City, CA) using BigDye terminator sequencing kit v2 (Applied Biosystems).

Sequenced clones were analyzed using the SAGE2000 v4.13 software kindly provided by K.W. Kinzler (John Hopkins Oncology Center, Baltimore, MD). Tags corresponding to linker sequences were discarded. SAGE tags derived from mitochondrial DNA were inferred from the human mitochondrial genome (RefSeq: NC_001807; [17]). Statistical analysis (Monte Carlo simulation) was performed using the SAGE2000 software. Differences in expression levels were determined at statistical significance levels of $P<0.05$ and $P<0.01$. For tag identification the libraries were compared with NCBI's reliable Unigene cluster to SAGE tag map (ftp://ncbi.nlm.nih.gov/pub/sage) [18] and with CGAP's SAGEgenie (http://cgap.nci.nih.gov/SAGE) [19]. Both databases were based on Unigene build#161. Combining these approaches decreased the degree of ambiguous mapping of SAGE tags.

To classify groups of genes showing similar changes in expression patterns across the libraries, self organizing maps (K-means clusters) of differentially expressed genes ($P<0.05$) were constructed using Spotfire Decision Site Software 7.1 (Spotfire, Göteborg, Sweden). Tag numbers were converted to relative expression values in order to visualize the change in expression rather than the absolute change in levels of expression.

For annotation of tags to biological processes and molecular function of genes, differentially expressed genes were matched to the Gene Ontology database (February 2004) using GoMiner [20] and AmiGO (http://www.godatabase.org/cgi-bin/amigo/go.cgi).

Quantitative real-time PCR (qPCR)

Quantitative real-time PCR (qPCR) was used to validate the SAGE data. For qPCR, samples derived from the same mRNA pool as used in SAGE were analyzed for expression of selected genes. For characterization and validation of normalization genes, mRNA samples from the individual donors were used to verify constant expression ratios of these genes in each sample. Single-stranded cDNA was synthesized using M-MLV Reverse Transcriptase primed with Oligo-dT (both from Invitrogen/Life Technologies, Breda, The Netherlands). Gene-specific primers were designed for two calcium-binding proteins

(S100A8, S100A9), for two proteinase inhibitors (Elafin and SLPI), and for members of the interleukin-1 family (IL-1β, IL-1RN, and IL-1F9). Primers were synthesized by Isogen (Maarssen, The Netherlands). To correct for differences in cDNA concentration of different samples, gene-specific primers for normalization genes (KRT6A, RPL5 and LMNA; see results for details) were used. Primer sequences and reaction conditions are listed in ***Table 1.*** qPCR analysis was performed on an iCycler PCR machine (BioRad, Hercules, CA) using SYBR green I chemistry [21]. Samples were analyzed in triplicate and threshold cycle numbers (C_T) were calculated using the iCycler v3.0a analysis software (BioRad, Hercules, CA). C_T values were used to calculate arbitrary mRNA concentrations using the relative standard curve method. The standard curve was generated in each reaction using a serial dilution of a cDNA sample containing message for the gene of interest. Relative mRNA concentrations for both the selected genes and normalization genes were determined and were used to calculate the expression ratios. Fold changes and standard errors of the mean for the validated genes were derived by averaging the ratios obtained from the independent normalizations for KRT6A, RPL5 and LMNA. Significance levels were determined at $P<0.05$ using a two-tailed paired Student's T-test.

Table 1: Sequences of primers for quantitative real-time PCR. Gene sequences used for primer design were retrieved from the ensembl website (http://www.ensembl.org). Primer sequences, annealing temperature and $MgCl_2$ concentration used in PCR reactions are listed.

Target	Ensembl ID	*Sense*	*Antisense*	Annealing T(°C)	MgCl2 (mM)
S100A8	ENSG00000143546	5'-TTT CCA TGC CGT CTA CAG-3'	5'-ACG CCC ATC TTT ATC ACC-3'	57	3
S100A9	ENSG00000163220	5'-GCT GGA ACG CAA CAT AGA G-3'	5'-GGT CCT CCA TGA TGT GTT C-3'	59	3,5
IL1β	ENSG00000125538	5'-CGA CAC ATG GGA TAA CGA-3'	5'-CGC AGG ACA GGT ACA GAT TC-3'	59	3
IL1RN	ENSG00000136689	5'-GTG CCT GTC CTG TGT CAA-3'	5'-TGT CTG AGC GGA TGA AGG-3'	63	3
IL1F9	ENSG00000136688	5'-TGG GAA TCC AGA ATC CAG-3'	5'-TTG GCA CGG TAG AAA AGG-3'	61	3.5
SLPI	ENSG00000124107	5'-CCA GGG AAG AAG AGA TGT TG-3'	5'-CCT CCA TAT GGC AGG AAT C-3'	61	2.5
PI3	ENSG00000124102	5'-CCG CTG CTT GAA AGA TAC TG-3'	5'-GAA TGG GAG GAA GAA TGG AC-3'	59	3
KRT6A	ENSG00000074729	5'-CTG AGG CTG AGT CCT GGT AC-3'	5'-GTT CTT GGC ATC CTT GAG G-3'	56	2.5
RPL5	ENSG00000122406	5'-TGG AGG TGA CTG GTG ATG-3'	5'-GCT TCC GAT GTA CTT CTG C-3'	63	3
LMNA	ENSG00000160789	5'-GTG GAG GAG GTG GAT GAG-3'	5'-ACC GGT AAG TCA GCA AGG-3'	63	3

RESULTS

SAGE expression profiles

Subcultures of human PBEC were stimulated for 6 hours with either a mixture of the pro-inflammatory cytokines IL-1β and TNFα or heat-inactivated *P. aeruginosa*. Expression profiles of stimulated cultured PBECs were constructed using SAGE. The data discussed in this publication have been deposited in the NCBI Gene Expression Omnibus [22] (GEO; http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE2056. In three libraries, a total number of 86.166 tags corresponding to 29.249 different transcripts were identified. ***Table 2*** shows the number of analyzed and unique tags in each library, including the number of duplicate dimers and linker tags encountered in each library. A relatively large number of tags (~75%) occurred only once in each library ***(Figure 1)***. Tags that were detected at least 4 times corresponded to 5-

Table 2: Numbers of analyzed and unique tags. Number of analyzed and unique tags in each library after discarding tags corresponding to linker sequences. For each library, the number of duplicate dimers and encountered linker tags are listed.

SAGE libraries	Tags analyzed	Unique tags	Diplicate dimers	Linkers tags
Control	28529	15020	493	281
IL1β/TNFα	28275	12436	1773	313
P. aeruginosa	28150	10204	503	618

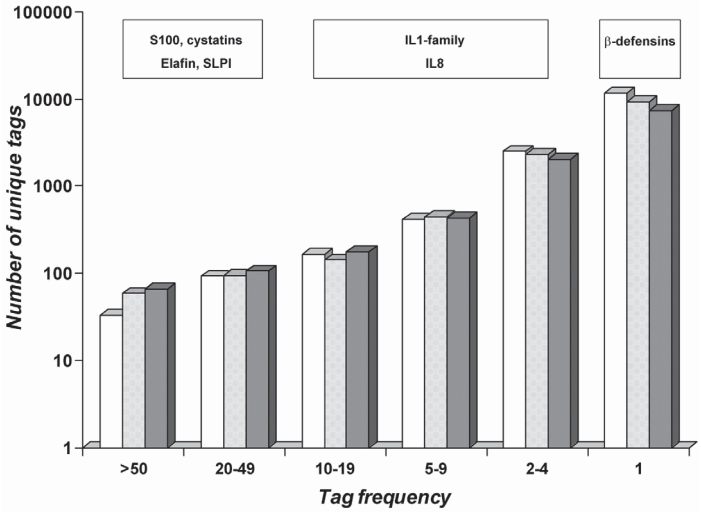

Figure 1: Tag frequency in the three libraries. Unique tags were classified into frequency classes based on their abundance in the three libraries. Open bars represent the frequency classes found in the control library, the grey bars the IL1β/TNFα library and the dark grey bars the *P. aeruginosa* library. Selected genes represented by corresponding tags in the different frequency classes are indicated at the top of the figure. Results show that transcripts encoding S100 proteins and proteinase inhibitors are abundant, those encoding cytokines intermediate, whereas β-defensin transcripts are rare.

8% of the unique tags, while contributing to approximately 50% of the total number of analyzed tags. Unique tags with an abundance of at least 50 copies contributed for approximately ~19% to the total number of sequenced tags. The difference in the number of unique tags between the control and challenge groups is explained by an increased number of tags that were expressed with at least 50 copies after exposure to IL-1β/TNFα or *P. aeruginosa* ***(Figure 1*** and ***Table 2)***.

The frequency distribution of tags in the libraries reflected the general pattern of gene expression in mammalian cells. Only a limited number of genes was expressed at high copy numbers (5-10% of all expressed genes), whereas the majority of genes displayed low levels of expression. For comparison, selected tags identified in the libraries are indicated in different frequency classes ***(Figure 1)***.

Tag identity was assigned using CGAP's SAGEgenie "Hs best gene to tag" and to NCBI's "UniGene cluster to SAGE tag reliable database" (NCBI SAGEmap). This allowed the assignment of 76.8% unique tags. The remaining 6,803 tags (23.2%) did not match to any genetic database. These unidentified tags might represent novel sequences expressed by PBEC. Most of these unidentified transcripts were expressed at low levels, as illustrated by the fact that 6,705 unidentified tags appeared only once, whereas a limited number of 98 tags showed expression levels of at least 3 tags in a single library.

Differentially expressed genes

Differential expression of tags between libraries was determined using the Monte Carlo simulation method. Using two cut-off values, a total number of 652 genes ($P<0.05$) and 292 genes ($P<0.01$) were found to be differentially expressed in one or more library comparisons. The number of genes found to be differentially expressed at a significance level of $P<0.01$ are shown in the Venn diagram in ***Figure 2***. The largest number of differentially expressed genes was found in the comparison of the control group versus the *P. aeruginosa* group. The number of genes showing different levels of expression between the IL-1β/TNFα library and the *P. aeruginosa* library is relatively low, most likely because these stimuli activate similar cellular processes. ***Table 3*** shows the number of differentially expressed genes increased or decreased in expression after challenge.

Among the 50 most differentially expressed genes, 45 were increased in expression, whereas only 5 genes were decreased in expression after *P. aeruginosa* or cytokine chal-

Table 3: Comparison of the SAGE libraries. Number of differentially expressed tags ($P<0.01$) subdivided into tags that were increased or decreased in expression.

Comparison	Differentially	Increased	Decreased
Control – IL1β/TNFα	152	102	52
Control – *P. aeruginosa*	234	155	79
IL1β/TNFα – *P. aeruginosa*	40	25	15

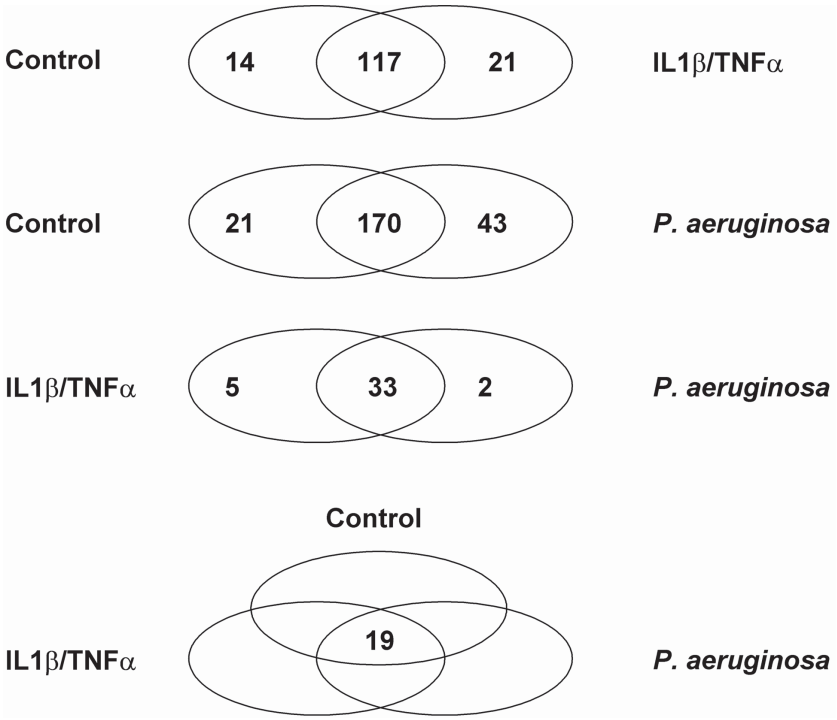

Figure 2: Venn-diagram of differentially expressed tags. Differentially expressed tags were subdivided into tags that were expressed in both libraries (overlapping areas) in a comparison or in one of the libraries (left or right ovals) in the comparison.

lenge. The 50 most differentially expressed genes sorted on tag abundance in the *P. aeruginosa* library are listed in ***Table 4***. Among these, several genes were present that encode for proteins that are directly or indirectly involved in epithelial defense, such as the S100 calcium binding proteins S100A6 and S100A9, the proteinase inhibitors elafin (PI3) and cystatin B (CSTB) and the cytokine interleukin-1 family member 9 (IL-1F9), a novel member of the IL-1 family of cytokines. Genes encoding structural components of the cytoskeleton that are also involved in differentiation including keratins, gamma actin (ACTG), small proline-rich protein 1B (SPRR1B), were also well represented in the top-50 list.

In order to structure the SAGE data, self organizing maps of differentially expressed genes ($P<0.05$) were constructed using the K-means clustering method. Nine clusters were created visualizing the change in direction of expression in groups of genes ***(Figure 3)***. The majority of differentially expressed genes was allocated to cluster 1. The pattern of expression shows that genes allocated to cluster 1 increased after IL-1β/TNFα exposure, which further increased after exposure with *P. aeruginosa*. This is in line with the data presented in ***Table 3***. Among the Top-50 list, 42 genes were allocated to K-means cluster 1 including S100A6, S100A9, elafin and FAU. Small proline-rich protein 2A

Table 4: 50 Most differentially expressed genes in P. aeruginosa treated PBEC. Tag counts of the 50 most differentially expressed genes in *P. aeruginosa*-treated PBEC. Listed from left to right: tag sequence, tag count in the control, IL1β/TNFα and *P. aeruginosa* SAGE libaries, Unigene accession (build#161), HUGO Gene Symbol, gene description and molecular function according to the Gene Ontology Consortium annotation.

Sequence	Control	IL1β/TNFα	*P. aeruginosa*	Unigene	SYMBOL	DESCRIPTION	GO: molecular function
GTGGCCACGG	121	419	668	Hs.112405	S100A9	S100 calcium binding protein A9 (calgranulin B)	calcium ion binding;signal transducer activity
CCCCCTGGAT	61	133	283	Hs.275243	S100A6	S100 calcium binding protein A6 (calcyclin)	calcium ion binding;cyclin-dependent protein kinase intrinsic regulator activity;growth factor activity;protein binding
CCCATCGTCC	134	188	213	Hs.193989	TARDBP	TAR DNA binding protein	RNA binding;microtubule binding;transcription factor activity
TGTGTTGAGA	73	141	205	Hs.422118	EEF1A1	eukaryotic translation elongation factor 1 alpha 1	protein-synthesizing GTPase activity\, elongation
GCCCCTGCTG	96	180	203	Hs.433845	KRT5	keratin 5	structural constituent of cytoskeleton
GCCGAGGAAG	32	83	159	Hs.434029	RPS12	ribosomal protein S12	RNA binding;structural constituent of ribosome
GGGGAAATCG	48	131	152	Hs.76293	TMSB10	thymosin, beta 10	actin modulating activity
TGGTGTTGAG	41	91	133	Hs.275865	RPS18	ribosomal protein S18	rRNA binding;structural constituent of ribosome
ATGAGCTGAC	64	135	129	Hs.695	CSTB	cystatin B (stefin B)	cysteine proteinase inhibitor activity
ATGGCTGGTA	38	82	126	Hs.406341	RPS2	ribosomal protein S2	RNA binding;structural constituent of ribosome
TAAACCTGCT	40	68	125	Hs.99923	LGALS7	lectin, galactoside-binding, soluble, 7 (galectin 7)	sugar binding
TTGAATCCCC	30	80	121	Hs.112341	PI3	proteinase inhibitor 3, skin-derived (SKALP)	peptidase activity;protein binding;serine proteinase inhibitor activity
GATCCCAACT	12	55	108	Hs.118786	MT2A	metallothionein 2A	helicase activity;zinc ion binding
CCCTTGAGGA	27	56	100	Hs.1076	SPRR1B	small proline-rich protein 1B (cornifin)	structural molecule activity

AAGGTGGAGG	23	64	99	Hs.337766	RPL18A	ribosomal protein L18a	RNA binding;structural constituent of ribosome
CCGTCCAAGG	16	54	96	Hs.397609	RPS16	ribosomal protein S16	RNA binding;structural constituent of ribosome
GCGACCGTCA	30	61	89	Hs.273415	ALDOA	aldolase A, fructose-bisphosphate	fructose-bisphosphate aldolase activity
AGCAGGAGCA	26	54	89	Hs.8182	MGC17528	hypothetical protein MGC17528	calcium ion binding
CCCGTCCGGA	27	60	87	Hs.431392	RPL13	ribosomal protein L13	RNA binding;structural constituent of ribosome
CGCCGGAACA	32	66	86	Hs.286	RPL4	ribosomal protein L4	RNA binding;structural constituent of ribosome
AGCTCTCCCT	36	71	84	Hs.82202	RPL17	ribosomal protein L17	RNA binding;structural constituent of ribosome
CTAGCCTCAC	18	56	83	Hs.14376	ACTG1	actin, gamma 1	motor activity;structural constituent of cytoskeleton
GGGCTGGGGT	20	60	82	Hs.430207	RPL29	ribosomal protein L29	RNA binding;heparin binding;structural constituent of ribosome
GACGACACGA	16	40	79	Hs.153177	RPS28	ribosomal protein S28	RNA binding;structural constituent of ribosome
CTGTCACCCT	9	25	75	Hs.355542	SPRR2A	small proline-rich protein 2a	molecular_function unknown;structural molecule activity
GCAGCCATCC	21	42	74	Hs.356371	RPL28	ribosomal protein L28	RNA binding;structural constituent of ribosome
GGCCGCGTTC	12	45	74	Hs.439420	RPS17	ribosomal protein S17	RNA binding;structural constituent of ribosome
GGGGCAGGGC	18	66	73	Hs.406397	GFAP	glial fibrillary acidic protein	structural constituent of cytoskeleton
CGAATGTCCT	21	60	65	Hs.432677	KRT6B	keratin 6B	structural constituent of cytoskeleton
AGCACCTCCA	11	34	60	Hs.75309	EEF2	eukaryotic translation elongation factor 2	GTP binding;translation elongation factor activity
AGGTCCTAGC	9	41	59	Hs.226795	GSTP1	glutathione S-transferase pi	glutathione transferase activity
AATCCTGTGG	18	29	57	Hs.178551	RPL8	ribosomal protein L8	RNA binding;structural constituent of ribosome
CTCAACATCT	9	34	55	Hs.136470	RPLP0	ribosomal protein, large, P0	RNA binding;structural constituent of ribosome

AGAAAGATGT	102	62	52	Hs.78225	ANXA1	annexin A1	calcium ion binding;calcium-dependent phospholipid binding;phospholipase A2 inhibitor activity;receptor binding
GCCTACCCGA	19	42	52	Hs.23582	TACSTD2	tumor-associated calcium signal transducer 2	receptor activity
TGGGCAAAGC	14	34	49	Hs.256184	EEF1G	eukaryotic translation elongation factor 1 gamma	translation elongation factor activity
AGGAAAGCTG	14	37	47	Hs.433411	RPL36	ribosomal protein L36	structural constituent of ribosome
GTTCCCTGGC	6	20	46	Hs.177415	FAU	Finkel-Biskis-Reilly murine sarcoma virus	RNA binding;structural constituent of ribosome
CCAGTGGCCC	7	16	41	Hs.180920	RPS9	ribosomal protein S9	RNA binding;structural constituent of ribosome
TACCATCAAT	8	34	38	Hs.169476	GAPD	glyceraldehyde-3-phosphate dehydrogenase	glyceraldehyde 3-phosphate dehydrogenase (phosphorylating) activity
TGGCCCCACC	7	32	34	Hs.198281	PKM2	pyruvate kinase, muscle	magnesium ion binding;pyruvate kinase activity
GTTAACGTCC	3	15	28	Hs.178391	RPL36A	ribosomal protein L36a	
GTCTGGGGCT	1	12	22	Hs.406504	TAGLN2	transgelin 2	actin binding
GTGCGGAGGA	1	18	21	Hs.332053	SAA1	serum amyloid A1	
CAGCTATTTC	55	21	18	Hs.408061	FABP5	fatty acid binding protein 5 (psoriasis-associated)	fatty acid binding;transporter activity
ACCTGGAGGG	0	12	16	Hs.2853	PCBP1	poly(rC) binding protein 1	RNA binding;single-stranded DNA binding
ACTAGCACAG	0	7	15	Hs.211238	IL1F9	interleukin 1 family, member 9	interleukin-1 receptor antagonist activity
CTGCCCAGTG	27	3	2	Hs.169793	RPL32	ribosomal protein L32	RNA binding;structural constituent of ribosome
GAGGACCTGG	66	1	1	Hs.250469	LoopADR	aldo-keto reductase loopADR	aldehyde reductase activity
CTTTCCCTCA	15	2	0	Hs.184510	SFN	stratifin	protein domain specific binding;protein kinase C inhibitor activity

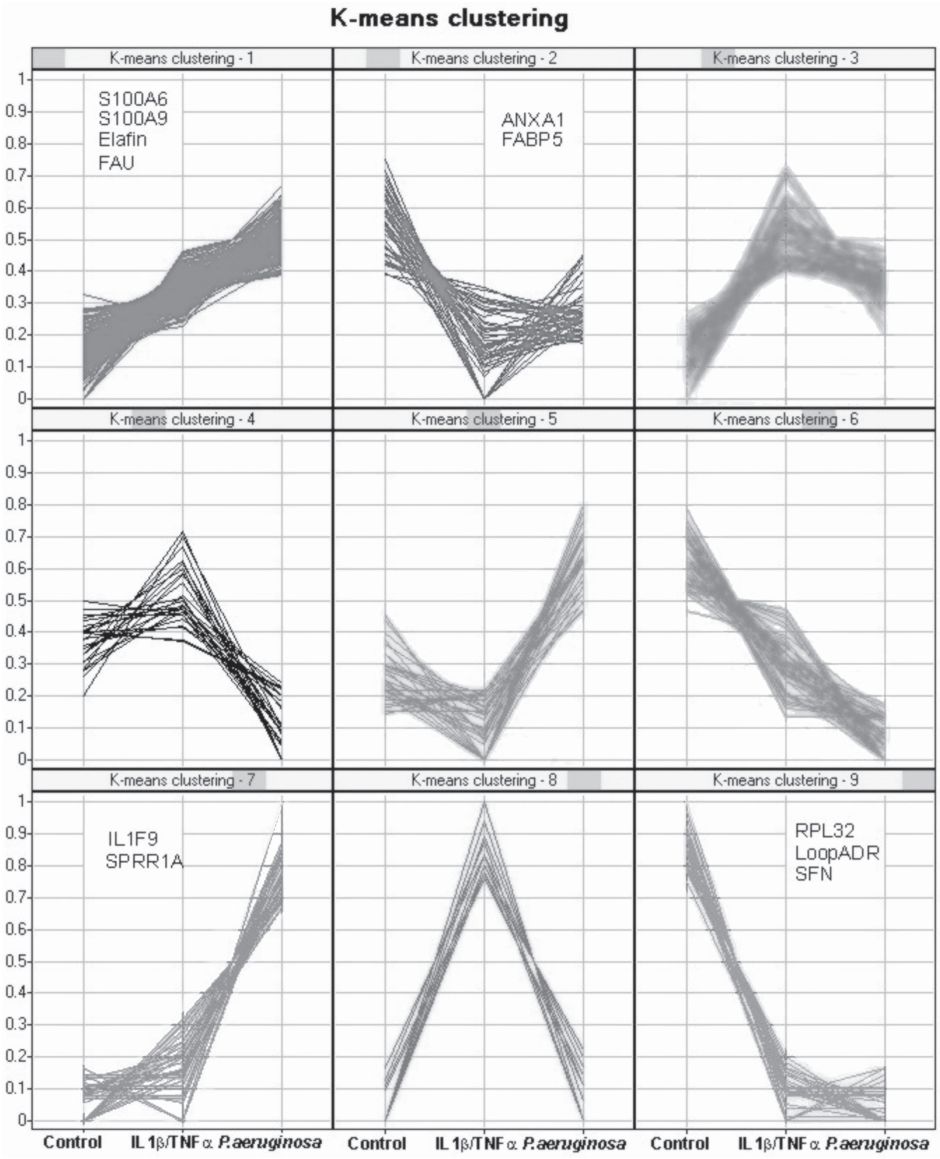

Figure 3: K-means clustering of differentially expressed genes (P<0.05). Eleven genes derived from the top-50 list are indicated in their corresponding cluster. The data shown indicates the change in direction of expression between control and challenge groups.

(SPRR2A) and IL-1F9 were allocated to cluster 7 since there was little or no expression in the control group and cytokine group, whereas high levels of expression were found in the *P. aeruginosa* group, indicating that these genes might be specific for *P. aeruginosa*-initiated response in PBEC. Three of the five downregulated genes, ribosomal protein L32 (RPL32), aldo-keto reductase loop ADR (LoopADR) and stratifin (SFN) were allocated to cluster 9. In this cluster, gene expression was decreased to a similar extent after both

IL-1β/TNFα and *P. aeruginosa* exposure. The two remaining downregulated genes, annexin A1 (ANXA1) and fatty acid binding protein 5 (FABP5) were allocated to cluster 2. The general expression pattern of genes allocated to cluster 2 showed high expression in the control group, low levels of expression in the IL-1β/TNFα group and moderate expression in the *P. aeruginosa* group.

Classification of differentially expressed genes

Differentially expressed genes after *P. aeruginosa* exposure were annotated to their corresponding biological processes using Gene Ontology. This analysis showed that the response of epithelial cells is characterized by the expression of genes encoding for proteins involved in biological processes such as metabolism, cell growth/maintenance, development, cell communication and response to stimulus ***(Figure 4)***. The cell growth/maintenance category refers to genes that exert molecular functions that are related to cell cycle, cell adhesion and cytoskeletal architecture.

Following this functional analysis of the gene expression profiles, we selected four classes/families of genes that were affected (P<0.01) in stimulated PBEC: keratins, proteinase inhibitors, interleukin-1 family members and S100 calcium-binding proteins ***(Table 5)***. The best represented family of genes involved in development in our data is the family of keratins. These genes encode for structural component proteins of epithelial cells. With 19 members expressed, representing 3.5% of all sequenced tags, and with

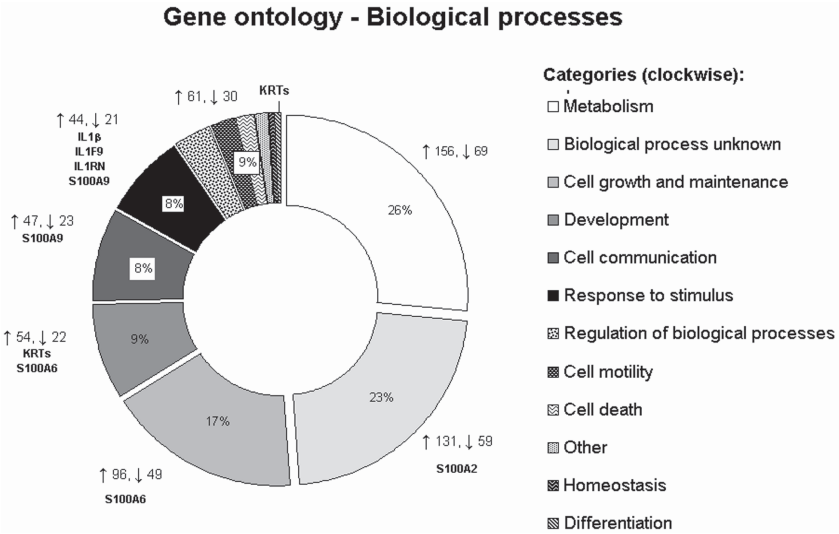

Figure 4: Gene Ontology: Biological processes. Differentially expressed genes in PBEC after *P. aeruginosa* exposure were subjected to gene ontology annotation to identify the corresponding biological processes. The major biological processes, represented by wedges are shown, including the percentage of differentially expressed genes annotated to the biological process. The number of genes that is increased or decreased for each of the processes is indicated outside each wedge. The smallest categories are singled out and summed. Redundancy in gene ontology exists due to the fact that genes may be involved in multiple biological processes.

Table 5: Classification of differentially expressed genes Tag counts of the 4 identified genes families that were affected by *P. aeruginosa*. Listed from left to right: tag sequence, HUGO Gene Symbol, tag count in the control, IL1β/TNFα and *P. aeruginosa* SAGE libaries, Unigene accession (build#161), RefSeq accession and molecular function according to the Gene Ontology Consortium annotation. The expression of bold printed genes was validated by qPCR. Multiple transcript variants exist for marked (*) Genbank accessions.

Tag	Symbol	Control	IL1β/TNFα	*P. aeruginosa*	Unigene	Genbank	Molecular Function (GO)
Keratins							
GCCCCTGCTG	KRT5	96	180	203	Hs.433845	NM_000424	structural constituent of cytoskeleton
CGAATGTCCT	KRT6B	21	60	65	Hs.432677	NM_005555	structural constituent of cytoskeleton
GATGTGCACG	KRT14	12	26	35	Hs.355214	NM_000526	structural constituent of cytoskeleton, structural constituent of epidermis
CTTCCTTGCC	KRT17	181	216	253	Hs.2785	NM_000422	structural constituent of cytoskeleton
GACATCAAGT	KRT19	20	48	45	Hs.182265	NM_002276	structural constituent of cytoskeleton
Proteinase inhibitors							
TGTGGGAAAT	**SLPI**	**18**	**40**	**47**	**Hs.251754**	**NM_003064**	**serine proteinase inhibitor activity**
TTGAATCCCC	**Elafin**	**30**	**80**	**121**	**Hs.112341**	**NM_002638**	**peptidase activity;protein binding ;serine proteinase inhibitor activity**
ATGAGCTGAC	CSTB	64	135	129	Hs.695	NM_000100	cysteine protease inhibitor activity
Interleukin 1 family							
CAATTTGTGT	**IL1β**	**2**	**12**	**15**	**Hs.126256**	**NM_000576**	**interleukin-1 receptor activity**
ACTCGTATAT	**IL1RN**	**2**	**3**	**14**	**Hs.81134**	**NM_000577***	**interleukin-1 receptor activity**
ACTAGCACAG	**IL1F9**	**0**	**7**	**15**	**Hs.211238**	**NM_019618**	**interleukin-1 receptor activity**

S100 proteins							
GATCTCTTGG	S100A2	123	170	189	Hs.38991	NM_005978	calcium ion binding
CCCCCTGGAT	S100A6	61	133	283	Hs.275243	NM_014624	calcium ion binding;cyclin-dependent protein kinase\, intrinsic regulator activity;growth factor activity; protein binding
TACCTGCAGA	**S100A8**	**139**	**214**	**158**	**Hs.416073**	**NM_002964**	**calcium ion binding**
GTGGCCACGG	**S100A9**	**121**	**419**	**668**	**Hs.112405**	**NM_002965**	**calcium ion binding;signal transducer activity**
Normalization genes							
AAAGCACAAG	**KRT6A**	**267**	**252**	**250**	**Hs.367762**	**NM_005554**	**structural constituent of cytoskeleton**
CTGCTATACG	**RPL5**	**13**	**11**	**12**	**Hs.180946**	**NM_000969**	**5S RNA binding;rRNA binding; structural constituent of ribosome**
GGAGGGGGCT	**LMNA**	**10**	**9**	**9**	**Hs.377973**	**NM_170707***	**structural molecule activity**

five members among the most differentially expressed genes (P<0.01), the family of keratins is likely to play a crucial role in forming and maintaining the physical epithelial barrier. In addition to keratins involved in barrier formation, epithelial cells expressed large amounts of transcripts encoding proteinase inhibitors including the serine proteinase inhibitors SLPI and elafin (PI3/SKALP) and the cysteine proteinase inhibitor CSTB. Increased release of proteinase inhibitors has been associated with inflammation. These molecules protect the airway epithelium from proteinase activity of endogenous and exogenous proteinases. Roughly 2.5% of all sequenced tags encoded for various proteinase inhibitors under control conditions increasing to up to 3.6% of all tags after *P. aeruginosa* exposure. Assigning biological processes to these molecules using Gene Ontology remained difficult since only a limited number of proteinase inhibitors have been correlated to a biological process.

Among the immune signaling molecules, a number of cytokines and chemokines is expressed. The most abundantly expressed cytokines are the members of the interleukin-1 family. According to GO, these molecules are involved in the biological process "response to stimulus". Six out of ten members were found to be expressed by bronchial epithelial cells of which three were differentially expressed (P<0.01) including IL-1β, IL-1RN and IL-1F9, a novel member of this family.

Finally, S100 calcium-binding proteins were abundantly expressed by PBEC and increased after microbial exposure. These molecules exert diverse functions, including antimicrobial activity, and involvement in differentiation and intracellular signaling [23,24]. In total, 13 out of 21 members of this family were encountered. SAGE tags for S100A2, S100A6, S100A8 and S100A9, were detected at high frequencies in our SAGE libraries and were also among the 292 most differentially expressed genes (P<0.01, ***Table 4***). With 668 tags in the SAGE library of *P. aeruginosa* challenged PBEC, the tag corresponding to S100A9 was the most frequently encountered tag in the three SAGE libraries. In total, these four families of genes contributed for 13% to the total number of sequenced tags in the SAGE library of the *P. aeruginosa* challenge group.

Generally, genes involved in metabolism were abundantly expressed and the expression fluctuated upon changes in the environment. Indeed, a large proportion of differentially expressed genes were classified to be involved in metabolism. In total, 17 tags corresponding to transcripts encoding for ribosomal proteins were found among the 50 most differentially expressed genes in PBEC after *P. aeruginosa* exposure. In addition, the Gene Ontology analysis showed that 26% of the differentially expressed genes were correlated to processes related to metabolism. The general trend in expression was that the number of tags encoding for ribosomal proteins increased after stimulation with IL-1β/TNFα and increased more markedly after *P. aeruginosa* challenge. The overall increase in expression of ribosomal proteins suggests an increased protein synthesis after challenge with pro-inflammatory cytokines and *P. aeruginosa*.

Validation of SAGE results by quantitative real-time PCR (qPCR)

The SAGE results were validated using quantitative real-time PCR. Primers were designed for selected genes that were found to be differentially expressed by SAGE. For normalization of the qPCR data, we used a panel of three endogenous control genes instead of a single normalization gene. All of the commonly used normalization genes such as β-actin (ACTB) or glyceraldehyde 3-phosphate dehydrogenase (GAPDH) showed variable levels of expression under the experimental conditions or were not expressed at all. We selected 3 genes from our SAGE libraries that showed similar levels of expression for use as normalization genes: keratin 6A (KRT6A), ribosomal protein L5 (RPL5) and lamin A/C (LMNA) ***(Table 5)***. The expression of KRT6A, RPL5 and LMNA was assessed by qPCR in 24 cDNA samples. These samples were derived from PBEC of the 7 individual donors, separately exposed to medium, cytokines or *P. aeruginosa* and the corresponding pooled RNA samples from the 7 donors (as used for SAGE). The results showed minimal variation in the ratios of the three normalization genes used (KRT6A/LMNA, KRT6A/RPL5 and LMNA/RPL5; ***Figure 5***), indicating that these genes were suitable for use as normalization genes in PBEC exposed to the stimuli used in this study.

In general, the change in direction and pattern of expression of the selected genes after exposure to *P. aeruginosa* or IL-1β/TNFα was observed using both SAGE and qPCR ***(Figure 6)***. The best correlation between SAGE and qPCR data was observed for those genes that were abundantly expressed such as S100A8, S100A9, SLPI and elafin. For low abundant transcripts such as the IL-1 family members, the observed change in direction of expression correlated well between SAGE and qPCR. However, the magnitude of the response showed more variation.

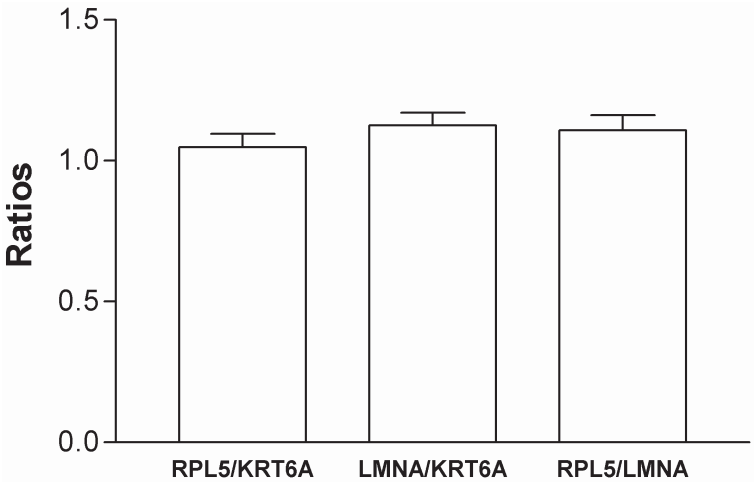

Figure 5: Ratios of normalization genes. Three genes were used to accurately normalize the qPCR data. Each bar represents the mean expression ratio of a pair of normalization genes in the 24 samples tested. Variation in expression ratios is displayed as standard error of the mean.

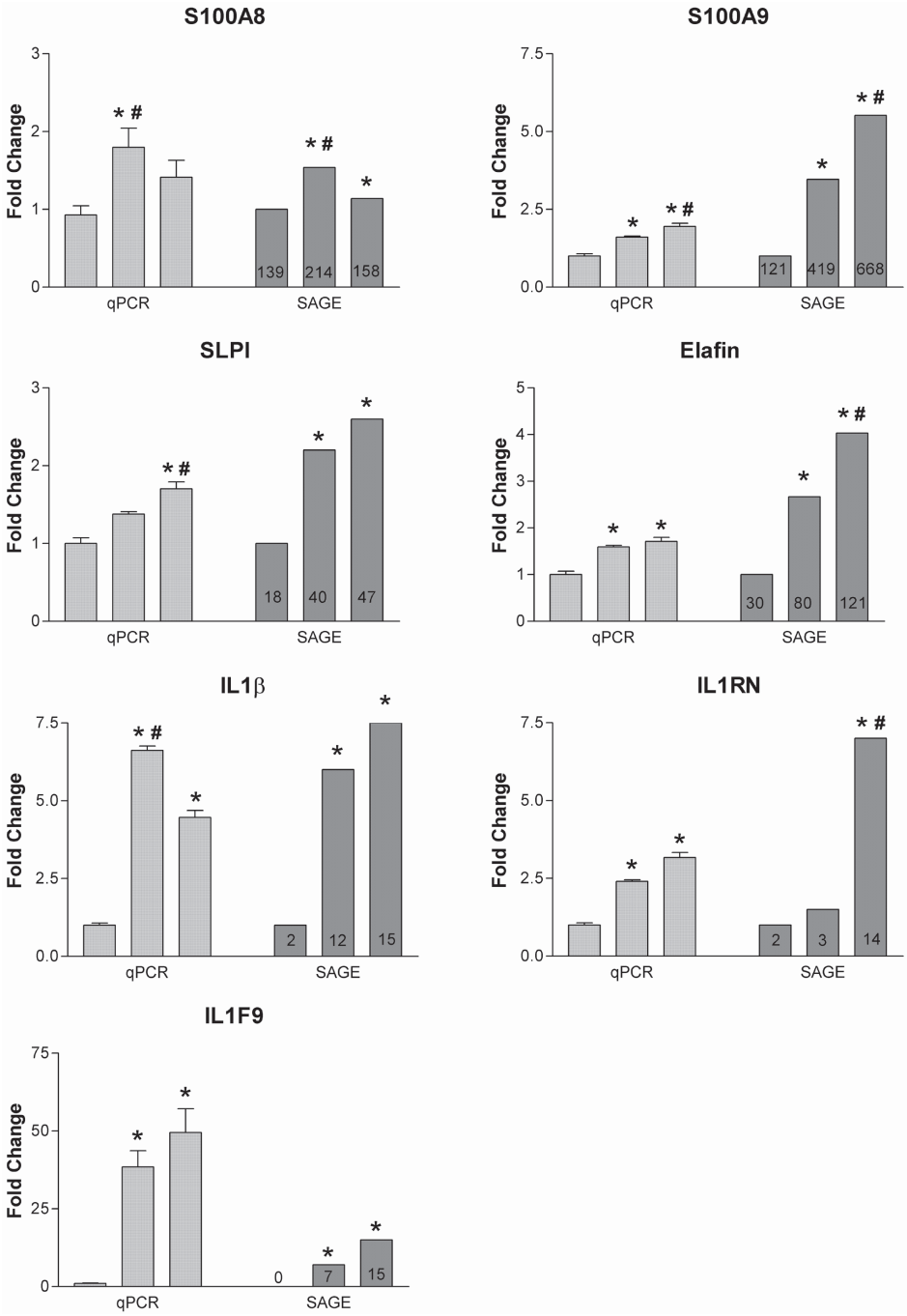

Figure 6: Validation of 7 target sequences using quantitative real-time PCR. From left to right are depicted the expression levels of control, IL-1β/TNFα and *P. aeruginosa*, respectively, as determined by qPCR and SAGE. Statistical significance was determined at $P < 0.05$. *Significant difference in expression between control and challenge groups. #Statistical difference between challenge groups.

DISCUSSION

The innate immune system of the airways is a complex and dynamic system that continuously adapts itself to the changing environment to prevent infection by respiratory pathogens. Although epithelial innate immunity gained renewed interest over the past two decades, little is still known about the underlying molecular mechanisms that lead to the onset of the host defense response in epithelial cells upon microbial exposure. In the present study, genes affected by IL-1β/TNFα and *P. aeruginosa* exposure were identified using SAGE. Differentially expressed genes in PBEC both after IL-1β/TNFα and *P. aeruginosa* exposure were found to be mainly implicated in biological processes such as metabolism, cell growth/maintenance, development and response to stimulus. In this study, we show that at least four families of genes are involved in the epithelial response to the pro-inflammatory cytokines IL-1β and TNFα or *P. aeruginosa:* keratins, proteinase inhibitors, IL-1 family members and S100 calcium-binding proteins. The expression profiles were validated using quantitative real-time PCR. To our knowledge, this is the first study in which a large-scale expression profiling technique was used to assess the epithelial response to pro-inflammatory cytokines or heat-inactivated microorganisms, providing a comprehensive view on the epithelial response to these stimuli at the mRNA level of thousands of transcripts simultaneously.

Previously, two expression profiling studies have been published using various lung-derived epithelial cell lines that were exposed to microorganisms. Inherent to the use of tumor or transformed (immortalized) cell lines is that these cells may respond differently as compared to primary cells such as the cells used in our study. In the study of Ichikawa *et al.* [25], the lung carcinoma cell line A549 was exposed to *P. aeruginosa* over a period of 1-3 hours after which changes in expression profiles were determined. This study showed that interferon regulatory factor 1 (IRF-1) is essential in the host defense response of A549 cells to the microorganism. A substantial drawback of this study is that a custom-made microarray was used that contained a limited number of 1.506 probe sets of which the identity has not been made publicly available. Furthermore, this microarray was developed in an earlier study to assess transcriptional changes in $CD4^+$ T cells after HIV-infection [26]. It is therefore not clear to what extent the probe sets present on this chip are relevant for studying epithelial gene expression. In the other published expression profiling study, Belcher *et al.* [27] assessed the transcriptional changes in the SV40-virus transformed bronchial epithelial cell line BEAS-2B upon exposure to *Bordetella pertussis* using an Affymetrix HU6800 microarray containing approximately 6.800 probe sets. Transcriptional changes were assessed one and three hours after *B. pertussis* exposure. The response after three hours was most robust in this model system, whereas after 1 hour exposure minimal changes in the transcriptome of *B. pertussis*-infected BEAS-2B cells were observed. At three hours, a modest number of 33 genes showed a relative

change of ± 3-fold in expression, containing cytokines such as among others IL-1β and IL-6 and chemokines such as IL-8, CCL2/MCP-1, CXCL1/Groα and CXCL2/Groβ. A number of the genes that were identified by Belcher *et al.* to be important in the defense against *B. pertussis* were also detected in our analysis. However, it should be noted that the levels of expression of these genes are close to the detection limit that can be reliably measured by large-scale gene expression profiling methods such as SAGE and microarray. These studies support our conclusion that substantial changes in the transcriptome of epithelial cells occur at least after 3-6 hours after microbial exposure. However, our study differs from these studies because of the use of primary bronchial epithelial cells instead of immortalized or tumor cell lines and the choice of expression profiling technique.

One of the main advantages of SAGE is its open-ended approach of assessing gene expression in an unbiased fashion. This was demonstrated in a similar *in vitro* model of epithelial inflammation using TNFα–stimulated keratinocytes [11]. This study revealed novel insights into cornification and the barrier function of the skin under inflammatory conditions. Using SAGE, the proteinase inhibitor cystatin M/E (CST6) was identified to be one of the key genes that was overexpressed in TNFα–stimulated keratinocytes. Previously, this gene had not been associated with barrier formation in this cell type. Further investigation unveiled a central role for cystatin M/E in keratinization of the skin [28,29]. Comparison of the SAGE results of the keratinocyte study and the present study showed a marked similarity in the response of airway epithelial cells and keratinocytes regarding barrier function and host defense under inflammatory conditions. For both cell types, highly expressed classes of genes were keratins, proteinase inhibitors and S100 proteins. The epithelial antimicrobial proteinase inhibitors SLPI and elafin were found at high levels in both keratinocytes and PBEC. Among the S100 calcium binding proteins, the tag corresponding to S100A9 was the most abundantly expressed S100 protein in both keratinocytes and PBEC.

Also in the present study, SAGE revealed the transcription of numerous genes that were previously not known to be expressed by bronchial epithelial cells after exposure to pro-inflammatory cytokines or *P. aeruginosa*. A selection of the 292 differentially expressed genes could be classified into four processes based on their function in innate immunity. First, barrier formation is an important feature of innate immunity. Genes encoding for structural components of the cytoskeleton are abundantly expressed and are involved in assembly and disassembling of the cytoskeleton, controlling cell size and shape, cell-cell or cell-matrix interactions. Keratin family members, that form intermediate filaments in epithelial cells (reviewed in [30]), were frequently encountered among the list of differentially expressed genes. Co-expression of a number of keratins such as KRT5 and KRT14 [31] that was previously reported in skin was also observed in bronchial epithelial cells. In the skin, mutations in KRT5 and KRT14 cause a number of skin disorders [32]. Whether these mutations affect the bronchial mucosa is unclear. Co-expression of

KRT6 isoforms (except for KRT6A which is expressed at steady-state levels) and KRT17 was increased after stimulation, whereas the expression of KRT16 was not affected. Co-expression of KRT6 isoforms, KRT16 and KRT17 has been associated with wound healing after cutaneous skin injury [33]. These results suggest that keratins are important players in the recovery process of epithelial cells after microbial insult. Furthermore, these data indicate similarities between the regulation of certain keratins in bronchial airway epithelial cells and keratinocytes.

The second class of differentially expressed genes consists of genes encoding proteinase inhibitors. These molecules are produced to protect the epithelium from enzymatic attack by microbial or endogenous host proteinases. The protective role of elafin for epithelial tissues has been described previously [34]. Proteinase inhibitors including SLPI, elafin and cystatin family members were expressed at high levels under control conditions and were increased in expression upon stimulation. Interestingly, SLPI and elafin also display antimicrobial activity against a range of microorganisms including *P. aeruginosa* [35-37] and may be involved in proliferation and repair of epithelial tissues [35,38].

Our finding that several IL-1 gene family members were expressed upon exposure of bronchial epithelial cells to pro-inflammatory cytokines and a microbial stimulus is interesting. Differential expression was observed for three IL-1 family members (IL-1β, IL-1RN and IL-1F9; $P<0.01$). Both IL-1β and IL-1 receptor antagonist (IL-1RN) exert their function through the IL-1 type I receptor (reviewed in [39] and [40]). Recently, a novel IL-1 signaling cascade has been described that employs a novel IL-1 receptor, the IL-1 receptor related protein 2 (IL-1Rrp2) and two novel IL-1 family members, IL-1F5 and IL-1F9 [41]. Binding of IL-1F9 to IL-1Rrp2 leads to activation of NFκB[42] while IL-1F5 functions as antagonist of the receptor [41]. The inducible expression of IL-1F9 in bronchial epithelium as observed in the present study is remarkable since expression of this tag, as determined by SAGE, has only been detected in three other tissue types at very low levels. This finding suggests a role for both IL-1β and IL-1F9 signaling in bronchial epithelial cell function, which may allow fine-tuning of the onset and boost of the inflammatory response.

S100 calcium-binding proteins form the fourth group of differentially expressed genes. These proteins exert a wide range of functions including signal transduction, cell-cell communication, regulation of cell cycle, cell differentiation, gene transcription and antimicrobial activities [43]. Four members of the S100 family, S100A2, A6, A8 and A9 were found to be differentially expressed ($P<0.01$). The SAGE tag corresponding to S100A9 was the most abundant transcript detected in both the IL-1β/TNFα library and *P. aeruginosa* library. This protein forms a heterodimer with S100A8 in a calcium-dependent manner. The complex, also known as cystic fibrosis antigen or calprotectin, displays antimicrobial activity that has been ascribed to its ability to chelate Zn^{2+} [44,45], thereby depriving the micro-environment of zinc-ions that are essential for microbial growth. A prominent role for S100A8/A9 in inflammation has been suggested since the complex is

abundantly detected in a number of inflammatory conditions such as abscesses, cystic fibrosis, chronic bronchitis [46] and Crohn's disease [47]. S100A2 may function as a tumor suppressor gene, although this is a matter of debate since both a decrease [48-50] as well as an increase in expression [51-53] of this gene has been observed in various types of tumors. Alternatively, Zhang *et al.* have suggested that the primary function of S100A2 is the participation in the oxidative stress response [54]. For S100A6, it has been demonstrated that this protein is involved in the cytoskeletal organization, morphology and the regulation of proliferation of lung fibroblasts in vitro [55]. Additionally, S100A6 may be involved in tumor progression since the levels of expression of S100A6 are proportional with the increase of malignancy as was demonstrated in human colon tissue [56].

Remarkably, expression of genes encoding antimicrobial peptides was limited, which was also demonstrated by the absence of expression of these molecules in the studies of Belcher *et al.* [27] and Ichikawa *et al.* [25]. This suggests that these molecules may not predominate in the early host defense response by bronchial epithelial cells. The most abundant transcript encoding an antimicrobial peptide in our data set, FAU, was found to be expressed at increased levels by PBEC after stimulation with IL-1β/TNFα or *P. aeruginosa* whereas little expression was found under control conditions ***(Table 4)***. Posttranslational modification of the FAU gene product yields the antimicrobial peptide ubiquicidin [57]. Among the antimicrobial defensins, human β-defensin 2 (DEFB4), was detected in our SAGE study. However, the levels of expression for the tag corresponding to DEFB4 were at the lower detection limit of SAGE. Using qPCR, we confirmed the absence of DEFB4 expression in PBEC under control conditions and observed an induction of expression after IL-1β/TNFα or *P. aeruginosa* exposure (data not shown).

To validate our SAGE results, qPCR was performed on selected target sequences. To optimize the accuracy of validation by qPCR, we carefully selected three genes for normalization to correct for differences in cDNA input. Widely used normalization genes such as GAPDH and β-actin showed variable expression levels in our SAGE libraries. Based on the SAGE data, KRT6A, RPL5 and LMNA were selected for normalization. Accurate normalization of qPCR data is essential for detecting small differences in gene expression or studying differential expression of low abundant genes and can be achieved by using multiple internal control genes, an approach described by Vandesompele *et al*. [58]. The best correlation between qPCR and SAGE was observed for genes that were represented by large tag counts such as S100A8 and S100A9. In contrast, whereas the direction of change in expression of low abundant genes such as IL-1β and IL-1F9 was the same using SAGE and qPCR data, the magnitude of the effect determined by SAGE and qPCR differed slightly. Since the SAGE technique is based on random sampling of transcripts, the more copies of the same SAGE tag is encountered, the more accurate the fold change in expression can be estimated. In addition, the more frequent a SAGE tag is encountered, the better the result can be reproduced using independent detec-

tion techniques, such as qPCR. Therefore, low copy numbers of SAGE tags can be used to predict the change of direction in gene expression, but should not be qualified as quantitative [59].

Experimental data acquired by using large-scale gene expression profiling methods such as SAGE provide a wealth of information. A drawback of these techniques is that very large quantities of starting material (microgram range of mRNA) are required to perform the gene expression analysis. Acquiring large amounts of RNA is problematic in many model systems and therefore immortalized or tumor cell lines are often used for large-scale gene expression profiling studies. Instead of using immortalized or tumor cell lines, we used subcultures of primary bronchial epithelial cells because these cells are more likely to reflect the cells lining the airways. Furthermore, to limit donor variation in the expression profiles, cells derived from seven different donors were used. Generated expression profiles were validated on the on the pooled material as used in SAGE as well as on individual donors.

SAGE was chosen as large-scale gene expression profiling method since this technique provides the ability of gene discovery. This advantage is illustrated by the observation that 6,803 tags could not be annotated. These tags may represent novel sequences including transcripts encoding for antimicrobial peptides, cytokines and chemokines. Further investigation of these unknown tags may lead to the discovery of previously unidentified genes or single nucleotide polymorphisms (SNP) in genes that are involved in the epithelial defense. Since mRNA of seven donors was used for the generation of our libraries, we can not exclude the possibility that SNP noise has been introduced in our libraries. SNPs may in part account for a number of no match tags [60]. In addition to SNP noise, unmatched tags could also have been generated due to technical limitations inherent to the SAGE technology. For example, internal priming of transcripts in the construction of SAGE libraries may result in contamination of libraries with tags that are not derived from the most 3' restriction site of *NlaIII*. Based on the Ludwig Transcript Viewer, provided with SAGEgenie [19], it is estimated that approximately 10% of unique tags is derived from internally primed transcripts as determined for the 652 differentially expressed genes ($P<0.05$). Furthermore, sequencing errors may have introduced false tags into the SAGE libraries that do not represent the original mRNA sample. However, it has been demonstrated that the number of tags resulting from sequencing errors is limited [61,62].

From our data and the literature, we hypothesize that the transcriptional response of epithelial cells upon stimulation with pro-inflammatory cytokines or microorganisms is characterized by two events that proceed subsequently. First, epithelial cells respond to pro-inflammatory cytokines and microbial exposure by affecting the expression of genes that are mainly involved in strengthening the physical epithelial barrier as demonstrated by the increased expression of keratins, proteinase inhibitors and the calcium-

binding proteins S100A8/A9. Furthermore, epithelial cells are urged into a repair process to limit damage caused by the microbial insult. To provide host defense during the early response, defense strategies such as depriving the micro-environment of essential nutrients, mediated by e.g. S100A8/A9, provide a simple and elegant mechanism to inhibit microbial growth. The involvement of cell communication, crucial for both the onset and boost of the innate immune response, is illustrated by expression of components of the IL-1 signaling route. The IL-1β signaling cascade has been previously described to initiate and mediate the inflammatory response in keratinocytes [3]. It is hypothesized that in a later stage of the response of PBEC upon exposure to IL-1β/TNFα or *P. aeruginosa*, expression of more specialized antimicrobial agents such as β-defensins occurs. In keratinocytes, it has been previously described that expression of β-defensin-2 occurs at a relatively late stage peaking at 24 hours after IL-1β exposure [3].

In summary, using SAGE we identified a number of genes and pathways that have not previously been implicated in lung inflammation. Gene Ontology analysis indicated the involvement of a number of biological processes such as metabolism, cell growth, development, and cell communication is crucial in the early response of epithelial cells to microbial exposure. Collectively, the differentially expressed genes may play functional roles in the strengthening of the epithelial barrier, initiation of epithelial repair and in the induction and regulation of the inflammatory response. However, extrapolation of the results presented in this study to the *in vivo* situation remains challenging. Verification of our findings in *in vivo* studies will not only further our understanding of the molecular mechanisms underlying the epithelial innate immune response, but may also provide new treatment targets in inflammatory lung diseases.

ACKNOWLEDGMENTS

We thank André Wijfjes and Katerina Pispilli (Leiden Genome Technology Center, LUMC, Leiden, The Netherlands) for providing sequencing facilities and technical support. We also thank Sylvia Lazeroms and Renate Verhoosel (Department of Pulmonology, LUMC, Leiden, The Netherlands) for their assistance in the generation of SAGE libraries and qPCR validation. This research was supported by Netherlands Organization for Scientific Research program grants: 902-11-092, 903-42-085 and 903-68-320.

REFERENCES

1. 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.
2. Bals R. & Hiemstra P.S. (2004) Innate immunity in the lung: how epithelial cells fight against respiratory pathogens. *Eur.Respir.J.* **23**, 327-333.
3. Liu L., Roberts A.A., & Ganz T. (2003) By IL-1 signaling, monocyte-derived cells dramatically enhance the epidermal antimicrobial response to lipopolysaccharide. *J.Immunol.* **170**, 575-580.
4. Tsutsumi-Ishii Y. & Nagaoka I. (2003) Modulation of human beta-defensin-2 transcription in pulmonary epithelial cells by lipopolysaccharide-stimulated mononuclear phagocytes via pro-inflammatory cytokine production. *J.Immunol.* **170**, 4226-4236.
5. Hertz C.J., Wu Q., Porter E.M., Zhang Y.J., Weismuller K.H., Godowski P.J., Ganz T., Randell S.H., & Modlin R.L. (2003) Activation of toll-like receptor 2 on human tracheobronchial epithelial cells induces the antimicrobial peptide human beta defensin-2. *J.Immunol.* **171**, 6820-6826.
6. Takeda K., Kaisho T., & Akira S. (2003) Toll-like receptors. *Annu.Rev.Immunol.* **21**, 335-376.
7. Bals R., Wang X., Wu Z., Freeman T., Bafna V., Zasloff M., & Wilson J.M. (1998) Human beta-defensin 2 is a salt-sensitive peptide antibiotic expressed in human lung. *J.Clin.Invest* **102**, 874-880.
8. Ganz T. (2004) Antimicrobial polypeptides. *J.Leukoc.Biol.* **75**, 34-38.
9. Velculescu V.E., Zhang L., Vogelstein B., & Kinzler K.W. (1995) Serial analysis of gene expression. *Science* **270**, 484-487.
10. 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.
11. Jansen B.J., van Ruissen F., de Jongh G., Zeeuwen P.L., & Schalkwijk J. (2001) Serial analysis of gene expression in differentiated cultures of human epidermal keratinocytes. *J.Invest Dermatol.* **116**, 12-22.
12. Nacht M., Dracheva T., Gao Y., Fujii T., Chen Y., Player A., Akmaev V., Cook B., Dufault M., Zhang M., Zhang W., Guo M., Curran J., Han S., Sidransky D., Buetow K., Madden S.L., & Jen J. (2001) Molecular characteristics of non-small cell lung cancer. *Proc.Natl.Acad.Sci.U.S.A* **98**, 15203-15208.
13. Nacht M., Ferguson A.T., Zhang W., Petroziello J.M., Cook B.P., Gao Y.H., Maguire S., Riley D., Coppola G., Landes G.M., Madden S.L., & Sukumar S. (1999) Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. *Cancer Res.* **59**, 5464-5470.
14. 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.
15. van Wetering S., van der Linden A.C., van Sterkenburg M.A., de Boer W.I., Kuijpers A.L., Schalkwijk J., & Hiemstra P.S. (2000) Regulation of SLPI and elafin release from bronchial epithelial cells by neutrophil defensins. *Am.J.Physiol Lung Cell Mol.Physiol* **278**, L51-L58.
16. Datson N.A., van der P.J., de Kloet E.R., & Vreugdenhil E. (2001) Expression profile of 30,000 genes in rat hippocampus using SAGE. *Hippocampus* **11**, 430-444.
17. Ingman M., Kaessmann H., Paabo S., & Gyllensten U. (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708-713.

18. 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.

19. 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.

20. Zeeberg B.R., Feng W., Wang G., Wang M.D., Fojo A.T., Sunshine M., Narasimhan S., Kane D.W., Reinhold W.C., Lababidi S., Bussey K.J., Riss J., Barrett J.C., & Weinstein J.N. (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.* **4**, R28.

21. Simpson D.A., Feeney S., Boyle C., & Stitt A.W. (2000) Retinal VEGF mRNA measured by SYBR green I fluorescence: A versatile approach to quantitative PCR. *Mol.Vis.* **6:178-83.**, 178-183.

22. Edgar R., Domrachev M., & Lash A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207-210.

23. 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.

24. Donato R. (2003) Intracellular and extracellular roles of S100 proteins. *Microsc.Res.Tech.* **60**, 540-551.

25. 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.

26. Geiss G.K., Bumgarner R.E., An M.C., Agy M.B., 't Wout A.B., Hammersmark E., Carter V.S., Upchurch D., Mullins J.I., & Katze M.G. (2000) Large-scale monitoring of host cell gene expression during HIV-1 infection using cDNA microarrays. *Virology* **266**, 8-16.

27. 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.

28. Zeeuwen P.L., Vlijmen-Willems I.M., Jansen B.J., Sotiropoulou G., Curfs J.H., Meis J.F., Janssen J.J., van Ruissen F., & Schalkwijk J. (2001) Cystatin M/E expression is restricted to differentiated epidermal keratinocytes and sweat glands: a new skin-specific proteinase inhibitor that is a target for cross-linking by transglutaminase. *J.Invest Dermatol.* **116**, 693-701.

29. Zeeuwen P.L., Vlijmen-Willems I.M., Olthuis D., Johansen H.T., Hitomi K., Hara-Nishimura I., Powers J.C., James K.E., Op Den Camp H.J., Lemmens R., & Schalkwijk J. (2004) Evidence that unrestricted legumain activity is involved in disturbed epidermal cornification in cystatin M/E deficient mice. *Hum.Mol.Genet.*

30. Kirfel J., Magin T.M., & Reichelt J. (2003) Keratins: a structural scaffold with emerging functions. *Cell Mol.Life Sci.* **60**, 56-71.

31. Moll R., Franke W.W., Schiller D.L., Geiger B., & Krepler R. (1982) The Catalog of Human Cytokeratins - Patterns of Expression in Normal Epithelia, Tumors and Cultured-Cells. *Cell* **31**, 11-24.

32. Ishida-Yamamoto A., McGrath J.A., Chapman S.J., Leigh I.M., Lane E.B., & Eady R.A. (1991) Epidermolysis bullosa simplex (Dowling-Meara type) is a genetic disease characterized by an abnormal keratin-filament network involving keratins K5 and K14. *J.Invest Dermatol.* **97**, 959-968.

33. McGowan K. & Coulombe P.A. (1998) The wound repair-associated keratins 6, 16, and 17. Insights into the role of intermediate filaments in specifying keratinocyte cytoarchitecture. *Subcell.Biochem.* **31**, 173-204.

34. Pfundt R., van Ruissen F., Vlijmen-Willems I.M., Alkemade H.A., Zeeuwen P.L., Jap P.H., Dijkman H., Fransen J., Croes H., van Erp P.E., & Schalkwijk J. (1996) Constitutive and inducible expression of SKALP/elafin provides anti-elastase defense in human epithelia. *J.Clin.Invest* **98**, 1389-1399.

35. Sallenave J.M. (2002) Antimicrobial activity of antiproteinases. *Biochem.Soc.Trans.* **30**, 111-115.

36. Simpson A.J., Wallace W.A.H., Marsden M.E., Govan J.R.W., Porteous D.J., Haslett C., & Sallenave J.M. (2001) Adenoviral augmentation of elafin protects the lung against acute injury mediated by activated neutrophils and bacterial infection. *Journal of Immunology* **167**, 1778-1786.

37. Wiedow O., Harder J., Bartels J., Streit V., & Christophers E. (1998) Antileukoprotease in human skin: An antibiotic peptide constitutively produced by keratinocytes. *Biochemical and Biophysical Research Communications* **248**, 904-909.

38. Ashcroft G.S., Lei K., Jin W., Longenecker G., Kulkarni A.B., Greenwell-Wild T., Hale-Donze H., McGrady G., Song X.Y., & Wahl S.M. (2000) Secretory leukocyte protease inhibitor mediates non-redundant functions necessary for normal wound healing. *Nat.Med.* **6**, 1147-1153.

39. Arend W.P. (2002) The balance between IL-1 and IL-1Ra in disease. *Cytokine Growth Factor Rev.* **13**, 323-340.

40. Dinarello C.A. (1998) Interleukin-1, interleukin-1 receptors and interleukin-1 receptor antagonist. *Int.Rev.Immunol.* **16**, 457-499.

41. Debets R., Timans J.C., Homey B., Zurawski S., Sana T.R., Lo S., Wagner J., Edwards G., Clifford T., Menon S., Bazan J.F., & Kastelein R.A. (2001) Two novel IL-1 family members, IL-1 delta and IL-1 epsilon, function as an antagonist and agonist of NF-kappa B activation through the orphan IL-1 receptor-related protein 2. *J.Immunol.* **167**, 1440-1446.

42. Towne J.E., Garka K.E., Renshaw B.R., Virca G.D., & Sims J.E. (2004) Interleukin (IL)-1F6, IL-1F8, and IL-1F9 signal through IL-1Rrp2 and IL-1RAcP to activate the pathway leading to NF-kappaB and MAPKs. *J Biol.Chem.* **279**, 13677-13688.

43. 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.

44. Clohessy P.A. & Golden B.E. (1995) Calprotectin-mediated zinc chelation as a biostatic mechanism in host defence. *Scand.J.Immunol.* **42**, 551-556.

45. Loomans H.J., Hahn B.L., Li Q.Q., Phadnis S.H., & Sohnle P.G. (1998) Histidine-based zinc-binding sequences and the antimicrobial activity of calprotectin. *J.Infect.Dis.* **177**, 812-814.

46. Stockley R.A., Dale I., Hill S.L., & Fagerhol M.K. (1984) Relationship of neutrophil cytoplasmic protein (L1) to acute and chronic lung disease. *Scand.J.Clin.Lab Invest* **44**, 629-634.

47. Tibble J., Teahon K., Thjodleifsson B., Roseth A., Sigthorsson G., Bridger S., Foster R., Sherwood R., Fagerhol M., & Bjarnason I. (2000) A simple method for assessing intestinal inflammation in Crohn's disease. *Gut* **47**, 506-513.

48. Feng G., Xu X., Youssef E.M., & Lotan R. (2001) Diminished expression of S100A2, a putative tumor suppressor, at early stage of human lung carcinogenesis. *Cancer Res.* **61**, 7999-8004.

49. Nagy N., Brenner C., Markadieu N., Chaboteaux C., Camby I., Schafer B.W., Pochet R., Heizmann C.W., Salmon I., Kiss R., & Decaestecker C. (2001) S100A2, a putative tumor suppressor gene, regulates in vitro squamous cell carcinoma migration. *Lab Invest* **81**, 599-612.

50. Liu D., Rudland P.S., Sibson D.R., Platt-Higgins A., & Barraclough R. (2000) Expression of calcium-binding protein S100A2 in breast lesions. *Br.J.Cancer* **83**, 1473-1479.

51. Maelandsmo G.M., Florenes V.A., Mellingsaeter T., Hovig E., Kerbel R.S., & Fodstad O. (1997) Differential expression patterns of S100A2, S100A4 and S100A6 during progression of human malignant melanoma. *Int.J.Cancer* **74**, 464-469.

52. Shrestha P., Muramatsu Y., Kudeken W., Mori M., Takai Y., Ilg E.C., Schafer B.W., & Heizmann C.W. (1998) Localization of Ca(2+)-binding S100 proteins in epithelial tumours of the skin. *Virchows Arch.* **432**, 53-59.

53. Hough C.D., Cho K.R., Zonderman A.B., Schwartz D.R., & Morin P.J. (2001) Coordinately up-regulated genes in ovarian cancer. *Cancer Res.* **61**, 3869-3876.

54. Zhang T., Woods T.L., & Elder J.T. (2002) Differential responses of S100A2 to oxidative stress and increased intracellular calcium in normal, immortalized, and malignant human keratinocytes. *J.Invest Dermatol.* **119**, 1196-1201.

55. Breen E.C. & Tang K. (2003) Calcyclin (S100A6) regulates pulmonary fibroblast proliferation, morphology, and cytoskeletal organization in vitro. *J.Cell Biochem.* **88**, 848-854.

56. Bronckart Y., Decaestecker C., Nagy N., Harper L., Schafer B.W., Salmon I., Pochet R., Kiss R., & Heizman C.W. (2001) Development and progression of malignancy in human colon tissues are correlated with expression of specific Ca(2+)-binding S100 proteins. *Histol.Histopathol.* **16**, 707-712.

57. Hiemstra P.S., van den Barselaar M.T., Roest M., Nibbering P.H., & van Furth R. (1999) Ubiquicidin, a novel murine microbicidal protein present in the cytosolic fraction of macrophages. *Journal of Leukocyte Biology* **66**, 423-428.

58. Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A., & Speleman F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, RESEARCH0034.

59. 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.

60. 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.

61. Chen J., Lee S., Zhou G., & Wang S.M. (2002) High-throughput GLGI procedure for converting a large number of serial analysis of gene expression tag sequences into 3' complementary DNAs. *Genes Chromosomes.Cancer* **33**, 252-261.

62. Chen J.J., Rowley J.D., & Wang S.M. (2000) Generation of longer cDNA fragments from serial analysis of gene expression tags for gene identification. *Proc.Natl.Acad.Sci.U.S.A* **97**, 349-353.