Hypervariable genes—experimental error or hidden dynamics

Igor Dozmorov*, Nicholas Knowlton, Yuhong Tang, Alan Shields, Parima Pathipvanich, James N. Jarvis[1] and Michael Centola

Department of Arthritis and Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA and [1]Department of Pediatrics, University of Oklahoma College of Medicine, Oklahoma City, OK 73104, USA

ABSTRACT

In a homogeneous group of samples, not all genes of high variability stem from experimental errors in microarray experiments. These expression variations can be attributed to many factors including natural biological oscillations or metabolic processes. The behavior of these genes can tease out important clues about naturally occurring dynamic processes in the organism or experimental system under study. We developed a statistical procedure for the selection of genes with high variability denoted hypervariable (HV) genes. After the exclusion of low expressed genes and a stabilizing log-transformation, the majority of genes have comparable residual variability. Based on an *F*-test, HV genes are selected as having a statistically significant difference from the majority of variability stabilized genes measured by the 'reference group'. A novel *F*-test clustering technique, further noted as '*F*-means clustering', groups HV genes with similar variability patterns, presumably from their participation in a common dynamic biological process. *F*-means clustering establishes, for the first time, groups of co-expressed HV genes and is illustrated with microarray data from patients with juvenile rheumatoid arthritis and healthy controls.

INTRODUCTION

Even when working with samples from a homogeneous group, we are able to observe a portion of genes that have high variability among individuals, which cannot be explained by experimental error. Genes determined to be significantly more variable compared to other genes in the same sample, bear information about some non-synchronized dynamic events in an otherwise homogeneous group. We developed a method for selection of this group of genes of higher variability and named them the 'hypervariable genes (HV genes)' based on analysis of residuals of normalized expression using an *F*-criterion.

The accuracy of the procedure will be discussed from both statistical and biological viewpoints. HV genes are determined with a threshold of $P < 1/N$ (where N is the number of genes expressed above background) such that the probability for their appearance by chance is negligible. The additional validation of their biological relevance is obtained with a clustering procedure, which demonstrates the existence of co-expressed genes whose biological interconnection indicates that the appearance of the clustering could not be due to chance.

Our clustering methodology, the '*F*-means' clustering, is based on the use of a statistical criterion (*F*-test) and produces as many different clusters as it is possible to discriminate at the accuracy of the experimental technique, i.e. microarray. The number of clusters and constituents of each cluster appeared independent of subjective decisions, such as in the case of hierarchical and *k*-means cluster methods.

The interrelationships between the clusters are studied here by a visual representation of gene correlation called a 'mosaic'. The correlation in expression levels across different samples are demonstrated to help identify genes that are regulated by a common mechanism or have similar function. Observed alterations and modulations of this mosaic in patient samples against controls visualize key changes in gene regulatory interrelations and could lead to new insight into pathological genesis.

MATERIALS AND METHODS

Patient selection and preparation of clinical specimens

We studied children newly diagnosed with polyarticular juvenile rheumatoid arthritis (JRA). Children were excluded if they had been treated with corticosteroids, methotrexate or therapeutic doses of nonsteroidal anti-inflammatory drugs for more than 3 weeks. Patients with active disease ranged in age from 3 to 15 years, and presented with proliferative synovitis of multiple joints and erythrocyte sedimentation rates ranging from 35 to 100 mm/h. Control subjects were laboratory volunteers under 25 years of age. Leukocyte buffy coat preparations were made from peripheral blood and total RNA was extracted with Trizol-reagent (Invitrogen, Carlsbad, CA). Fluorescent-labeling of cDNA was undertaken using the Micromax

*To whom correspondence should be addressed at Department of Arthritis and Immunology, Oklahoma Medical Research Foundation, 825 NE, 13th Street, Oklahoma City, OK 73104, USA. Tel: +1 405 271 7052; Fax: +1 405 271 1339; E-mail: igor-dozmorov@omrf.ouhsc.edu

TSA-labeling kit (Perkin Elmer Life Sciences, Boston, MA). Labeled cDNAs were hybridized with Perkin Elmer Micromax cDNA arrays containing 2400 human genes, and arrays were scanned using an Affymetrix 428 Array Scanner. Designations for the sample groups: AD—acute disease; AP—acute disease treated, persistent; PR—partially responsive to treatment; FR—fully responsive to treatment; HD—healthy donors. Further details of the patient and control populations have been previously published (1).

Establishment of the internal standard, the 'reference group'

After data normalization as our previous publication (2), residuals are created from the control group and they approximate a normal distribution, based on the Kolmogorov–Smirnov criterion. The expressed genes are selected out of the group of residuals whose gene expression is above background noise levels by a t-test ($P < 0.05$). Following steps are used to create the 'reference group'. First, the SD of all residuals taken together is calculated along with the SD for every gene individually. Next, an F-test is performed for every gene against the reference group to determine whether each gene should be included in the reference group. All genes whose F-statistic was significant ($\alpha = 0.05$) or whose SD was higher than the total group's SD are removed to reduce variability in the group. This process is repeated with the new smaller subset until the group is homoscedastic. These residuals present an internal standard of measurement (ISM) for the baseline variations introduced by instrumental errors and stochastic fluctuations present among samples.

Selection of 'hyper-variable genes' (HV genes)

By comparing a gene's variability to the ISM's variability through an F-test ($P = 1/N$), we are able to select genes that exhibit variation above the predetermined baseline measurement. The threshold selected, $P = 1/N$, is a less conservative version of the Bonforonni correction, $P < \alpha/N$, for multiple-hypothesis testing. It is important to note that there are HV genes even within homogeneous group of samples such as the control or treatment groups. Their hypervariability can significantly exceed the established ISM and therefore may reflect some non-synchronized gene expression dynamics. Their expressions in given samples are considered snapshots of some biological process in which they participate. Correlation of these expressions reflects some functional interconnections in the aforementioned dynamical processes.

F-means clustering algorithm

After establishing the ISM, all HV genes are clustered using an additional F-test. Begin by comparing the variability of HV gene 1 to the variability of every other HV gene on a one-by-one basis. Repeat for genes 2 through N. Next, sort all genes based on the number of associations to other HV genes. Association can be defined as follows: Gene X is associated to Gene Y if and only if the difference between the variances of Gene X and Gene Y is less than the ISM variance as measured by an F-test with a user-specified alpha level. Next, let the number of genes associated be denoted 'connectivity'. Cluster 1 comprises the gene with the highest connectivity and all genes

that are 'associated'. The gene of the second highest connectivity and all of its peers comprise cluster 2. This sequence is repeated until all genes are examined. Genes that appear in more than one cluster are considered to be likely functional links among these clusters. Genes that have zero connectivity do not belong to any cluster.

The clustering procedure can be summarized in the following steps:

- Gene expression normalization, log-transformation and rescaling as noted above.
- Identification of and limiting subsequent analyses to gene whose expressions determined to be significantly different from normally distributed background spots through a Student's t-test with $\alpha = 0.05$.
- Determination of the ISM.
- Identification of HV genes by comparing a gene's residual variability with the ISM through an F-test with $\alpha = 1/N$.
- Determination of connectivity for each of hypervariable gene.
- HV genes contained within each group were sorted by their connectivity and the clustering process is started.
- Gene co-expression and cluster relationships are represented by a correlative mosaic.

RESULTS

Analysis of residuals for selection of hypervariable genes

Array expression data among groups of samples were first normalized and log-transformed (see Materials and Methods). Residuals—deviations of the expression values from common regression line in robust regression analysis—were calculated as described previously (2). After log-transformation and exclusion of weakly expressed genes, the residuals have a nearly normal distribution (Figure 1A). Besides that the residuals of the majority of genes have relatively homogeneous variations following closely to an F-distribution (Figure 1B). However, a portion of these residuals has enormously high variability as judged by F-test based on the SD of replicated residuals for a given gene against the SD of all other genes. We used a very high threshold level ($P < 0.0005$ for approximately 2000 genes expressed distinctively from background) for selection of hypervariable genes, which makes it unlikely for the selections to appear just by chance. The level of variability of these genes as well as the proportion of genes being hypervariable exceeds significantly the statistically expected ranges. It intrigues us to consider this group of genes as distinctive from the majority of biologically stable genes, and to look for some biological rationale for their appearance in the homogeneous group of samples.

HV genes exist in all groups of samples. Hypervariations appearing from experimental errors (influence of dirty spots, etc.) were excluded from this analysis statistically—comparing the variability of the residuals in replicated group of samples with the variability obtained after excluding either the maximum or minimum one at a time. A statistical decrease in variability after excluding one replicate provides evidence of possible error in that particular replicate. Such genes were excluded from the family of HV genes as being falsely selected.
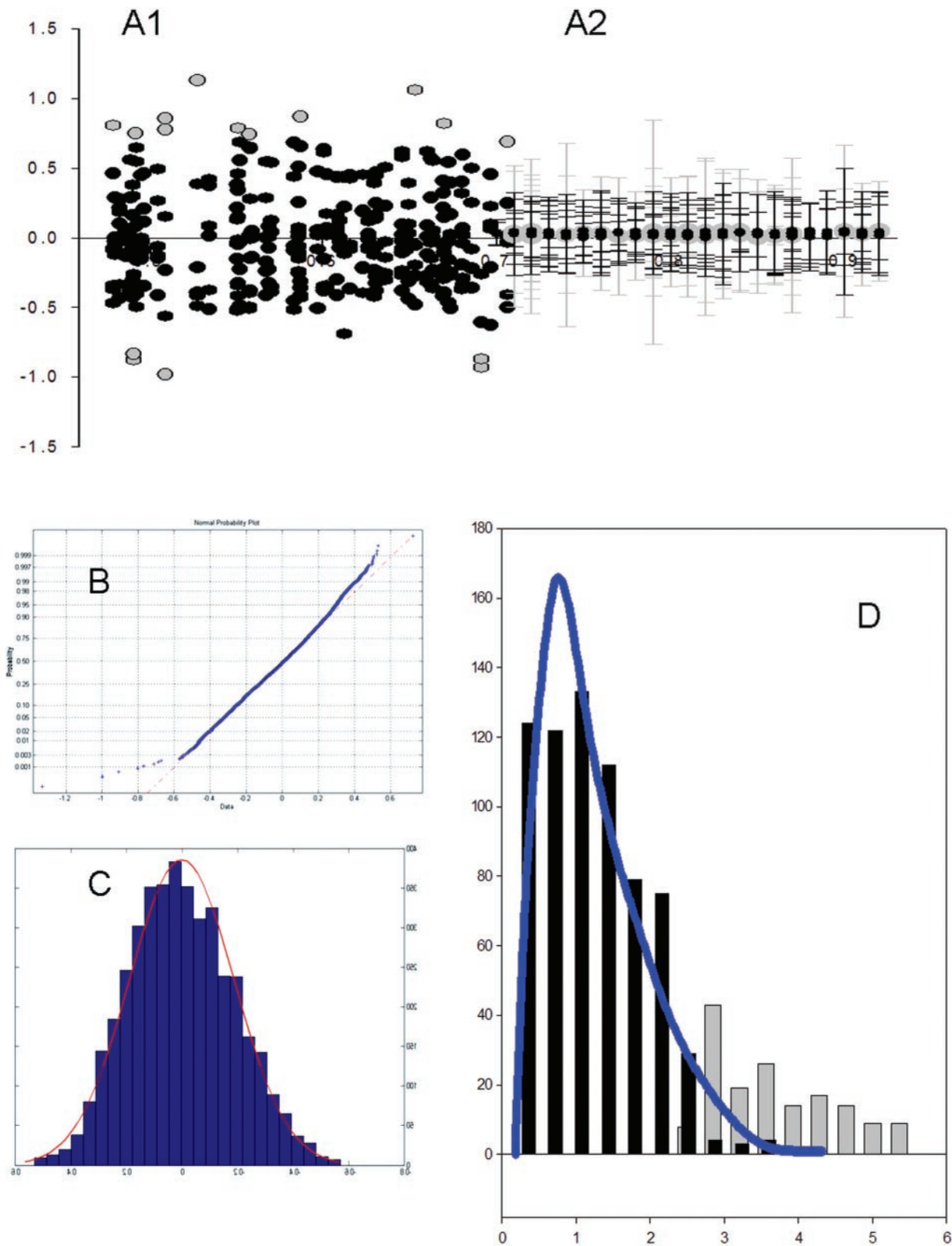
Figure 1. (**A1**) Scatter plot of residuals derived from common regression line in a robust regression analysis as described in Materials and Methods. The points represent residuals where the black are of normal variability and the grey are outliers. (**A2**) A continuation of A1 where instead of points, the variability is expressed by error bars. The black represent expected variability and the grey signify HV genes. The determination of HV is different for each expression level and is determined by an *F*-test. (**B**) A normality plot revealing the normality of the log-transformed non-background spots. (**C**) A histogram of the log-transformed non-background spots superimposed by a red normal distribution line. (**D**) Vertical bar chart of the variance ratio between individual genes and the variability of the reference group. The black lines represent the ratio frequencies of non-HV genes. The blue line is a superimposed *F*-distribution showing expected frequencies. The grey bars show the frequency of HV genes as they distort the tail of the *F*-distribution.

Co-expression of HV genes

To search for biological meaning of the HV genes, we carried out 'F-means' clustering analysis of 27 expression profiles derived from different stages of JRA and healthy donors—AD, AP, PR, FR and HD groups. The abbreviation for each group was described in Material and Methods. This procedure selects genes whose expression is different from the reference group, but similar to one another. The simulated data was generated using a Monte Carlo simulation for each of the 27 samples based on its expression average and variability. Clustering for the simulation data provides a statistical validation of the F-means clustering procedure for HV genes:

- No cluster was identified in the simulation data from all 27 samples. In contrast, clustering of the real data produced 15 clusters with 2 and more members (maximum—5 members). Two biggest clusters are presented in Figure 2A and B.

- Only one cluster with 3 members was obtained in simulation of control data of 10 samples (HD and FR patients). Real data analysis gave 26 such clusters (maximum 17 members). Two biggest clusters are presented in Figure 2C and E.

- No clusters with 3 and more genes were produced in simulation of data from AD and AP patients with 13 samples. Clustering of the real data produced 12 clusters with 3 and more members (maximum 7 members). The biggest cluster of this group is presented in Figure 2F.

It is necessary to emphasize that as opposed to a longitudinal study, the shapes of the clustered profiles presented in each plot of Figure 2 are the result of an arbitrary arrangement of samples. The co-expression of the selected genes in all 27 samples bear the biological significance rather than the clusters' shape, which can be manipulated through sample arrangement as demonstrated in Figure 3.

The presence of gene clusters independent of sample source in the control and patient data but not in the simulated data
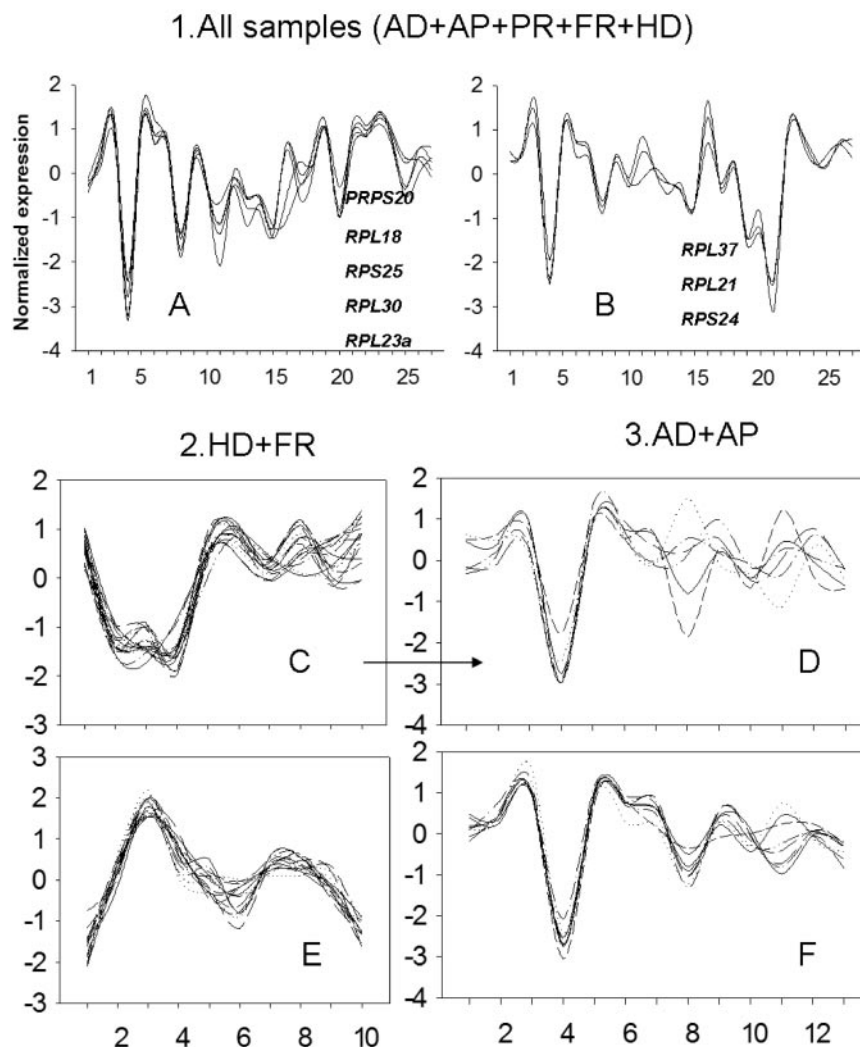

Figure 2. (**A** and **B**) F-means clustering of genes from 27 patients and healthy controls [patients with acute disease (AD), non-responsive to the treatment (persistent) (AP), partially responsive to treatment (PR), demonstrating full response (FR) and from control group of healthy donors (HD)]. Here the two largest clusters are presented consisting exclusively of ribosomal genes. (**C** and **E**) The two largest clusters in combined HD and FR groups. (**D** and **F**) Gene clustering of two groups with acute disease. (D) Genes from largest cluster HD and FR (C) having similar patterns in AD and AP groups. (F) Largest cluster in AD and AP groups.
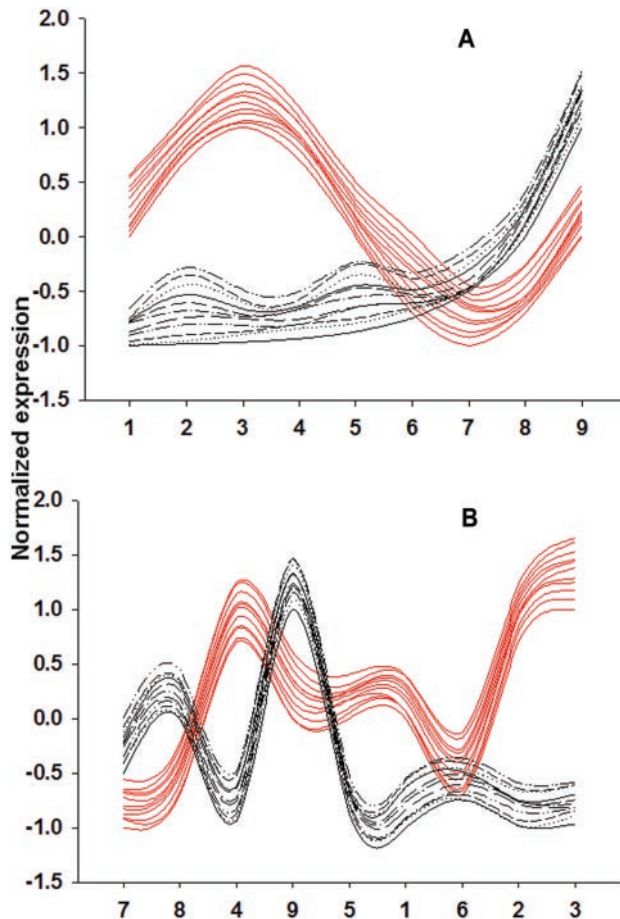
Figure 3. Diagrams illustrating the formation of the cluster profiles for HV genes in a homogeneous group. (**A**) Possible assortment of nine samples representing two dynamical processes with participation of several genes each whose profiles are shown in either red or black. (**B**) Variant of (**A**) in which only the order of samples was changed. This is valid since all samples were collected simultaneously and are part of a homogeneous group. The gene co-expression is preserved in all possible arrangements of samples and in these co-expressions is where *F*-means clustering can tease out the involvement of these genes in common dynamical processes.

provide compelling evidence that the appearance of the cluster is not random, and there are potential functional interconnections between genes in the same cluster. The common functionality of genes in a single cluster can be seen in the clusters presented in Figure 2A and B as both clusters consist of solely ribosomal proteins.

There is evidence of the uniformity in these clusters' constituents. A significant portion of genes clustered in the biggest cluster in HD and FR group of samples (Figure 2C) had a very similar profile to the profile in AD and AP samples (Figure 2D) akin to the expression profile for the bigger cluster in this group (Figure 2F).

The fact that co-expression of some genes is reproducible across different groups of samples between patients and controls, such as in Figure 2A and B indicates that these clusters represent genes involved in some biological processes independent of the disease pathology under investigation. In the mean time, some other clusters, which could not be reproduced between control and patient groups (data not shown), could be involved in the pathogenesis of the disease under study and will be manifested in the next section.

Comparative analysis of co-expressed genes in different groups

With the use of colored correlation mosaics, complicated interdependencies between genes can be visualized and differences between subgroups quickly assessed (Figure 4). The image is a visual representation of a correlation coefficient matrix with different colors symbolizing different strength of correlation with strongest positive correlation to be 1 and the strongest negative correlation being −1. For this presentation, we selected genes from the two largest HV clusters obtained from healthy donors. Genes HV in control, but stable or non-expressed in disease were excluded from this analysis. The HD group reveals the presence of two highly correlated clusters of genes represented by two orange squares. These highly correlated genes in a single group of HD exhibit some differences when studied in both patient groups of non-treated (AD) and treated partially responding (PR) patients as demonstrated by the change of color on the mosaic. Gene orders are kept the same between these two groups and along the axes.

When attention was focused on the genes with altered functional interconnections within each clusters taking place in AD and PR patients, their involvement in the pathology become obvious. Information from these genes and their relationship to known pathology (here JRA) was obtained from literature and is presented in Table 1.

DISCUSSION

Our analysis of microarray expression data suggests that after background correction and log-transformation, the majority of the genes expressed significantly above background have stable normal errors, and this result is in good agreement with the approximate analysis in (16). After that it seems to be possible to apply an *F*-test for the selection of the small portion of genes having statistically significant increase in variability, which in our terminology are HV genes. HV genes are selected as genes expressed significantly higher than background and having statistically higher variations based on use standard *F*-test with threshold high enough to prevent the appearance of false-positive selections. To validate the biological significance of these variations, HV genes were clustered at a threshold of $P < 0.0005$ to exclude random affiliations. Clusters of HV genes provide empirical evidence that the observed variations are not random fluctuations of individual gene expressions (errors), but the reflections of multi-participatory dynamical processes. This is especially evident when cluster analysis is able to reveal group of genes reproducibly co-expressed across sampling conditions (control and patients).

A special provision is made within our algorithm for exclusion from HV gene selections those genes whose high variability is produced by the erroneous signal in one replicate. These precautions take into account the opportunity for the appearance of a single mistake in the set of gene measurements within group because the probability for more than one mistake is negligibly low in our experience.
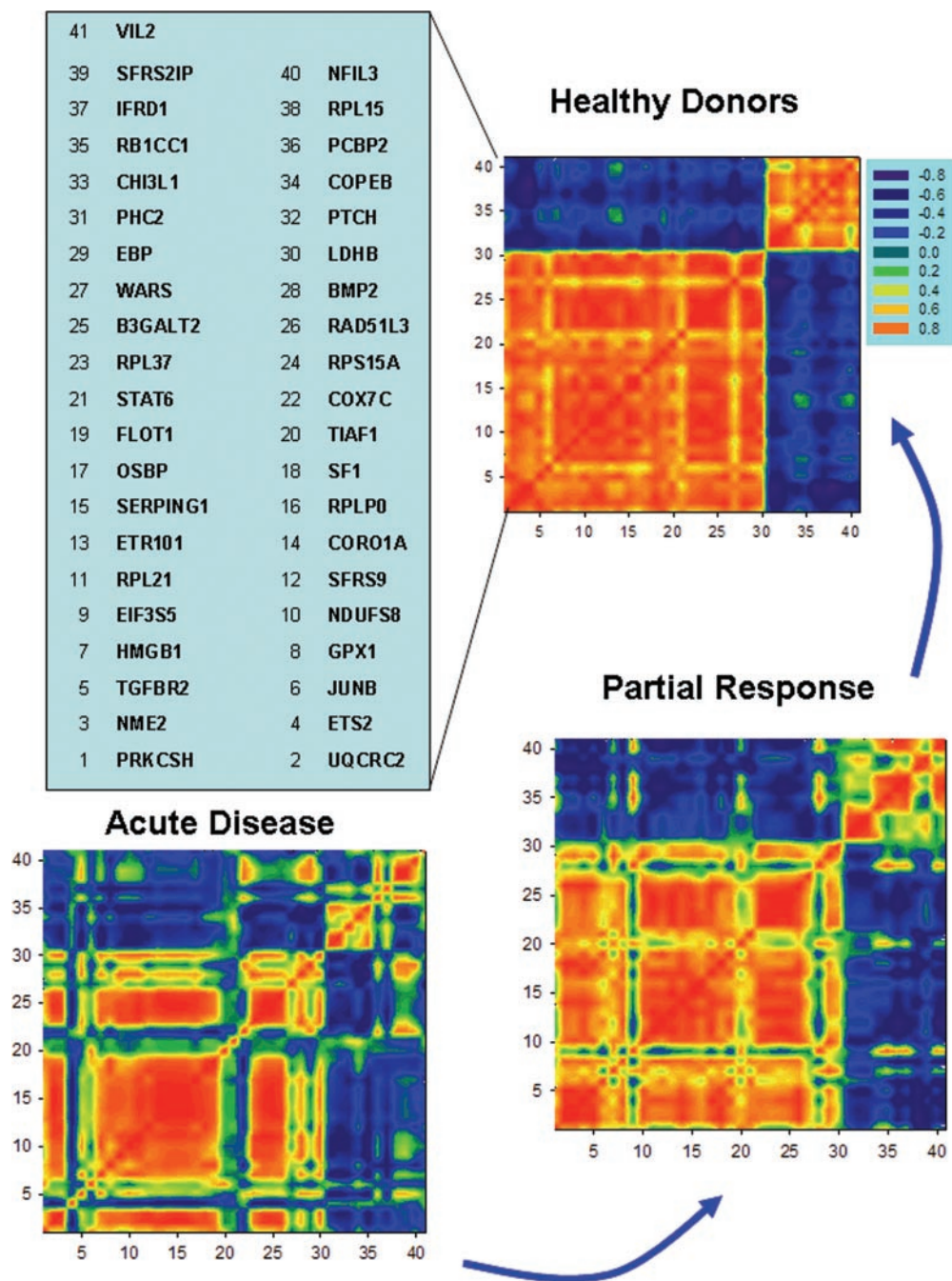
Figure 4. Correlation mosaics for genes from the two largest clusters in the control group. Each spot in the plot presents correlation coefficients of expressions for genes along the axes. A red spot is highly correlated, conversely a blue spot is highly anti-correlated. Gene order is chosen to present joined co-expressed genes in two largest clusters of the HD samples. The same order of the genes along axis is used for all three mosaics.

Identification of HV genes within homogeneous groups of samples provided evidences for some dynamical processes normally taking place in an organism's physiology. These processes are not static; consequently, each sample gives a snapshot of the same process much like an observer gazing into a hall of mirrors sees an object from many vantage points, but is unable to discern the 'true' image. Due to the third person view of the dynamics in play, we are unable to tell the order of the phases, only that they all belong to the same biological process. Moreover, these dynamical processes are not chaotic, but in a coordinated system as genes involved in a common process fall into a single cluster. Changes in the dynamical coordination evidenced by the alteration of cluster constituents between control and patients may contribute to the pathology of a disease by destroying the sensitive balance between biologic pathways.

Here we used a novel clustering technique based upon empirical variance estimates. This procedure is free from common drawbacks of traditional clustering algorithms because the decision determining the number of clusters and their

Table 1. Genes of different co-expression patterns in pathology compared with health donors as displayed by mosaics of correlations in Figure 4

Order on Figure 4/symbol/name	References and description
4/ETS2/V-ets erythroblastosis virus E26 oncogene homolog 2	(3) There was an increased expression of ets-2 in 30% of rheumatoid arthritis samples and an up to 30-fold decreased expression of the potential metastasis suppressor gene nm23-H1 in 90% of RA tissues, compared with control tissues.
	(4) demonstrated a role for the ETS1 and ETS2 transcription factors in regulating the expression of p16(INK4A) in these different contexts based on their ability to activate the p16(INK4A) promoter through an ETS binding site and their patterns of expression during the life span of human diploid fibroblasts.
	(5) They studied the expression of cyclin-dependent kinase inhibitors in rheumatoid synovial cells as a means of suppressing synovial cell proliferation. Synovial cells derived from hypertrophic synovial tissue readily expressed p16INK4a when they were growth-inhibited. This was not seen in other fibroblasts, including those derived from normal and osteoarthritis-affected synovial tissues. *In vivo* adenoviral gene therapy with the p16INK4a gene efficiently inhibited the pathology in an animal model of rheumatoid arthritis. Thus, the induction of p16INK4a may provide a new approach to the effective treatment of rheumatoid arthritis.
6/JUNB/Jun B proto-oncogene	(6) Differential junB mRNA expression in synovial membranes and isolated synovial fibroblasts from rheumatoid arthritis patients
20/ TIAF1/TGFB1-induced anti-apoptotic factor 1	(7) TGFB1-induced anti-apoptotic factor 1 inhibits tumor necrosis factor cytotoxicity.
	(8) TIAF1 binds to Jak3 and inhibits apoptosis induced by IL-2 deprival.
	(9) TIAF1 protects murine L929 fibroblasts from TNF cytotoxicity. TIAF1 induced growth inhibition and apoptosis of monocytic U937 and other types of cells. TIAF1 increased the expression of p53, Cip1/p21, and Smad proteins; suppressed ERK phosphorylation; and altered TGF-beta1-mediated Smad2/3 phosphorylation in U937 cells. Antisense TIAF1 mRNA significantly enhanced the proliferation of mink lung Mv1Lu epithelial cells. Together, these observations indicate that TIAF1 participates in the TGF-beta-mediated growth regulation.
	(10) T helper cells are important modulators of the allograft immune response. A small number of genes are already known to be differentially expressed in T helper 1 (Th1) and T helper 2 (Th2) cells, but it is likely that many other genes are functionally important. To study gene expression in Th1 and Th2 cells, we used serial analysis of gene expression. One of the differentially expressed genes was TIAF-1, which is a TGF-beta 1-induced antiapoptotic factor, known to inhibit the cytotoxic effects of tumor necrosis factor-alpha on mouse fibroblasts. We hypothesized that TIAF-1 differentially expressed in Th1 and Th2 cells and plays a protective role against apoptosis during allograft rejection.
	TIAF-1 mRNA and protein were not detectable in normal kidney and liver; however, the expression of TIAF-1 was up-regulated in most biopsy specimens with chronic and a few with acute allograft rejection. Immunohistochemistry for TIAF-1 revealed expression in the inflammatory infiltrate and in tubular epithelial cells. CONCLUSIONS: TIAF-1 mRNA and protein are predominantly up-regulated in kidney and liver allografts with chronic rejection. Expression of TIAF-1 in the lymphocytes during chronic allograft rejection may be related to the predominance of a Th2 response in this condition. The expression in the transplanted tissue may protect these cells from apoptosis.
21/STAT6/signal transducer and activator of transcription 6, interleukin-4 induced	(11) CD23 is atypically highly expressed in various chronic diseases, including B-CLL, lupus erythematodes and rheumatoid arthritis. Cooperation between NF-AT and STAT factors might be one of the molecular mechanisms responsible for high-level expression of CD23 on the surface of B-CLL cells.
	(12) Serum soluble CD23 levels and CD23 expression on peripheral blood mononuclear cells in juvenile chronic arthritis.
	(13) The finding that STAT6(−/−) T(h)2 cells did not show any down-regulation of VLA-2 expression and expressed the same levels of VLA-2 as T(h)1 cells indicated a critical role for the IL-4 receptor/STAT6 signaling pathway in IL-4-dependent down-regulation of VLA-2 on T(h)2 cells. Our results indicate that VLA-2 is a novel functional marker that dissociates T(h)1 from T(h)2 cells, and thus might be useful for therapeutic monitoring of T(h)1-dependent immune diseases such as rheumatoid arthritis or Crohn's disease.
27/WARS/tryptophanyl-tRNA synthetase	(14) Sera of patients bearing autoimmune diseases (rheumatoid arthritis and systemic lupus erythematosus) show the presence of autoantibodies against tryptophanyl-, tyrosyl- and phenylalanyl-tRNA synthetases.
28/BMP2/bone morpho-genetic protein 2	(15) Bone morphogenetic proteins 2 and 6, expressed in arthritic synovium, are regulated by proinflammatory cytokines and differentially modulate fibroblast-like synoviocyte apoptosis.

constituents is based on statistical estimates derived from the data not determined a priori by the researcher.

Clustering procedures were broadly applied for identification of differential gene expressions under a wide range of conditions (17). Here, we demonstrated the application of this clustering procedure to characterize hypervariable genes in homogeneous samples. Validation of this method comes from the following:

- significantly more cluster constituents than can be expected by chance;
- functional homogeneity of the cluster constituents [biggest clusters in Figure 2A and B consist exclusively of ribosomal genes; co-expression of ribosomal genes was also described in (18,19)];

- reproducibility of the cluster constituents of biggest clusters between different groups of samples (Figure 2C and D);

The fact that the majority of HV genes with altered pathological interrelationships (as seen from comparison of the correlation mosaics—Figure 4) are characterized in literature as associated with various manifestations of the inflammation and rheumatoid arthritis [Table 1, please refer to our previous publication (1)]. These findings indicate that appearance of high variation genes in homogeneous groups could be the manifestation of some biological processes taking place in normal tissues and modulated by pathological processes in patients samples.

ACKNOWLEDGEMENTS

REFERENCES

1. Jarvis,N.J., Dozmorov,I., Jiang,K., Frank,M.B., Szodoray,P., Alex,P. and Centola,M. (2003) Novel approaches to gene expression analysis of active polyarticular juvenile rheumatoid arthritis. *Arthritis Res. Ther.*, **6**, R15–R31.
2. Dozmorov,I.M. and Centola,M. (2003) An associative analysis of gene expression array data. *Bioinformatics*, **19**, 204–211.
3. Dooley,S., Herlitzka,I., Hanselmann,R., Ermis,A., Henn,W., Remberger,K., Hopf,T. and Welter,C. (1996) Constitutive expression of c-fos and c-jun, overexpression of ets-2, and reduced expression of metastasis suppressor gene nm23-H1 in rheumatoid arthritis. *Ann. Rheum. Dis.*, **55**, 298–304.
4. Ohtani,N., Zebedee,Z., Huot,T.J.G., Stinson,J.A., Sugimoto,M., Ohashi,Y., Sharrocks,A.D., Peters,G. and Hara,E. (2001) Opposing effects of Ets and Id proteins on p16 (INK4A) expression during cellular senescence. *Nature*, **409**, 1067–1070.
5. Taniguchi,K., Kohsaka,H., Inoue,N., Terada,Y., Ito,H., Hirokawa,K. and Miyasaka,N. (1999) Induction of the p16INK4a senescence gene as a new therapeutic strategy for the treatment of rheumatoid arthritis. *Nature Med.*, **5**,760–767.
6. Huber,R., Kunisch,E., Gluck,B., Egerer,R., Sickinger,S. and Kinne,R.W. (2003) Comparison of conventional and real-time RT-PCR for the quantitation of jun protooncogene mRNA and analysis of junB mRNA expression in synovial membranes and isolated synovial fibroblasts from rheumatoid arthritis patients. *Z Rheumatol.*, **62**, 378–389.
7. Chang,N.S., Mattison,J., Cao,H., Pratt,N., Zhao,Y. and Lee,C. (1998) Cloning and characterization of a novel transforming growth factor-beta1-induced TIAF1 protein that inhibits tumor necrosis factor cytotoxicity. *Biochem. Biophys. Res. Commun.*, **253**, 743–749.
8. Ji,H., Zhai,Q., Zhu,J., Yan,M., Sun,L., Liu,X. and Zheng,Z. (2000) A novel protein MAJN binds to Jak3 and inhibits apoptosis induced by IL-2 deprival. *Biochem. Biophys. Res. Commun.*, **270**, 267–271.
9. Khera,S. and Chang,N.S. (2003) TIAF1 participates in the transforming growth factor beta1-mediated growth regulation. *Ann. NY Acad. Sci.*, **995**, 11–21.
10. van der Leij,J., van den Berg,A., Albrecht,E.W., Blokzijl,T., Roozendaal,R., Gouw,A.S., de Jong K.P., Stegeman,C.A., van Goor,H., Chang,N.S. and Poppema,S. (2003) High expression of TIAF-1 in chronic kidney and liver allograft rejection and in activated T-helper cells. *Transplantation*, **75**, 2076–2082.
11. Kneitz,C., Goller,M., Tony,H., Simon,A., Stibbe,C., Konig,T., Serfling,E. and Avots,A. (2002) The CD23b promoter is a target for NF-AT transcription factors in B-CLL cells. *Biochim. Biophys. Acta*, **1588**, 41–47.
12. Massa,M., Pignatti,P., Oliveri,M., De Amici,M., De Benedetti,F. and Martini,A. (1998) Serum soluble CD23 levels and CD23 expression on peripheral blood mononuclear cells in juvenile chronic arthritis. *Clin. Exp. Rheumatol.*, **16**, 611–616.
13. Sasaki,K., Tsuji,T., Jinushi,T., Matsuzaki,J., Sato,T., Chamoto,K., Togashi,Y., Koda,T. and Nishimura,T. (2003) Differential regulation of VLA-2 expression on Th1 and Th2 cells: a novel marker for the classification of Th subsets. *Int. Immunol.*, **15**, 701–710.
14. Bolgarin,R.N., Rodnin,N.V. and Sidorik,L.L. (1998) Autoantibodies to tryptophanyl-tRNA-synthetase in systemic autoimmune diseases. *Mol. Biol.*, **32**, 745–749.
15. Lories,R.J., Derese,I., Ceuppens,J.L. and Luyten,F.P. (2003) Bone morphogenetic proteins 2 and 6, expressed in arthritic synovium, are regulated by proinflammatory cytokines and differentially modulate fibroblast-like synoviocyte apoptosis. *Arthritis Rheum.*, **48**, 2807–2818.
16. Rocke,D.M. and Durbin,B. (2003) Approximate variance-stabilizing transformations for gene-expression microarray data. *Bioinformatics*, **19**, 966–972.
17. Horimoto,K. and Toh,H. (2001) Statistical estimation of cluster boundaries in gene expression profile data. *Bioinformatics*, **17**, 1143–1151.
18. Eisen,M.B., Spellman,P.T., Brown,P.O. and Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.
19. Alon,U., Barkai,N., Notterman,D.A., Gish,K., Ybarra,S., Mack,D. and Levine,A.J. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl Acad. Sci. USA*, **96**, 6745–6750.