BMC Genomics

Research article

Molecular signature of clinical severity in recovering patients with severe acute respiratory syndrome coronavirus (SARS-CoV)

Yun-Shien Lee[1,2], Chun-Houh Chen[3], Angel Chao[4,5], En-Shih Chen[1], Min-Li Wei[1], Lung-Kun Chen[1], Kuender D Yang[6], Meng-Chih Lin[7], Yi-Hsi Wang[7], Jien-Wei Liu[8], Hock-Liew Eng[9], Ping-Cherng Chiang[10], Ting-Shu Wu[10], Kuo-Chein Tsao[11], Chung-Guei Huang[11], Yin-Jing Tien[12], Tzu-Hao Wang*[1,4], Hsing-Shih Wang[4,5] and Ying-Shiung Lee[5]

Address: [1]Genomic Medicine Research Core Laboratory, Chang Gung Memorial Hospital (CGMH), Tao-Yuan, Taiwan, [2]Department of Biotechnology, Ming Chuan University, Tao-Yuan, Taiwan, [3]Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, [4]Department of Obstetrics and Gynecology, Lin-Kou Medical Center, CGMH, Tao-Yuan, Taiwan, [5]Graduate Institute of Clinical Medical Sciences, College of Medicine, Chang Gung University, Tao-Yuan, Taiwan, [6]Department of Pediatrics, Kaohsiung Medical Center, CGMH, Kaohsiung, Taiwan, [7]Department of Internal Medicine, Division of Pulmonary and Critical Care Medicine, Kaohsiung Medical Center, CGMH, Kaohsiung, Taiwan, [8]Department of Internal Medicine, Division of Infectious Diseases, Kaohsiung Medical Center, CGMH, Kaohsiung, Taiwan, [9]Department of Pathology, Kaohsiung Medical Center, CGMH, Kaohsiung, Taiwan, [10]Department of Internal Medicine, Division of Infectious Diseases, Lin-Kou Medical Center, CGMH, Tao-Yuan, Taiwan, [11]Clinical Virology Laboratory, Department of Clinical Pathology, CGMH, Tao-Yuan, Taiwan and [12]Institute of Statistics, National Central University, Tao-Yuan, Taiwan

Email: Yun-Shien Lee - bojack@mcu.edu.tw; Chun-Houh Chen - cchen@stat.sinica.edu.tw; Angel Chao - angel945@adm.cgmh.org.tw; En-Shih Chen - chenenshih@adm.cgmh.org.tw; Min-Li Wei - ricewei@pchome.com.tw; Lung-Kun Chen - kervin@adm.cgmh.org.tw; Kuender D Yang - yangkd@adm.cgmh.org.tw; Meng-Chih Lin - mengchih@adm.cgmh.org.tw; Yi-Hsi Wang - yhwang@cgmh.org.tw; Jien-Wei Liu - jwliu@adm.cgmh.org.tw; Hock-Liew Eng - eng6166@adm.cgmh.org.tw; Ping-Cherng Chiang - pcchiang@adm.cgmh.org.tw; Ting-Shu Wu - tswu@adm.cgmh.org.tw; Kuo-Chein Tsao - kctsao@adm.cgmh.org.tw; Chung-Guei Huang - cghung@adm.cgmh.org.tw; Yin-Jing Tien - gary@stat.sinica.edu.tw; Tzu-Hao Wang* - knoxtn@cgmh.org.tw; Hsing-Shih Wang - hswang@cgmh.org.tw; Ying-Shiung Lee - ysln@adm.cgmb.org.tw

* Corresponding author

Published: 21 September 2005

BMC Genomics 2005, **6**:132 doi:10.1186/1471-2164-6-132

Received: 11 March 2005
Accepted: 21 September 2005

This article is available from: http://www.biomedcentral.com/1471-2164/6/132

Abstract

Background: Severe acute respiratory syndrome (SARS), a recent epidemic human disease, is caused by a novel coronavirus (SARS-CoV). First reported in Asia, SARS quickly spread worldwide through international travelling. As of July 2003, the World Health Organization reported a total of 8,437 people afflicted with SARS with a 9.6% mortality rate. Although immunopathological damages may account for the severity of respiratory distress, little is known about how the genome-wide gene expression of the host changes under the attack of SARS-CoV.

Results: Based on changes in gene expression of peripheral blood, we identified 52 signature genes that accurately discriminated acute SARS patients from non-SARS controls. While a general suppression of gene expression predominated in SARS-infected blood, several genes including those involved in innate immunity, such as defensins and eosinophil-derived neurotoxin, were upregulated. Instead of employing clustering methods, we ranked the severity of recovering SARS patients by generalized associate plots (GAP) according to the expression profiles of 52 signature genes. Through this method, we discovered a smooth transition pattern of severity from normal

controls to acute SARS patients. The rank of SARS severity was significantly correlated with the recovery period (in days) and with the clinical pulmonary infection score.

Conclusion: The use of the GAP approach has proved useful in analyzing the complexity and continuity of biological systems. The severity rank derived from the global expression profile of significantly regulated genes in patients may be useful for further elucidating the pathophysiology of their disease.

Background

SARS-CoV is a single-stranded, plus-sense RNA virus with a genome of ~30 kb. Its sequence does not closely resemble any of the previously characterized coronaviruses [1-4]. Before SARS-CoV was recognized as the cause of the deadly SARS [1-3,5-7], other human coronaviruses had only been known to account for 15–30% of colds [8]. SARS-CoV appears to be new to humans, as supported by the finding that human sera collected before the SARS outbreak did not contain antibodies against this virus [3,9]. After an incubation period from 2 to 10 days, SARS patients might develop fever (>38°C), headache, dry cough, and pneumonia [3,5,9-14]. Most patients gradually recovered while some progressed to respiratory distress syndrome with ~10% mortality rate. The genome-wide changes in human gene expression when challenged by this novel pathogen are essentially unknown.

Profiles of gene expression patterns help define the complex biological processes associated with both health and disease *in vivo*. Investigation of host responses to infection with *in vitro* models have offered insights into mechanisms of pathogenesis, and have highlighted the potential for applications of microarray technology to diagnose infection *in vivo* [15]. Whitney et al. observed that the variation in gene expression patterns in the blood of healthy subjects was strikingly smaller than the significant changes induced by diseases either in patients with cancer or with bacterial infections [15]. It was conceivable that microarray profiling of gene expression in whole bloods exhibits the potential in monitoring the patients' responses to a disease, especially a novel infection such as SARS.

Many discriminative methods have been developed for analysis of microarray gene expression data in cancer patients and the resulting classifications have been correlated closely with clinical parameters [16-19]. For instance, the discovery of signature genes for breast cancers through microarray analysis of gene expression has provided us with a more precise clinical staging that will improve the outcome of treatment [20,21]. However, clinical parameters are not always in a discrete pattern but more likely in a continuous fashion, where an absolute classification may not be achievable. Herein we present the use of cDNA microarray analysis of gene expression in whole blood from a cohort of recovering SARS patients, of whom the disease severity appeared to be a continuum. After we had identified the molecular signature of 52 genes that accurately discriminated acute SARS patients from non-SARS controls, we ranked the disease severity of these patients using a generalized association plot (GAP) elliptical seriation algorithm [22] based on the expression profiles of the 52 genes. The derived severity rank of the patients proved to be closely correlated with their clinical parameters, namely, the recovery period (in days) and the clinical pulmonary infection score.

Results

Patient information

Using the cDNA microarrays spotted with duplicated 7,334 cDNA clone, we analyzed RNA specimens successfully amplified in 44 peripheral blood collected from 25 confirmed SARS patients (age ranged from 23 to 80 years old, mean = 41.8, SD = 17.2, median = 34), of whom 24 survived. Except for one patient who died on the 4th day, duration of hospitalization in this cohort ranged from 12 to 51 days (n = 24, mean = 24.5, SD = 10.1, median = 21) (Additional file 1). We defined 11 specimens as acute SARS (AS) using the following criteria: (i) the whole blood RNA from a hospitalized patient was PCR positive for SARS-CoV, or (ii) the specimen was collected within 10 days after the disease onset in patients whose blood was later diagnosed ELISA-positive for anti-SARS IgG. The rest of 33 RNA specimens from SARS patients were labelled as recovering SARS (RS). Our study included 11 normal control (NC) volunteers and 11 patients with bacterial infections (IN) as healthy and non-SARS infection controls, respectively (Additional files 2 and 3).

cDNA microarray analysis

When we compared the gene expression profiles among acute SARS (AS), recovering SARS (RS), bacterial infection (IN), and normal control (NC) groups, we observed the variances of gene expression in both SARS (AS, AS+RS) and bacterial (IN) groups to be equally higher than that in healthy controls (NC) (Fig. 1a). This result indicates that gene expression profiles of either SARS or bacterial groups differed significantly from that of normal controls.

A probe set of 885 genes with standard deviations greater than 0.5 across 66 arrays was selected for further analyses.

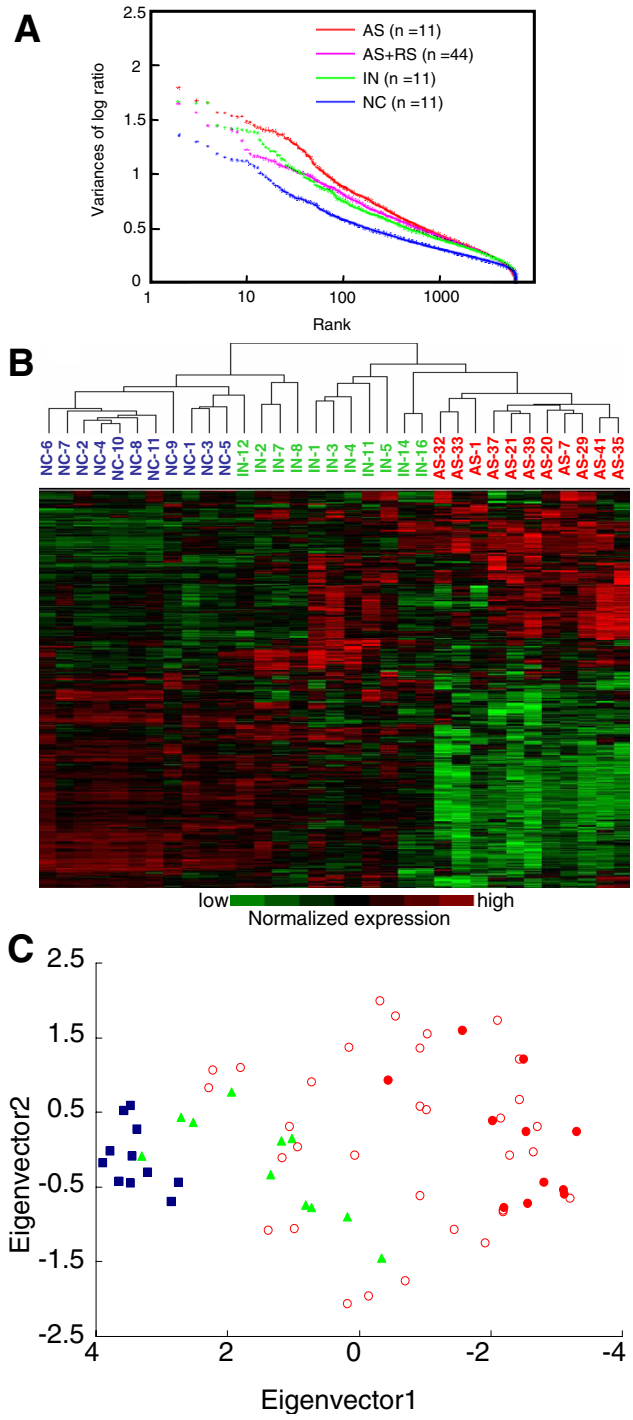

Figure 1

Significant differences in gene expression profiles in patients with SARS or bacterial infection. Using a probe set with 6,525 annotated genes, global gene expression was analyzed by **(A)** variation distribution in peripheral blood specimens from patients with acute SARS (AS), recovering SARS (RS), bacterial infections (IN), and normal controls (NC). **(B)** In the hierarchical clustering of relative change in gene expression using a probe set of 885 filtered genes (gene vector >0.5 SD), red indicates upregulation and green indicates downregulation in gene expression relative to a common reference that was the pooled amplified RNA from 11 normal controls. **(C)** Using the 885-gene set, singular value decomposition (SVD) analysis by two eigenvectors showed three distinguished clusters of AS (red ●), IN (green ▲), and NC (blue ■) groups, with the RS (red ❍) specimens scattering among AS and IN.

An average linkage hierarchical clustering tree with Pearson correlation proximity was built on the 33 arrays (11 NC, 11 IN, and 11AS) using these 885 genes (Fig. 1b). The AS and NC groups were well separated into two opposite coherent clusters. Singular value decomposition (SVD) analysis, a dimension reduction method to project gene expression profiles to fewer representative eigenvectors [23], also successfully separated AS, IN, and NC specimens into three clusters with first two eigenvectors (Fig. 1c). Interestingly, the recovering SARS (RS) samples are interspersed among the AS and IN samples.

To identify which genes were specifically regulated by SARS-CoV, we performed two sets of two-sample Student t-test for means with an unequal-variance assumption. In the first set, we contrasted 11 AS versus 22 non-AS (NC and IN) specimens on all 885 genes. The genes with significant testing results were considered to be specifically induced by SARS-CoV (Fig. 2a,b). For the second set of t-tests, we compared 11 NC with 22 non-NC (IN and AS) specimens. We considered that the change in significant genes identified by the second t-test was induced by both bacterial and viral infections (Fig. 2c,d). Genes identified from these two sets of test were then ranked separately according to the corresponding sets of P-values. Gene expression profiles for the top 20 and the bottom 20 genes from both sets are displayed as Figure 2.

Unexpectedly, most of the genes specifically upreguated by SARS-CoV are ESTs (13/20 genes) that were not annotated previously (Fig. 2a). On the other hand, SARS-CoV stimulated the host innate immunity by upregulating genes including defensins [24,25] and eosinophil derived neurotoxin [26,27], similar to that of bacterial infections (Fig. 2c).

Signature genes and GAP algorithms

A simple k-nearest-neighbour method was used to obtain a near optimal number of 30 genes from the 885 filtered gene set for discriminating specimens between acute SARS (AS) and non-SARS (NC and IN) (Additional file 4). The selected top 30 upregulated ($P < 6 \times 10^{-6}$) and the top 30 downregulated genes ($P < 4 \times 10^{-7}$) from the AS versus non-AS (IN and NC) Student's t-test were used as the specific probe set to assess the status of SARS infection. Eight genes that were also significant in the NC versus non-NC (AS and IN) t-test were excluded, resulting in a specific AS probe set of 52 genes. For the GAP analysis, we calculated pair-wise Euclidean distances among 55 samples (11 AS, 33 RS, 11 NC) using these 52 genes, aiming to identify a one-dimensional order that could reflect the severity structure of the disease (Fig. 3a). Using this GAP elliptical arrangement of 55 specimens (columns), we observed a transition of gene expression patterns of 52 genes (rows) from the left side where NC clustered to the right side where AS accumulated (Fig. 3b). Hierarchical clustering trees guided by self-organized map (SOM) and other clustering methods were also performed to sort the SARS patient samples using the same 52 genes. The Robinson criterion [22,28,29] is often employed to assess the performances of different seriation algorithms. Table 1 (and Additional file 5) shows that the GAP algorithm derived a smoother transition pattern than other methods in the Robinson sense. Thereby, we have derived the SARS severity rank according to the expression profile of 52 signature genes as a whole in each patient, as demonstrated by the smooth transition of expression levels in each (row) of these genes from NC to RS to AS (Fig. 3b).

For validation purposes, we further tested the stability of the rank (order) derived from GAP analysis on the 52 genes for the 55 specimens. The same GAP procedure was repeatedly applied to the top 20 to 200 genes (among the filtered 885 genes) with significant p-values (Student's t-test) between the AS versus non-AS (IN and NC). While the ranks for the 55 specimens obtained from the most significant 20 to 200 genes are highly correlated to each other, they are significantly different from the ranks derived from the 52-gene sets that were randomly selected from the 885 genes (data not shown).

We scrutinized the clinical courses of patients who donated the 10 RS specimens that were scattered among AS (Fig. 3a) and found evidence of underlying severity of the disease in the majority of patients. For example, sample RS43 from a patient who had been discharged from hospital for 2 weeks was still PCR-positive for SARS-CoV; RS54, a PCR-positive sample was not grouped as AS because of the negative ELISA result. RS38, RS40, and RS42 still represented acute SARS infections because they were collected only 1, 2, and 3 days after AS37, AS39, and AS41, respectively. Patients with RS78 and RS91 who had severe SARS courses were hospitalized for 41 and 51 days, respectively. The patient for RS8 was in the second week of disease. The only two unexplained specimens, RS18 and RS71 from the same patient, may represent a unique biological variability, accounting for the misclassification using this 52-gene molecular signature.

Molecular signature for severity and clinical correlations

To test the efficacy of using these 52 genes as the molecular signature for the severity of SARS patients, we identified a significant correlation ($P < 1 \times 10^{-6}$) between the derived rank of SARS severity and the number of days after the onset of disease (Fig. 4a). We further used this rank of SARS severity to examine the recovery trend in 17 recovering patients who had donated multiple specimens (Fig. 4b). Except for the one patient (5.3 % = 1/19, shown as the red line in Fig. 4b), similar trends existed in 18 out of 19 lines (94.7 %). Pugin et al. combined body

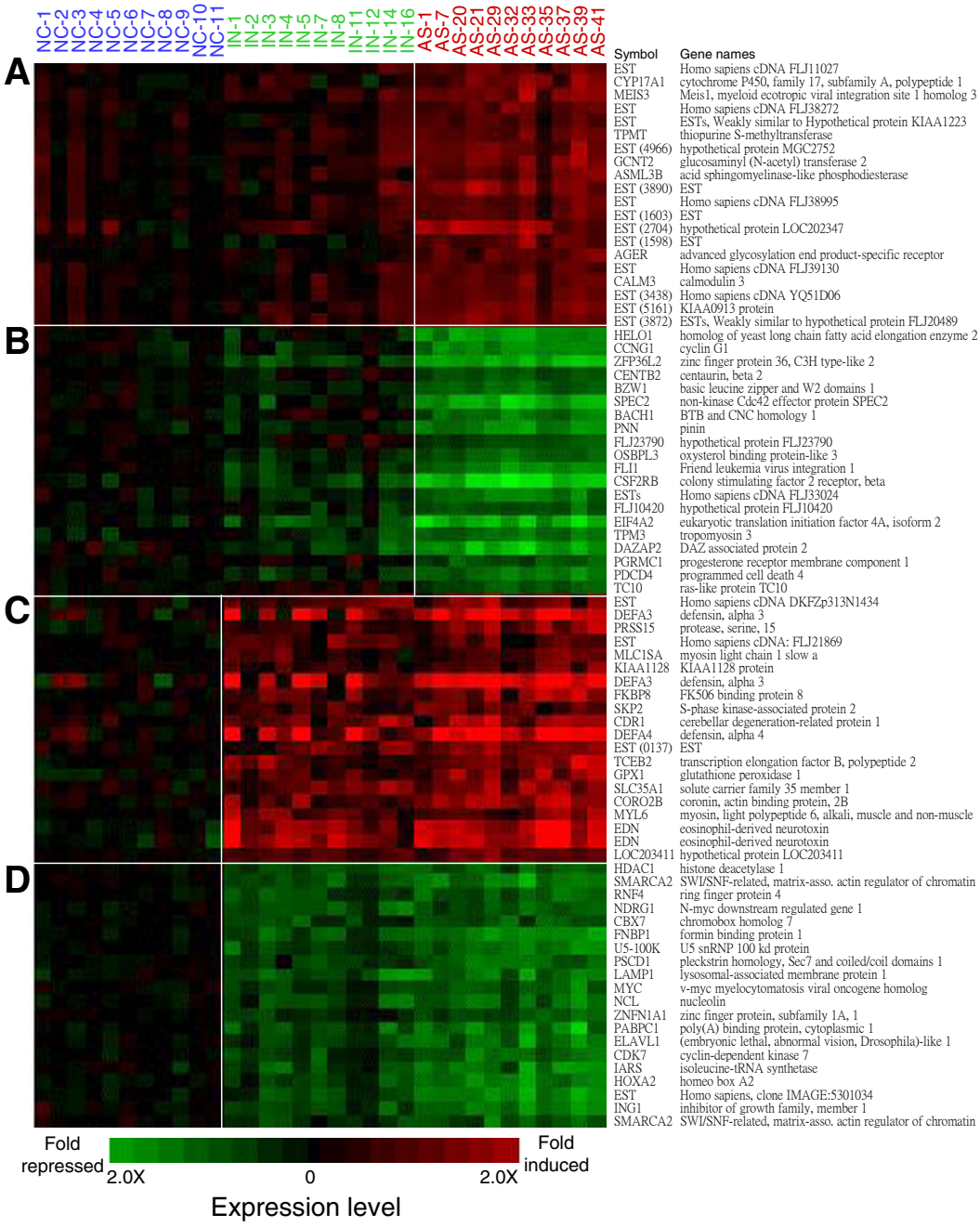

Figure 2

The top 40 discriminating genes with the highest distinction values for AS or NC groups. Twenty genes that were specifically **(A)** upregulated or **(B)** downregulated in patients with SARS. Another twenty genes that were non-specifically **(C)** upregulated or **(D)** downregulated by both bacterial infection and SARS. Each column represents an individual sample and each row represents a gene. The color range reflected relative change according to the scale shown. NC, normal control; IN, bacterial infection; AS, acute SARS. GMRCL clone numbers of some ESTs are also included in the parentheses.

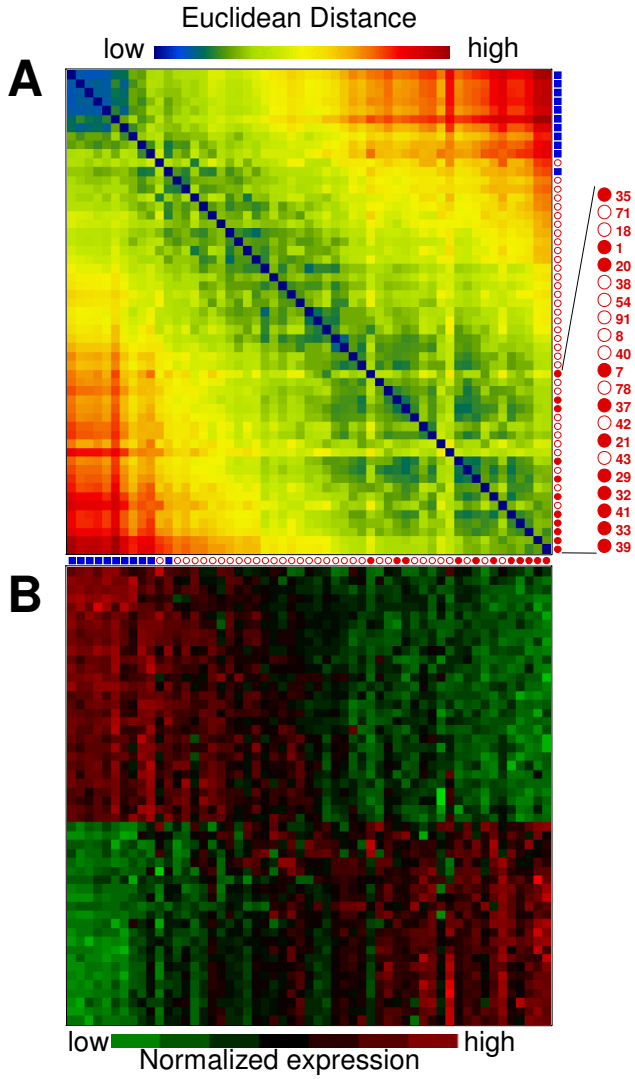

Figure 3
Generalized associated plots (GAP) analysis of SARS patients samples. **(A)** Pair-wise Euclidean distance matrix that was sorted by a GAP using 52 genes with the highest discriminating power for AS groups revealed the minimum anti-Robinson events in the matrix, resulting in a smooth transition order of the AS and RS specimens from severely diseased to healthy states. AS (red ●); RS (red ❍); NC (blue ▮). **(B)** Gene expression profile for the 52 discriminating genes displayed in the order obtained from the GAP method.

temperature, white blood cell count, volume and appearance of tracheal secretions, oxygenation, chest X-ray, and tracheal aspirate cultures into a clinical pulmonary infection score (CPIS) as a diagnostic tool for pneumonia [30]. We observed that the rank of SARS severity was also significantly ($P < 0.001$) correlated with the CPIS (Fig. 4c). Collectively, these results demonstrate a correlation between the molecular severity rank and clinical factors, suggesting the usefulness of the molecular signature as a genome-wide parameter for gauging the severity of SARS patients.

Discussion

Diverse infections can induce a shared core gene expression involving the human innate immune system; each infection may also trigger a pathogen-specific immune response of the host. The innate immune genes were upregulated in both acute SARS (AS) and bacteria infection (IN) patients (Fig. 2c). SARS was a novel viral infection that had not been encountered by the humans in the history before 2003. Intriguingly, most of the genes specifically upreguated in SARS patients were ESTs (13/20 genes) (Fig. 2a), suggesting that the first human encounter with SARS-CoV might provoke a set of human genes that were poorly annotated due to disuse. Annotation of these ESTs may lead to the discovery of novel genes.

Given the high cost of microarray analyses, the detection of a comprehensive gene expression profile may not be cost-effective for clinical diagnosis and evaluation of patients with infectious diseases. However, in a complex system such as the human body where genes interplay through intricate circuitries, it is inadequate to examine only a few routine parameters in biochemistry and blood cell counts for the global physiochemical status of a patient at the time of blood collection. In this report, we applied the GAP method to derive a smooth transition pattern among samples based on the molecular signature consisting of 52 genes, which in turn were used to monitor the severity of clinical courses of SARS patients. Instead of clustering samples into discrete groups in a method similar to commonly-used microarray classifications [31], GAP focuses more on a global orientation of the sample-to-sample relationship. For instance, the AS and RS samples were seriation ranked (Fig. 3), and the rank order proved to correlate well with clinical parameters (Fig. 4).

The GAP-derived rank of severity also provided us with a unique way, where expression of most relevant genes were all considered, to decipher the meaning of the changes in other genes obtained from the same microarray experiment. For instance, we have identified the correlative change in matrix metalloproteinase MMP-7 and MMP-9 (Additional file 6): both can stimulate α-defensin [32]. Importantly, these correlations could not be revealed with other parameters alone, such as number of days after disease onset or clinical score CPIS (data not shown).

In this study, however, there might be technical limitations during RNA isolation from some clinical specimens as well as an unavoidable sample-collecting bias. First, both RNA isolation from SARS specimens and RNA

Table 1: Performance of Robinson structure with different seriation algorithms.

	Anti-Robinson Events	
Seriation Algorithm	Counts	Scaled Counts (%)
Random Order (NC-AS-RS)	26,245	100.00
Original Order	23,030	87.75
Self Organizing Map (SOM) Order	11,499	43.81
Average Linkage Tree/Original	10,268	39.12
Average Linkage Tree/SOM	10,738	40.91
Average Linkage Tree/GAP	5,940	22.63
GAP Elliptical Order	5,022	19.14

amplification were performed in the Biosafety Level 3 laboratory, where the instrument for RNA quantitation was not available. This limitation resulted in the failed generation of aRNA from 10 out of 54 SARS specimens (Additional file 1). Unfortunately, these 10 specimens contained 7 specimens from patients at an early (i.e. first 2 weeks) stage [14]. Secondly, 25 SARS patients who donated blood specimens for this study may belong to the milder subgroup of a total of 44 SARS patients in Kaohsiung Medical Center of Chang Gung Memorial Hospital. According to a paper describing the complete cohort of SARS patients [33], intubation and mechanical ventilation were required in 20 out of these 44 patients. However, only two in our 25 patients needed intubation (Additional file 1). The aforementioned two potential limitations may account for why our microarray results could not detect a correlation with a possible worsening clinical course before recovering, which was described by Peiris et al [14].

In conclusion, we propose the use of a molecular signature reflecting the severity of SARS in order to interpret the trends of expression changes in groups of genes within particular functional categories. The use of GAP methodology proved to be instrumental in determining the severity of SARS. The derived severity ranking of SARS patients in turn formed a gradual basis for the analysis of the interaction patterns, providing us with a useful tool for understanding the molecular pathogenesis of this novel viral infection.

Conclusion

We illuminate the human gene expression profiles, in terms of gene expression in peripheral blood, to the unprecedented infection of SARS-CoV. We also discovered a smooth transition pattern of severity from normal controls to acute SARS patients based on the gene expression profiles by generalized associate plots (GAP). The rank of SARS severity was significantly correlated with other clinical parameters.

Methods

Patient information and specimen preparation

Blood specimens of 25 SARS patients (Additional file 1) were collected from 10 May to 4 July 2003 at Kaohsiung Medical Center of Chang Gung Memorial Hospital (CGMH) in Kaohsiung City of southern Taiwan. Two additional blood samples (RS94 and RS97) were collected from apparently healthy individuals who had recovered from SARS infection 3 months later. Diagnosis of SARS was based on the guidelines of World Health Organization (WHO) [34]. More comprehensive data of the SARS cohort were previously published [33]. This study was approved by the Institute Review Board of CGMH. Total RNA was isolated with the PAXgene Blood RNA System (Qiagen, USA) and stored at -80°C. After RNAs were further purified and concentrated into 15 μl BR5 solution with RNeasy MinElute kit (Qiagen, USA), 2 μl were used for linear RNA amplification using RiboAmp RNA Amplification Kit (Arcturus, California USA). Before the first Strand Nuclease Mix was added to the RNA samples, all of the RNA purification and amplification were performed inside a Biosafety Level 3 laboratory located in Lin-Kou Medical Center of CGMH. We analyzed the quality and quantity of amplified RNA with Bioanalyzer 2100 (Agilent, CA, USA).

Anti-SARS-CoV IgG ELISA and real-time quantitative PCR analysis

The antigen used for the SARS detection ELISA was the detergent-extracted and gamma irradiated Vero E6 cells infected with SARS-CoV. Identical preparations from uninfected Vero E6 cells were used as the control. Patients' sera were 1:10 diluted and added to the ELISA plates, and goat anti-human IgG antibody conjugated with horseradish peroxidase (DAKO, Cambridgeshire, UK) was added

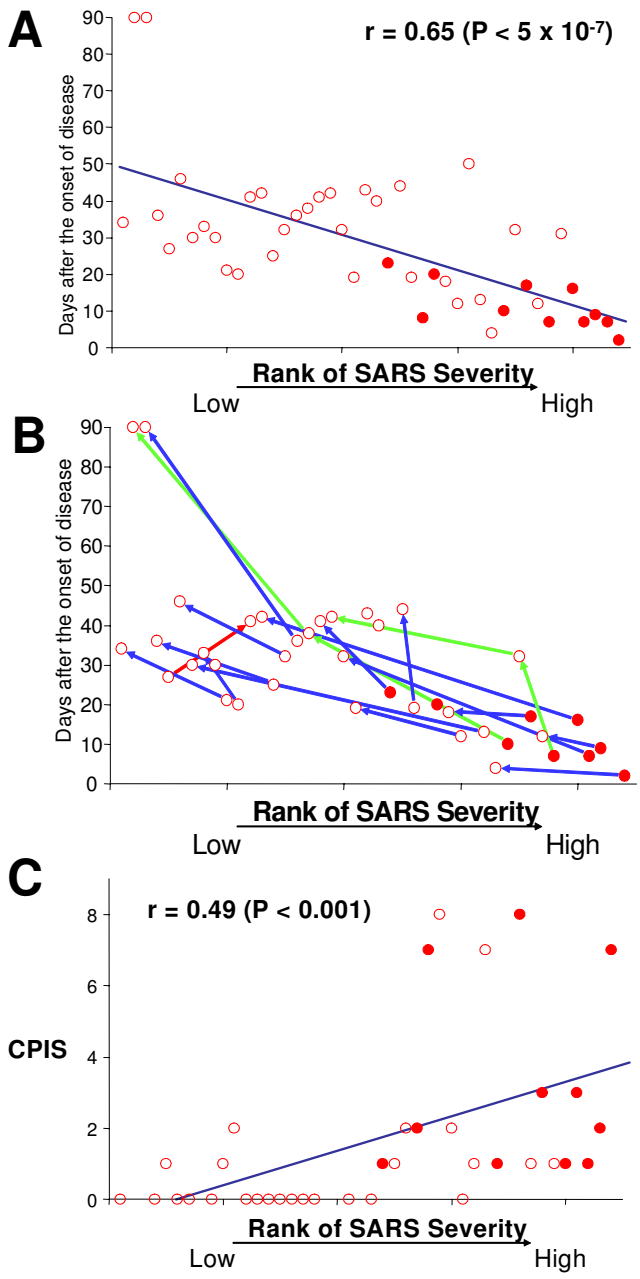

Figure 4
Correlations between the GAP-derived rank for SARS severity and clinical parameters. (A) The scatter plot of all SARS specimens with the order obtained from the GAP method and the days after the onset of disease showed a significant correlation ($P < 5 \times 10^{-7}$). **(B)** Sixteen out of 17 SARS patients who submitted multiple blood specimens showed a similar trend of changes in the GAP-derived severity rank along with the recovery from the disease. Patients with 2 (n = 15) and 3 specimens (n = 2) were labeled with blue and green lines, respectively. **(C)** The scatter plot of all AS and RS specimens with the order obtained from the GAP method and clinical pulmonary infection score (CPIS) showed a significant correlation ($P < 0.001$).

for enzymatic reaction. After adding the substrate, O-phenylenediamine, the optical density (O.D.) was measured at 450 nm wavelength. The cutoff value of O.D. for SARS-CoV IgG ELISA was 0.15. Sensitivity of this method was 100% (28/28 in confirmed SARS cases) and specificity was 98.4% (790/803) in the healthy control group.

Real-time quantitative PCR analysis for SARS-CoV was performed with Cor-p-F4, Cor-p-R4 and Cor-probe developed by CDC (GA, USA) with HT 7900 Sequence Detection System (Applied Biosystems, CA, USA).

Microarray procedures

In this study, we used the *GMRCL Human 7K set, Version 2* chips as previously described [35]. Twelve amplified RNA samples from healthy donors (Additional file 3) were pooled as the common reference for every array in this study. A total of 66 aRNA samples including 11 acute SARS (AS), 33 recovering SARS (RS), 11 non-SARS infection (IN), and 11 normal controls (NC) were analyzed with cDNA microarrays as tests against the pooled aRNA (the common reference). Among 66 aRNA preparations, 28 were analyzed with the dye-swapping microarray design. We averaged the log ratios of the duplicated spots on each slide. In the dye swapping experiments, we further averaged the log ratios derived from two slides. We used 400 ng of aRNA for labeling and hybridization using a 3DNA Array 350RP Detection kit (Genisphere, PA, USA), and scanned slides with a confocal scanner ChipReader (Virtek, Canada). We acquired the spot and background intensities with GenePix Pro 4.1 software (Axon Instruments, Inc., CA, USA), and carried out within-slide normalization using programs written with MATLAB 6.0 software (The MathWorks, Inc., MA, USA). To assure the reproducibility of our microarray system, we got the similar gene expression profiles from replicated samples (RS88) using the hierarchical clustering analysis and also got the highly correlated results ($r^2 = 0.84$) from two specimens (AS37 and RS38) that were collected from the same patient at a time interval of only one day. We consistently obtained identical results in each of 28 pairs dye-swapping experiments. The complete microarray data is available in Additional file 7.

Hierarchical clustering and singular value decomposition

We performed hierarchical clustering using Cluster and TreeView software [36] with the following parameters: (i) a standard deviation > 0.5 as the filtering cutoff point (885 genes with marked changes selected among 66 arrays), (ii) mean-centered genes and normalized genes, (iii) cluster analysis carried out with uncentered correlation of arrays. We also performed a singular value decomposition (SVD) [23] analysis of the correlation matrix for all 66 samples. The first two eigenvectors weighted by the

corresponding singular values (eigenvalues) of the 66 samples were plotted against each other.

Euclidean distance matrix by generalized association plots

Robinson criterion [22,28] is frequently used to assess the performances of sorting algorithms on symmetric proximity matrices. A Robinson Matrix, R = $[r_{ij}]$, is a symmetric matrix such that $r_{ij} \le r_{ik}$ if $j<k<i$ and $r_{ij} \ge r_{ik}$ if $i<j<k$. The GAP elliptical seriation [22] utilizing the ellipse structure from a singular value decomposition of a converged correlation coefficient matrix usually identifies permuted matrix with a near Robinson form. A brief review on GAP and some details of its applications are available [37].

Authors' contributions

Lee YS, Chen CH and Wang TH designed the study and prepared the manuscript. Lee YS, Tien-YJ and Chen CH carried out the statistical analysis. Lee YS, Wang TH, Chen ES, Wei ML, Chen LK carried out the microarray experiments. Chao A, Yang KD, Lin MC, Wang YH, Liu JW, Eng HW, Chiang PC, Wu TS, Tsao KC, Huang CG, Wang HS and Lee YS obtained the clinical materials and analyzed clinical information. All authors read and approved the final manuscript.

Additional material

Additional File 1

Demographics of SARS Patients.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-6-132-S1.doc]

Additional File 2

Demographics of patients with non-SARS infection.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-6-132-S2.doc]

Additional File 3

Demographics of healthy donor (information of the pooled reference).
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-6-132-S3.doc]

Additional File 4

K-nearest-neighbour methods in evaluating the best discriminating (classifying) accuracy for AS and non-SARS specimens.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-6-132-S4.doc]

Additional File 5

Euclidean distance matrix of 55 specimens with 52 selected genes using different seriation algorithms.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-6-132-S5.doc]

Additional File 6

Analyses of gene expression in MMP-7 and MMP-9, both of which are involved in innate immunity.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-6-132-S6.doc]

Additional File 7

Complete microarray data
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-6-132-S7.zip]

Acknowledgements

We thank Yalin Huang, Fong-Yee Chiu, Yu-Liey Tong, Wei-Hsiang Kong, Shihyee Mimi Wang (University of Illinois in Chicago), Hsiu-Chuan Liu, Rong-Fu Chen and Ling-Ling Huang for technical assistance, Shih-Tien Wang (Northwestern University) for editing the manuscript, and PC Huang (Johns Hopkins University) for critical comments. The authors also gratefully acknowledge the SARS team of Kaohsiung Chang Gung Memorial Hospital (Yun-Tze Chen, Ju-Hao Lee, Sui-Liong Wang, Tze-Yu Lee, Chao-Chien Wu, Sheung-Fat Ko, Chen-Hsiang Lee) and many more medical personnel who served courageously during the SARS episode. This study was supported by grants CMRPD32019S (YS Lee), CMRPG1008 (TH Wang), CMRPG32010S (TH Wang) from Chang Gung Memorial Hospital, NSC93-2320-B-130-001 (YS Lee) from National Science Council of Taiwan, and a generous donation of Mr. Yung-Ching Wang, Chairman of Formosa Plastic Corporation.

References

1. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Penaranda S, Bankamp B, Maher K, Chen MH, Tong S, Tamin A, Lowe L, Frace M, DeRisi JL, Chen Q, Wang D, Erdman DD, Peret TC, Burns C, Ksiazek TG, Rollin PE, Sanchez A, Liffick S, Holloway B, Limor J, McCaustland K, Olsen-Rasmussen M, Fouchier R, Gunther S, Osterhaus AD, Drosten C, Pallansch MA, Anderson LJ, Bellini WJ: **Characterization of a novel coronavirus associated with severe acute respiratory syndrome.** *Science* 2003, **300:**1394-1399.
2. Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, Khattra J, Asano JK, Barber SA, Chan SY, Cloutier A, Coughlin SM, Freeman D, Girn N, Griffith OL, Leach SR, Mayo M, McDonald H, Montgomery SB, Pandoh PK, Petrescu AS, Robertson AG, Schein JE, Siddiqui A, Smailus DE, Stott JM, Yang GS, Plummer F, Andonov A, Artsob H, Bastien N, Bernard K, Booth TF, Bowness D, Czub M, Drebot M, Fernando L, Flick R, Garbutt M, Gray M, Grolla A, Jones S, Feldmann H, Meyers A, Kabani A, Li Y, Normand S, Stroher U, Tipples GA, Tyler S, Vogrig R, Ward D, Watson B, Brunham RC, Krajden M, Petric M, Skowronski DM, Upton C, Roper RL: **The Genome sequence of the SARS-associated coronavirus.** *Science* 2003, **300:**1399-1404.
3. Peiris JS, Lai ST, Poon LL, Guan Y, Yam LY, Lim W, Nicholls J, Yee WK, Yan WW, Cheung MT, Cheng VC, Chan KH, Tsang DN, Yung RW, Ng TK, Yuen KY: **Coronavirus as a possible cause of severe acute respiratory syndrome.** *Lancet* 2003, **361:**1319-1325.
4. Ruan YJ, Wei CL, Ee AL, Vega VB, Thoreau H, Su ST, Chia JM, Ng P, Chiu KP, Lim L, Zhang T, Peng CK, Lin EO, Lee NM, Yee SL, Ng LF, Chee RE, Stanton LW, Long PM, Liu ET: **Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection.** *Lancet* 2003, **361:**1779-1785.
5. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA, Berger A, Burguiere AM, Cinatl J, Eickmann M, Escriou N, Grywna K, Kramme S,

Manuguerra JC, Muller S, Rickerts V, Sturmer M, Vieth S, Klenk HD, Osterhaus AD, Schmitz H, Doerr HW: **Identification of a novel coronavirus in patients with severe acute respiratory syndrome.** *N Engl J Med* 2003, **348:**1967-1976.
6. Fouchier RA, Kuiken T, Schutten M, Van Amerongen G, Van Doornum GJ, Van Den Hoogen BG, Peiris M, Lim W, Stohr K, Osterhaus AD: **Aetiology: Koch's postulates fulfilled for SARS virus.** *Nature* 2003, **423:**240.
7. Kuiken T, Fouchier RA, Schutten M, Rimmelzwaan GF, van Amerongen G, van Riel D, Laman JD, de Jong T, van Doornum G, Lim W, Ling AE, Chan PK, Tam JS, Zambon MC, Gopal R, Drosten C, van der Werf S, Escriou N, Manuguerra JC, Stohr K, Peiris JS, Osterhaus AD: **Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome.** *Lancet* 2003, **362:**263-270.
8. Holmes KV: **SARS-associated coronavirus.** *N Engl J Med* 2003, **348:**1948-1951.
9. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, Rollin PE, Dowell SF, Ling AE, Humphrey CD, Shieh WJ, Guarner J, Paddock CD, Rota P, Fields B, DeRisi J, Yang JY, Cox N, Hughes JM, LeDuc JW, Bellini WJ, Anderson LJ: **A novel coronavirus associated with severe acute respiratory syndrome.** *N Engl J Med* 2003, **348:**1953-1966.
10. Poutanen SM, Low DE, Henry B, Finkelstein S, Rose D, Green K, Tellier R, Draker R, Adachi D, Ayers M, Chan AK, Skowronski DM, Salit I, Simor AE, Slutsky AS, Doyle PW, Krajden M, Petric M, Brunham RC, McGeer AJ: **Identification of severe acute respiratory syndrome in Canada.** *N Engl J Med* 2003, **348:**1995-2005.
11. Tsang KW, Ho PL, Ooi GC, Yee WK, Wang T, Chan-Yeung M, Lam WK, Seto WH, Yam LY, Cheung TM, Wong PC, Lam B, Ip MS, Chan J, Yuen KY, Lai KN: **A cluster of cases of severe acute respiratory syndrome in Hong Kong.** *N Engl J Med* 2003, **348:**1977-1985.
12. Donnelly CA, Ghani AC, Leung GM, Hedley AJ, Fraser C, Riley S, Abu-Raddad LJ, Ho LM, Thach TQ, Chau P, Chan KP, Lam TH, Tse LY, Tsang T, Liu SH, Kong JH, Lau EM, Ferguson NM, Anderson RM: **Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong.** *Lancet* 2003, **361:**1761-1766.
13. Nicholls JM, Poon LL, Lee KC, Ng WF, Lai ST, Leung CY, Chu CM, Hui PK, Mak KL, Lim W, Yan KW, Chan KH, Tsang NC, Guan Y, Yuen KY, Peiris JS: **Lung pathology of fatal severe acute respiratory syndrome.** *Lancet* 2003, **361:**1773-1778.
14. Peiris JS, Chu CM, Cheng VC, Chan KS, Hung IF, Poon LL, Law KI, Tang BS, Hon TY, Chan CS, Chan KH, Ng JS, Zheng BJ, Ng WL, Lai RW, Guan Y, Yuen KY: **Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study.** *Lancet* 2003, **361:**1767-1772.
15. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO: **Individuality and variation in gene expression patterns in human blood.** *Proc Natl Acad Sci USA* 2003, **100:**1896-1901.
16. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R: **A gene-expression signature as a predictor of survival in breast cancer.** *N Engl J Med* 2002, **347:**1999-2009.
17. Ramaswamy S, Ross KN, Lander ES, Golub TR: **A molecular signature of metastasis in primary solid tumors.** *Nat Genet* 2003, **33:**49-54.
18. Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D, Levy R: **Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes.** *N Engl J Med* 2004, **350:**1828-1837.
19. Spentzos D, Levine DA, Ramoni MF, Joseph M, Gu X, Boyd J, Libermann TA, Cannistra SA: **Gene expression signature with independent prognostic significance in epithelial ovarian cancer.** *J Clin Oncol* 2004, **22:**4700-4710.
20. Cleator S, Ashworth A: **Molecular profiling of breast cancer: clinical implications.** *Br J Cancer* 2004, **90:**1120-1124.
21. Robison JE, Perreard L, Bernard PS: **State of the science: molecular classifications of breast cancer for clinical diagnostics.** *Clin Biochem* 2004, **37:**572-578.
22. Chen CH: **Generalized association plots: information visualization via iteratively generated correlation matrices.** *Statistica Sinica* 2002, **12:**7-29.
23. Alter O, Brown PO, Botstein D: **Singular value decomposition for genome-wide expression data processing and modeling.** *Proc Natl Acad Sci USA* 2000, **97:**10101-10106.
24. Zhang L, Yu W, He T, Yu J, Caffrey RE, Dalmasso EA, Fu S, Pham T, Mei J, Ho JJ, Zhang W, Lopez P, Ho DD: **Contribution of human alpha-defensin 1, 2, and 3 to the anti-HIV-1 activity of CD8 antiviral factor.** *Science* 2002, **298:**995-1000.
25. Ganz T: **Defensins: antimicrobial peptides of innate immunity.** *Nat Rev Immunol* 2003, **3:**710-720.
26. Rosenberg HF, Tenen DG, Ackerman SJ: **Molecular cloning of the human eosinophil-derived neurotoxin: a member of the ribonuclease gene family.** *Proc Natl Acad Sci USA* 1989, **86:**4460-4464.
27. Domachowske JB, Dyer KD, Bonville CA, Rosenberg HF: **Recombinant human eosinophil-derived neurotoxin/RNase 2 functions as an effective antiviral agent against respiratory syncytial virus.** *J Infect Dis* 1998, **177:**1458-1464.
28. Robinson WS: **A method for chronologically ordering archaeological deposits.** *American Antiquity* 1951, **16:**191-301.
29. Hurley CB: **Clustering visualization of multidimensional data.** *J Computational & Graphical Statistics* 2004, **13:**788-806.
30. Pugin J, Auckenthaler R, Mili N, Janssens JP, Lew PD, Suter PM: **Diagnosis of ventilator-associated pneumonia by bacteriologic analysis of bronchoscopic and nonbronchoscopic "blind" bronchoalveolar lavage fluid.** *Am Rev Respir Dis* 1991, **143:**1121-1129.
31. Hampton GM, Frierson HF: **Classifying human cancer by analysis of gene expression.** *Trends Mol Med* 2003, **9:**5-10.
32. Wilson CL, Ouellette AJ, Satchell DP, Ayabe T, Lopez-Boado YS, Stratman JL, Hultgren SJ, Matrisian LM, Parks WC: **Regulation of intestinal alpha-defensin activation by the metalloproteinase matrilysin in innate host defense.** *Science* 1999, **286:**113-117.
33. Wang YH, Lin AS, Chao TY, Lu SN, Liu JW, Chen SS, Lin MC: **A cluster of patients with severe acute respiratory syndrome in a chest ward in southern Taiwan.** *Intensive Care Med* 2004, **30:**1228-31. Epub 2004 Apr 23..
34. **The World Health Organization:Case Definitions for Surveillance of Severe Acute Respiratory Syndrome (SARS)** [http://www.who.int/csr/sars/casedefinition/en]
35. Wang TH, Lee YS, Chen ES, Gong WH, Chen LK, Hsueh DW, Wei ML, Wang HS, Lee YS: **Establishment of cDNA microarray analysis at the Genomic Medicine Research Core Laboratory (GMRCL) of Chang Gung Memorial Hospital.** *Chang Gung Medical Journal* 2004, **27:**243-260.
36. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95:**14863-14868.
37. **Generalized association plots (GAP)** [http://gap.stat.sinica.edu.tw/Software/GAP/]