

GENOME-WIDE SURVEY OF HOST RESPONSES: USE OF COMPUTATIONAL ANALYSIS TO CLASSIFY EXPOSURES AND EXTRACT SIGNATURES OF UNCONVENTIONAL VERSUS COMMON VIRAL EXPOSURES

Rasha Hammamieh* and Marti Jett

Molecular Pathology, Walter Reed Army Institute of Research, Silver Spring, MD 20910

ABSTRACT

Exposures to many unconventional pathogenic agents result in flu-like illness that are initially indistinguishable from common respiratory illnesses and early diagnosis to distinguish among the severe vs common viral infections depends on pathogen proliferation to dangerous, near-untreatable levels. Assessing exposure to a pathogen, in advance of onset of illness or at various stages post-exposure, is invaluable among the diagnostic options. Lymphocytes serve as “whistle blower” indicators as they encounter pathogenic agents even early during the course of infection, registering the encounters in their mRNA and developing patterns of expression that correspond to each specific pathogen. Time series of gene expression patterns relate to the stage or severity of the infection and are unique for each pathogen.

We are using the host blood for determination of whole genome regulation in response to various viral agents to extract features and signatures that can be used for point-of-care diagnosis of various viral infections (common respiratory, arena, flavi-, alpha- and other viruses). These data also have the potential to provide stage-appropriate therapeutic targets.

These studies utilized exposure time sequences of host gene expression. Series #1 contained common respiratory viruses (influenza A, parainfluenza, rhinovirus, respiratory syncytial virus). Series #2 focused on 5 arena viruses (the highly virulent hemorrhagic virus, Lassa, and 2 additional pairs of virulent/avirulent arena viruses). Series #3. includes flaviviruses (West Nile virus and 4 serotypes of Dengue), and Series #4, 2 alpha viruses, among which was Venezuelan equine encephalitis (VEE). The “training sets” were constructed from in vitro exposures to purified peripheral blood leukocytes from ~6 human leukapheresis donors for each virus described above. Numerous modeling / mathematical techniques were applied to these datasets in order to identify signature patterns indicative of each. The “shrink/grow” modeling approaches were used as well as other algorithms that have shown success for signature extraction. For the “grow” algorithm, genes are individually selected that have the best discriminating power and the first of those frequently show properties unique for specific viral infections.

Our first question related to the fact that, by necessity (due to the lack of human samples from exposures to biothreat pathogens), we were training on in vitro exposures to these many pathogenic agents. Our thesis was that such an exposure would provide a biochemical signature, related to the mechanistic course of action of each pathogenic agent and it should be reflective of at least early events from in vivo exposures (to non-human primates-NHP).

Our first “test” data set was from host gene expression responses in peripheral blood (leukocytes) from NHP exposed to VEE. This is a mosquito-borne viral disease characterized by fever and one or more of the following: severe headache, back pain, myalgias, prostration, chills, nausea, vomiting, weakness and other flu-like symptoms. The aerosol form of VEE is highly infectious, making VEE a potential biowarfare agent. This could be especially worrisome if strains are altered genetically to increase pathogenicity. If this virus was deployed efficiently, it could incapacitate thousands of people for a week or more and cause untold psychological stress as well as being physically debilitating.

Diagnosis of VEE relies on virus isolation from acute phase serum or from spinal fluid, or on detection of VEEV-specific IgM in the cerebrospinal fluid in cases of encephalitis. We carried out 2 different approaches using blood from 12-25 control (unexposed) NHP along with 8 samples from VEE-exposed NHP. Although no signs of illness were apparent by day 3, gene expression patterns identified these NHP as having mild exposure to the virus. To increase the level of confidence in the methodology, the NHP samples from 12 control and 5 VEE challenged NHP were blinded as “test” datasets. All of the 17 samples were correctly categorized as to the nature of the exposure. Using both types of training/test methods, one control NHP was categorized correctly as a “control” but had at least some indicators of common viral exposure.

Additional studies are underway to apply the most instructive of these algorithms to arena viruses (NHP exposed to Lassa virus), people exposed to dengue virus or West Nile virus or influenza A.

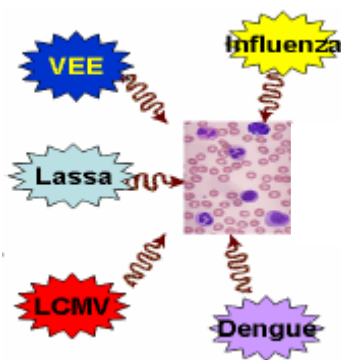
1. INTRODUCTION

Preparations for the Army of the future include reliance on high technology instrumentation for diagnostic and therapeutic approaches. This highly sophisticated hardware requires extensive computational capabilities and will have the potential to provide a wealth of

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 01 NOV 2006		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE Genome-wide Survey of Host Responses: Use of Computational Analysis to Classify Exposures and Extract Signatures of Unconventional Versus Common Viral Exposures				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Molecular Pathology, Walter Reed Army Institute of Research, Silver Spring, MD 20910				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES See also ADM002075., The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 50	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

important information to aid the warfighter. Our laboratory is focused on utilizing gene microarray technology as a tool for establishing rapid diagnostics and tailored therapeutic approaches.

We are using human peripheral mononuclear cells (PBMC) to study whole genome regulation in response to various viral agents and to identify diagnostic biomarkers for these viruses. We apply various algorithms to extract features and signatures that can be used for point-of-care diagnosis of various viral infections (common respiratory, arena, flavi-, alpha- and other viruses) (Figure 1).



We have identified host gene expression patterns that can discriminate exposure to various biological threat agents. Each of these gene patterns regulated by a specific agent reveals the cascade of events that occurs after the host encounters a pathogenic agent. Even though these pathogens initially cause similar symptoms, such as malaise, fever, headache, and cough, the course of illness induced by each of them differs in time frame of illness patterns. Using these signature gene profiles to assess possible exposure to pathogenic agents or to differentiate them from non lethal illnesses when the classical identification of a pathogen is not conclusive may fill a gap in the arsenal of diagnostic tools.

We have studied the kinetics of host responses to various groups of viruses. The first group is common respiratory viruses (influenza A, parainfluenza, rhinovirus, respiratory syncytial virus).

Second group constitute of 5 arena viruses (the highly virulent hemorrhagic virus, Lassa, and 2 additional pairs of virulent/avirulent arena viruses). The third group includes flaviviruses (West Nile virus and 4 serotypes of Dengue), and the forth group include 2 alpha viruses, among which was Venezuelan equine encephalitis (VEE).

We have used the microarray data from the in vitro exposures to various viral agents as a “training set”. This set was assembled using data from PBMC obtained from

~6 human leukapheresis donors for VEE and each of the respiratory and arena viruses.

We applied various algorithms to these datasets in order to identify signature patterns indicative of each.

Due to the lack of human samples from exposures to biothreat pathogens, we are using in vitro exposures to these many pathogenic agents. Our previous studies have shown that high throughput gene expression data obtained in vitro can, to some extent, represent in vivo exposures. Our thesis was that such an exposure would provide a biochemical signature, related to the mechanistic course of action of each pathogenic agent and it should be reflective of at least early events from in vivo exposures (to non-human primates-NHP).

We examined our hypothesis that in vitro exposures can predict exposures in vivo. We have used data from PBMC exposed to Venezuelan equine encephalitis (VEE) in vitro and in vivo using NHP.

Venezuelan equine encephalitis is a mosquito-borne viral disease characterized by fever and one or more of the following: severe headache, back pain, myalgias, prostration, chills, nausea, vomiting, weakness and other flu-like symptoms. The aerosol form of VEE is highly infectious, making VEE a potential biowarfare agent. This could be especially worrisome if strains are altered genetically to increase pathogenicity. If this virus was deployed efficiently, it could incapacitate thousands of people for a week or more and cause untold psychological stress as well as being physically debilitating.

Diagnosis of VEE relies on virus isolation from acute phase serum or from spinal fluid, or on detection of VEEV-specific IgM in the cerebrospinal fluid in cases of encephalitis.

We carried out microarray gene expression analysis to identify biomarkers for VEE. For the in vitro studies, we exposed PBMC for 6 healthy donors to VEE and RNA samples were isolated at various time point post exposure. For the in vivo, we carried out 2 different approaches using blood from 12-25 control (unexposed) NHP along with 8 samples from VEE-exposed NHP. Although no signs of illness were apparent by day 3, gene expression patterns identified these NHP as having mild exposure to the virus.

We have used microarray data from the in vivo exposures to apply predictive modeling and to identify genes that can be predictors/signatures of exposures to VEE. We identified genes that can be used as diagnostic biomarkers for VEE.

We have then used these data from VEE exposure and data from exposures to respiratory and arena viruses to apply feature extraction and predictive modeling. These data were used as a training set and the data from the in vivo NHP exposures were used as a test set. Applying feature extraction and predictive modeling algorithms, we were able to correctly predict the VEE exposures in >90% of the NHP samples.

We have also used samples from the NHP controls and VEE challenged that were blinded as “test” datasets. All of the blinded samples were correctly categorized as to the nature of the exposure.

Using the whole genome expression profiles, we were able to identify biomarkers that would distinguish VEE from arena and respiratory viruses and to determine exposures in blinded samples.

Rapid detection, before the symptoms appear or even at various stages of illness, offers the opportunity to initiate appropriate treatment. Furthermore, this technique may provide the means to identify new therapeutic approaches to ameliorate the devastating results of these pathogens.

2. EXPERIMENTAL APPROACH

Our laboratory is studying the complex interaction between various biological pathogens and the host in vivo and in vitro to understand molecular characteristics of illnesses induced by biological threat agents and to identify host defense strategies and by which they are regulated.

We have used a library of 40,000 human cDNA, of those 10,000 are human sequence verified cDNA, to construct customized microarray chips used in our studies. Experiments were carried out in replicates at each time point for each pathogen using cDNA microarrays.

RNA isolation

RNA isolation was carried out according to the Trizol[™] Reagent manufacturer's recommended instructions. RNA was ethanol-precipitated, air-dried and re-suspended in 20ul/sample of nuclease-free water. RNA quantity was measured via spectrophotometry followed by analysis with a Bioanalyzer 2100 (Agilent Technologies, CA)

Custom made cDNA Microarray Slide Preparation and Hybridization

The gene library for the present project was commercially obtained from Operon, containing ~37 K genes. Superamine coated Telechem slides (Telechem Inc., OR) were used for printing the cDNA clones using 12 X 4 pin format, on a Virtek chip writer professional microarrayer in KemTek, Inc, MD. The printed slides underwent UV cross-linking, followed by post-processed by succinic anhydride treatment. The Micromax[™] Tyramide Signal Amplification (TSA)[™] Labeling and

Detection Kit (PerkinElmer, Inc., MA) was used as directed by the manufacturer to determine relative gene expression of the collected samples. Sample hybridization was carried out at 55°C for sixteen hours. A laser detection system was used (GenePix 4000b, Axon Instruments, CA) to scan the finished slides. Intensity of the scanned images was digitalized through Genepix 4.0 software (Axon Inc., CA).

Microarray Analysis

Data cleansing and statistical analysis was carried out using Genespring® 7.0 (Agilent Tech., CA). Local background was subtracted from individual spot intensity. Genes that failed this ‘background check’ in any of the eight given experiments were eliminated from further analysis. Each chip was next subjected to intra-chip normalization (LOWESS). The genes that varied most between control and treated sample sets were selected via *t*-test analysis. The *p*-value cutoff was set at 0.05. Four hundreds and thirty two genes were differentially expressed between VEE –infected and control uninfected animals with *p*<0.05.

The pattern of gene expression variability of the experimental set having reduced dimension was evaluated using principal component analysis (PCA) classifying VEE positive and negative samples as the two variable classes.

We used the reference design where a reference RNA sample is co-hybridized with each sample on the slide. This design allows us to normalize between the slide for variations that can be due to hybridization, transcription and labeling efficiencies (technical variations). We used various modules to analyze the microarray data including GeneSpring, Partek Pro, SAM and Bioconductor. Using Analysis of Variance (ANOVA) we determined genes that exhibited variations in expression between the control samples. These variations may be due to many factors including biological and technical variations. These normally varying genes are excluded from further analysis to study gene regulation upon exposure to pathogens.

Selection of genes (biomarkers) and predictive analysis: Gene Selection

To select genes for use in the predictor, all genes are examined individually and ranked on their power to discriminate each class from all others, using the information on that gene alone. For each gene, and each class, all possible cutoff points on gene expression level for that gene are considered to predict class membership either above or below that cutoff. Genes are scored on the basis of the best prediction point for that class. The score function is the negative natural logarithm of the *p*-value for a hypergeometric test of predicted versus actual class membership for this class versus all others. A combined list containing the most discriminating genes for each class is produced as the predictor list. Each class is examined in turn, and the gene with the highest score for that class is added to the list, if it is not

already on the list. Then genes with the next highest scores for each class are added. This is continued in rotation among the classes until the specified number of predictor genes is obtained.

We applied the Fisher's Exact Test that looks for an association between expression level and class membership. Each gene is tested for its ability to discriminate between the classes. Genes with the lowest p-values are kept for the subsequent calculations. In this method, all the measurements for a given gene are ordered according to their normalized expression levels. For each class, the predictor places a mark in the list where the relative abundance of the class on one side of the mark is the highest in comparison to the other side of the mark. The genes that are most accurately segregated by these markers are considered to be the most predictive. A list of the most predictive genes is made for each class and an equal number of genes (lowest p-value using Fischer's exact test) are taken from each list.

Classifying the Test Samples

Based on the selected genes, classifications are then predicted for the independent test data, using the k-nearest-neighbors rule. A sample in the independent set is classified by finding the k nearest neighbors of the sample among the training set samples, based on Euclidean distance between the normalized expression ratio profiles of the samples. The class memberships of the neighbors are examined, and the new sample is assigned to the class showing the largest relative proportion among the neighbors after adjusting for the proportion of each class in the training set.

3. RESULTS

Study of host responses to VEE in vitro and in vivo in NHP:

We have used microarrays to study host response to VEE in vitro using human PBMC at various time points and in vivo using NHPs (Figure 2).

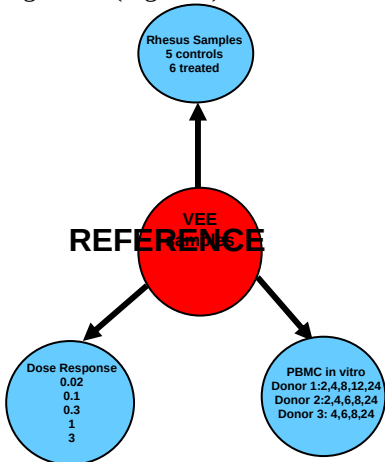


Figure 2: scheme of the experimental design for the VEE study.

Microarray analysis of host responses to VEE in vitro in PBMC:

We carried out microarray gene expression analysis of PBMC samples exposed to VEE at various time points. We identified genes that exhibited highly statistically significant differential expression in PBMC exposed to VEE at various time points when compared to control untreated cells (Figure 3).

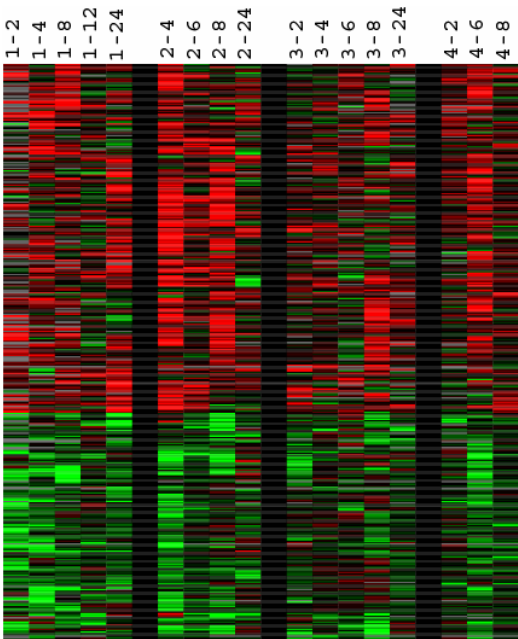


Figure 3: Cluster analysis of genes that were differentially expressed in PBMC exposed to VEE at various time points when compared to control untreated cells.

Microarray analysis of host responses to VEE in vivo in NHPs:

We carried out microarray analyses using blood samples obtained from 12 control (unexposed) NHPs along with 8 samples from VEE-exposed NHPs. Although no signs of illness were apparent by day 3, gene expression patterns identified these NHP as having mild exposure to the virus.

Genes that were significantly altered by VEE were clustered to show their expression profiles. These genes were consistently regulated by VEE (figure 4).

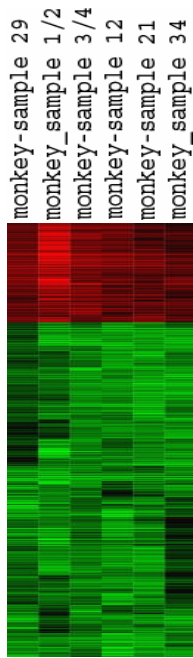


Figure 4. Cluster view and expression profiles of genes that were differentially expressed between NHPs exposed to VEE and control NHPs.

We carried out data mining and functional classification using our in-house developed tools for pathway and literature mining showed that some of the genes that were highly up-regulated by VEE were involved in apoptotic pathways and cell cycle (Figure 5).

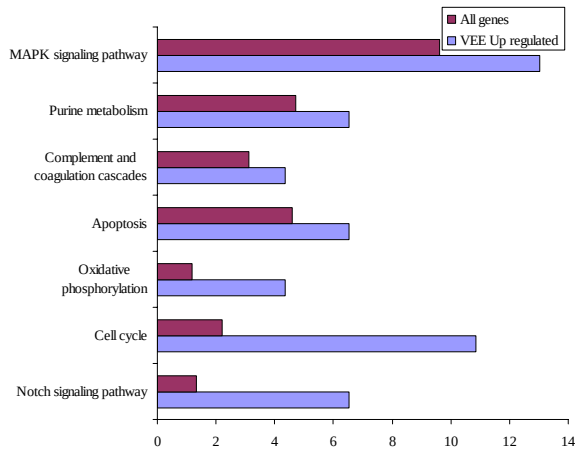


Figure 5. functional analysis of genes that were highly up regulated by VEE in the treated NHPs.

Use of predictive modeling and feature extraction to identify exposures in blinded samples:

To estimate the level of confidence in the methodology, we used our microarray data from the NHP study as a training set. The NHP samples from 12 control and 5 VEE challenged NHP were blinded as “test” datasets.

We applied principal component analysis to examine the behavior of these samples and plotted the PCA components (Figure 6).

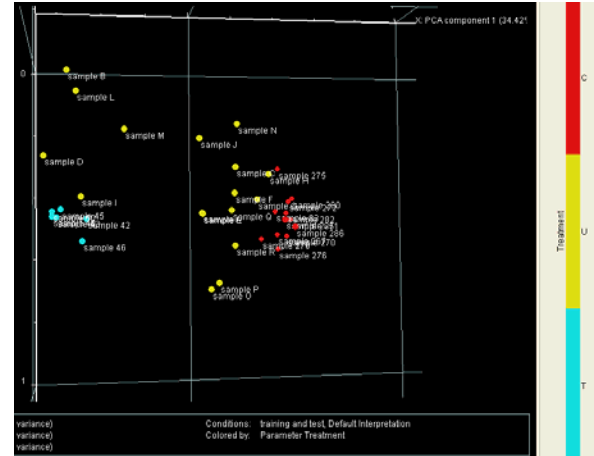


Figure 6. Principal component analysis of known and blinded sets. The blinded set is shown in yellow, the VEE treated samples are shown in blue and the control untreated samples were shown in red.

We used the k-nearest neighbors algorithm with a $p < 0.02$. We identified 50 genes that can be predictors of exposure to VEE.

All of the 17 samples were correctly categorized as to the nature of the exposure. Using both types of training/test methods, one control NHP was categorized correctly as a “control” but had at least some indicators of common viral exposure (Table 1).

Sample	Predicted Value	p value
13256433_sampleA.gpr	C	0.00151
13256434_sampleB.gpr	T	0.0113
13250846_sampleC.gpr	C	0.00151
13250847_sampleD.gpr	T	0.0113
13250848_sampleE.gpr	C	0.208
13250849_sampleF.gpr	C	0.00151
13256857_sampleG.gpr	C	0.00151
13256858_sampleH.gpr	C	0.00151
13256859_sampleI.gpr	T	0.0113
13256860_sampleJ.gpr	C	0.0317
13256862_sampleK.gpr	C	0.00151
13256424_sampleL.gpr	T	0.0113
13256425_sampleM.gpr	T	0.0113
13256426_sampleN.gpr	C	0.00151
13256427_sampleO.gpr	C	0.0317
13256856_sampleP.gpr	C	0.00151
13256855_sampleQ.gpr	C	0.00151
13256854_sampleR.gpr	C	0.00151

Identification of biomarkers that can distinguish VEE from common viruses:

Venezuelan equine encephalitis is mosquito-borne viral disease characterized by early flu-like symptoms. To identify biomarkers that are unique for VEE and can distinguish between exposure to VEE and other common viral agents, we studied host responses to representatives of respiratory and arena viruses to compare the host genome regulation to that of VEE. We carried out temporal analysis of gene expression to the 2 groups of viruses alongwith VEE. Group #1 contained common respiratory viruses (influenza A, parainfluenza, rhinovirus, respiratory syncytical virus). Group #2 focused on 5 arena viruses (the highly virulent hemorrhagic virus, Lassa, and 2 additional pairs of virulent/avirulent arena viruses) (figure 7).

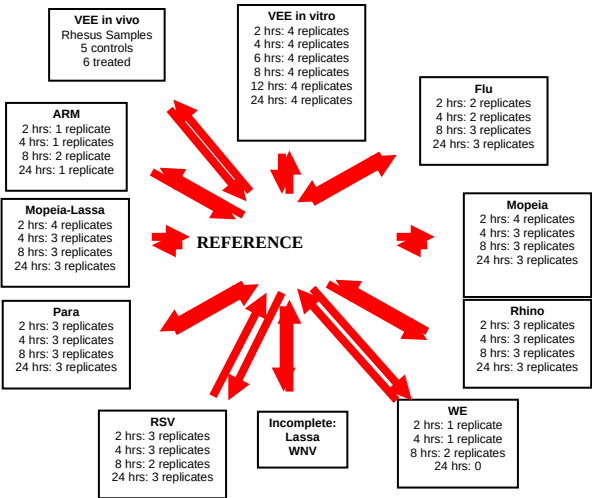


Figure 7. Experimental design of the study carried out to identify biomarkers that would distinguish VEE from other viral agents.

We applied a PCA analysis on the various samples and showed that the host response to VEE have some unique profile when compared to the arena and respiratory viruses (Figure 8).

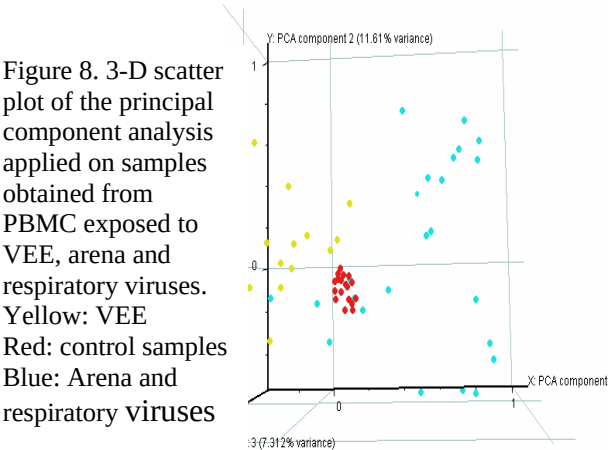


Figure 8. 3-D scatter plot of the principal component analysis applied on samples obtained from PBMC exposed to VEE, arena and respiratory viruses. Yellow: VEE Red: control samples Blue: Arena and respiratory viruses

We applied ANOVA followed by Benjamini correction to identify genes that are differentially expressed between VEE and the other 2 groups. We identified 53 genes that can distinguish exposure to VEE from arena and respiratory viruses (figure 9).

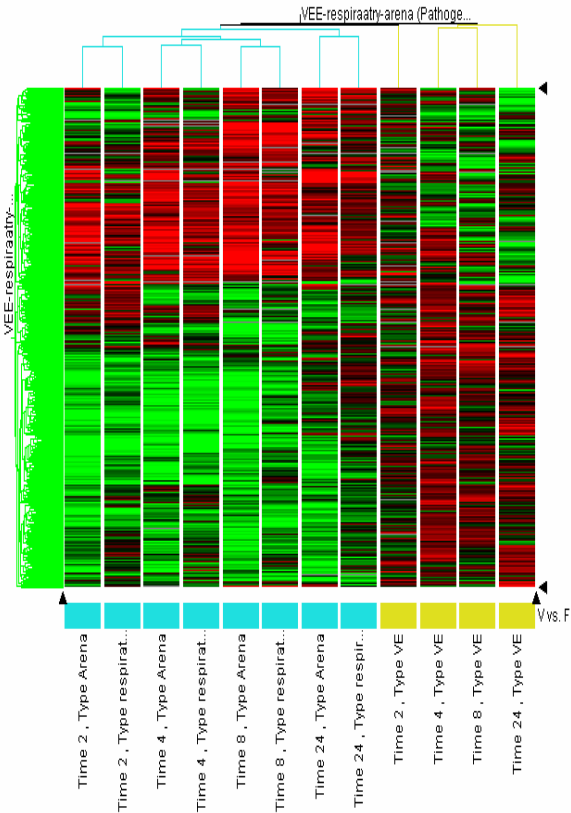


Figure 9. Cluster view of genes that were significantly differentially expressed between VEE (yellow) and arena and respiratory viruses (blue). These genes were obtained using ANOVA with $p < 0.05$ followed by Benjamini correction.

We determined the expression patterns of these genes and showed that they can be used as biomarkers for VEE (figure 10).

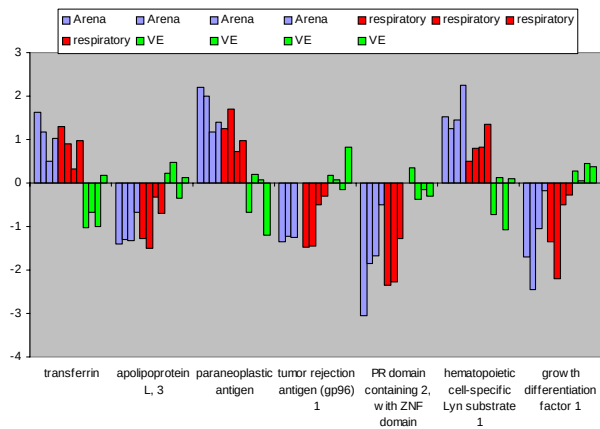
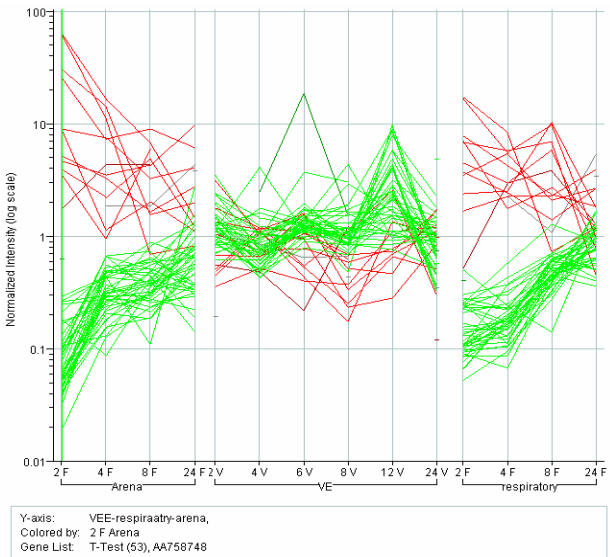


Figure 10. Expression patterns of genes differentially expressed between VEE and Arena and respiratory viruses. A. Genes colored in green are up regulated by VEE and down regulated by the arena and respiratory viruses while the red colored genes are down regulated by VEE and up regulated by the arena and respiratory viruses. B. some the gene are plotted to show patterns of expression in arena viruses (blue), respiratory viruses (red) and VEE (green).

Use of in vitro data to predict the exposure in vivo:

To determine whether these data can predict an exposure to VEE, we used the data from PBMC treated with each of the viruses (VEE, arena and respiratory viruses) as a training set and the data obtained from NHP in vivo VEE

study as a test set. We were able to correctly confirm an exposure to VEE in 15 out of 17 samples (table 2).

Sample	Predicted value
Agent VE , TIme 12 , Donor 1	V
Agent VE , TIme 2 , Donor 1	V
Agent VE , TIme 2 , Donor 3	V
Agent VE , TIme 2 , Donor 4	V
Agent VE , TIme 24 , Donor 1	V
Agent VE , TIme 24 , Donor 2	V
Agent VE , TIme 24 , Donor 3	F
Agent VE , TIme 4 , Donor 1	V
Agent VE , TIme 4 , Donor 2	V
Agent VE , TIme 4 , Donor 3	F
Agent VE , TIme 6 , Donor 2	V
Agent VE , TIme 6 , Donor 3	V
Agent VE , TIme 6 , Donor 4	V
Agent VE , TIme 8 , Donor 1	V
Agent VE , TIme 8 , Donor 2	V
Agent VE , TIme 8 , Donor 3	V
Agent VE , TIme 8 , Donor 4	V

4. CONCLUSION

cDNA microarray studies were effectively used for “training” datasets to address the following questions: 1) Are there common genes modulated under both NHP in vivo and human in vitro conditions in response to VEE exposure? 2) Does this common set of genes permit sufficient discrimination of a test data set based on an in vivo pathogen exposure against a database derived from in vitro exposures to several pathogens especially respiratory and arena viruses? Our successes in correct identification of exposures in NHP to VEE provide a framework for further work in validating and differentiating viral exposures from other pathogenic agents using blood from in NHP or human exposures. Furthermore, new technology for multiplexing genomic responses has decreased the time required for such analyses to nearly “point of care” applications. This study is part of our larger quest to create a library of host gene expression responses that broadly distinguish among common and unconventional pathogenic agents, showing severity of responses via a molecular pathogenesis blueprint that will enable us to design novel methods for intervention. Finding a small number of genes that can accurately classify samples is useful because these predictor genes could prove essential in designing future diagnostic assays for early detection upon exposure to such toxic agents.

5. REFERENCES

Das, R., Cummings, C., Mendis, C., Neill, R., Hoover, D., Lindler, L., Paravitana, C., Huang, X., Ludwig, G., Henchal, E., Yang, D.C.H., Jett, M., 2002. Identification of changes in gene expression induced by toxic agents:

Implications for therapy and rapid diagnosis. In: Wade, J. (Ed), NATO Conference: Operational Issues in Chemical and Biological Defense Human Factors in Medicine Panel. RTO-MP-075-019.

Das, R., Neill, R., Hammamieh, R., Ludwig, G., Jett, M., 2003. Host gene expression profiles in peripheral blood mononuclear cells: Detection of exposure to biological threat agents. ASM Biodefense Research Conference, March 9-12, 2003, Baltimore, MD, USA.

Hammamieh, R., Bi, S., Mani, S., Chakraborty, N., Mendis, C., Das, R., Jett, M., 2003. Genetic variations in peripheral blood mononuclear cells in piglets used as an animal model for staphylococcal enterotoxin exposures. *Omics*. 7(4), 401-409.

Henchal E. A., Teska J. D., Ludwig G. V., Shoemaker D. R and Ezzell J. W, *Clin Lab Med*. **21**, 661-678 (2001).

Jett M. *et al.*, Identification Of Changes In Gene Expression Induced By Toxic Agents: Implications for Therapy and Rapid Diagnosis, J. Wade, Ed., NATO: Portugal (2001).

Jett M. *et al.*, *Infect Immun* **62**, 3408-15 (1994).

Mattix, M.E., Hunt, R.E., Wilhelmsen, C.L., Johnson, A.J., Baze, W.B., 1995. Aerosolized staphylococcal enterotoxin B-induced pulmonary lesions in rhesus monkeys (*Macaca mulatta*). *Toxicol Pathol* 23(3), 262-268.

Genome-wide Survey of Host Responses: Use of Computational Analysis to Classify Exposures and Extract Signatures of Unconventional Versus Common Viral Exposures

Rasha Hammamieh and Marti Jett
Walter Reed Army Institute of Research



WRAIR

Dr. Marti Jett

Dr. Mohsen Barmada

Nabarun Chakraborty

USAMRIID

Dr. George Ludwig

Maj James Koterski



Department of Molecular Pathology

VEE

Dengue 1, 2, 3, 4

WNV

Influenza A

Parainfluenza I

Rhino

RSV

LCMV-WE

LCMV-ARM

Moopia

Moopia-Lassa

Systems Used:

In vitro:

Human peripheral
blood mononuclear
cells exposed to
Biological threat
agents.

In vivo:

Monkeys exposed to
VEE, LCMV-WE and
ARM

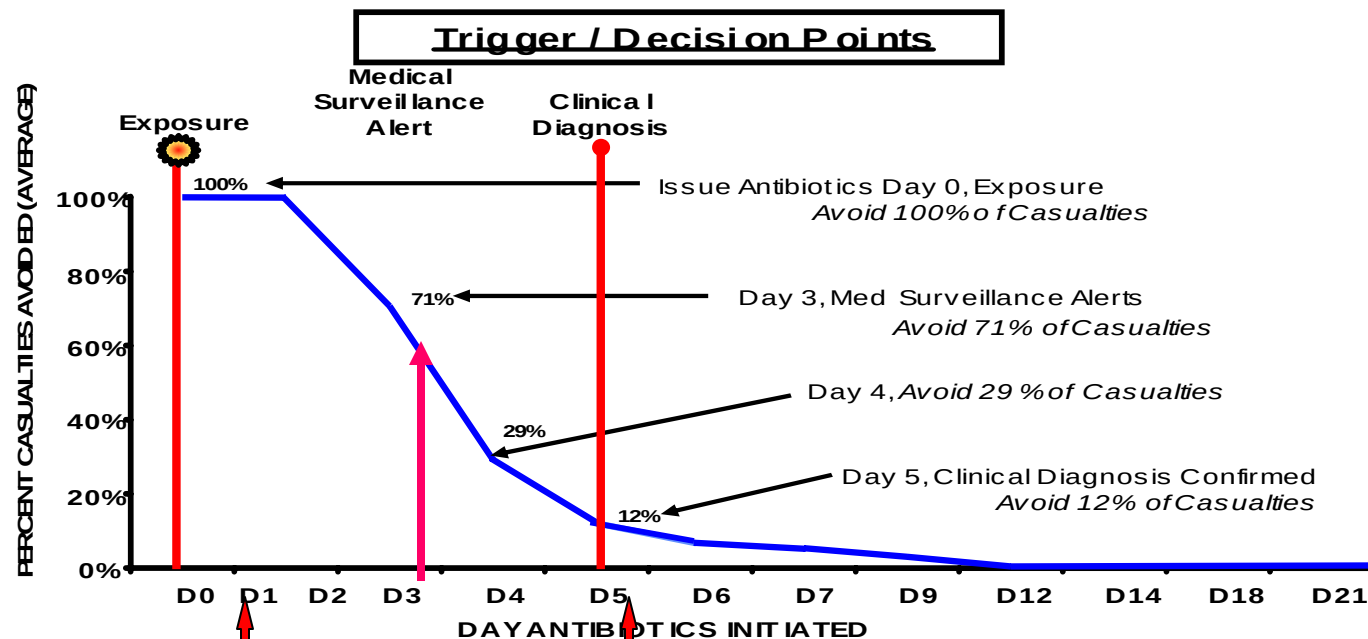
Human Volunteers for
Dengue exposure



- Rapid detection of exposure to biothreat agents is essential so that initiation of treatment can begin promptly.
- Gene profiling provides potential for very early detection strategies (for classic or emerging/modified threat agents).

OTSG Casualty Projections LTC Deborah Schnelle

Projected survival rates vs time treatment is initiated

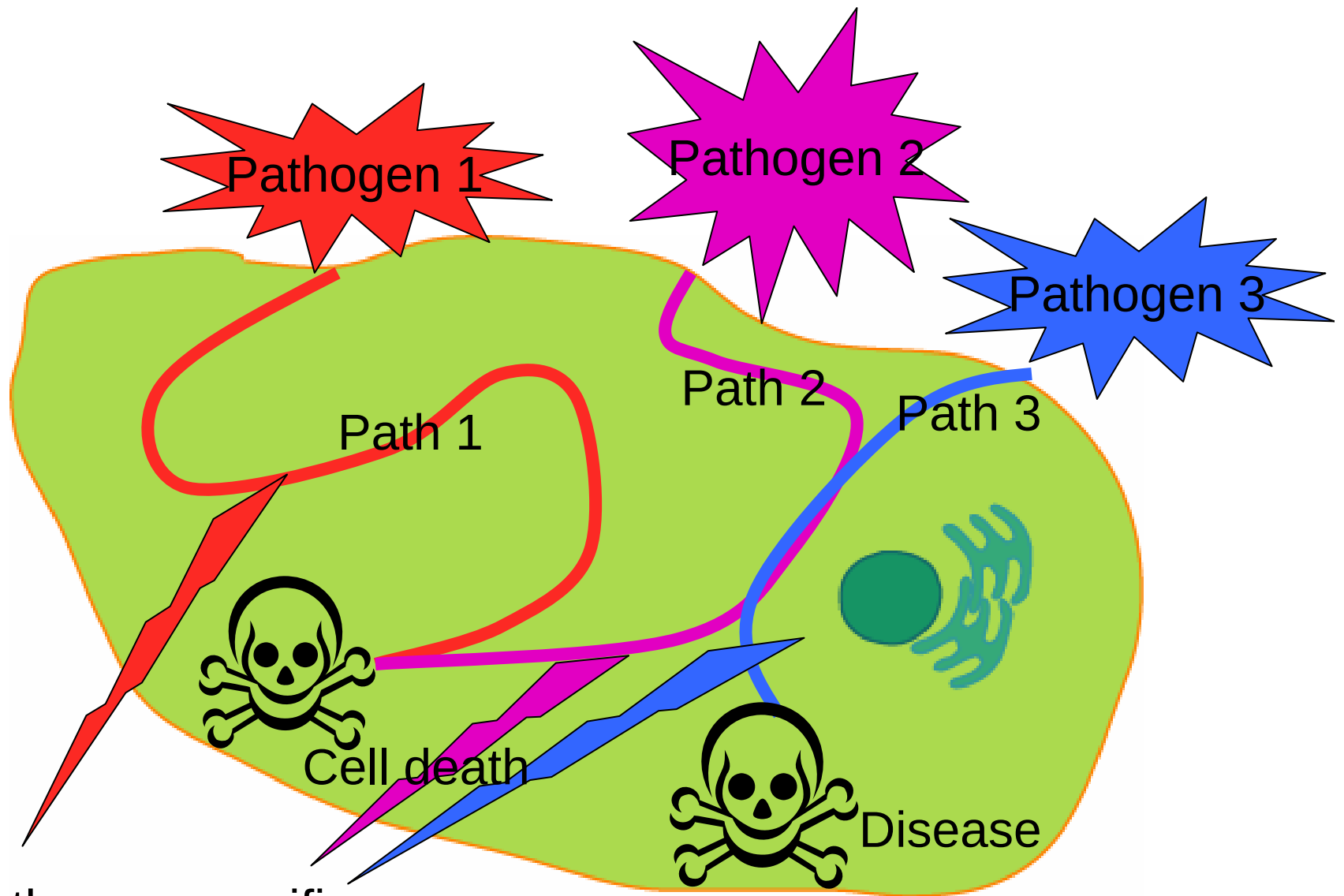


Detection by host
Gene profiling

Pathogen
detectable
in plasma

Clinical
diagnosis



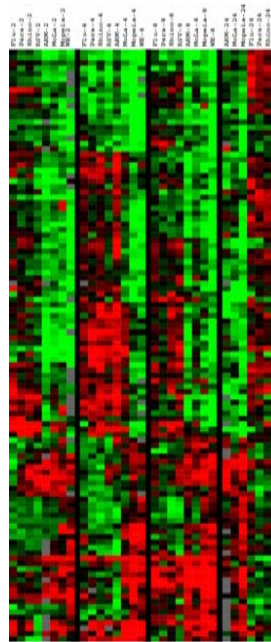
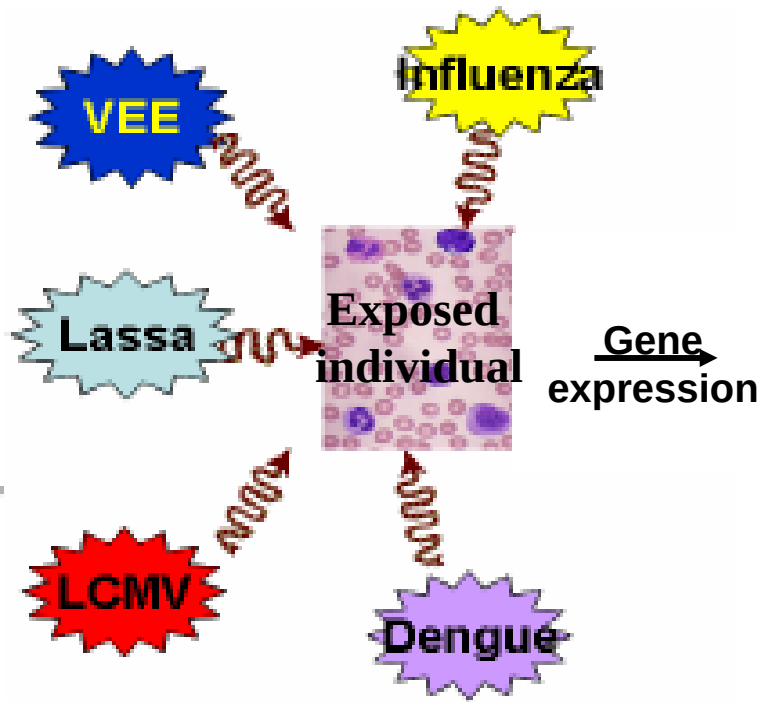


Pathogen-specific
Signal

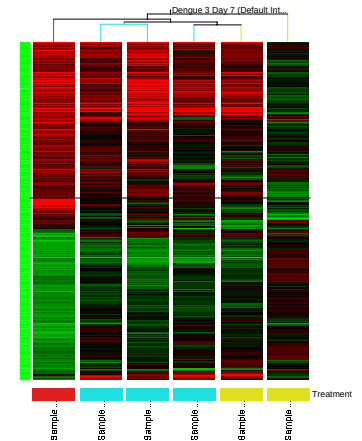
Microarray Technology? Piece of cake!!!



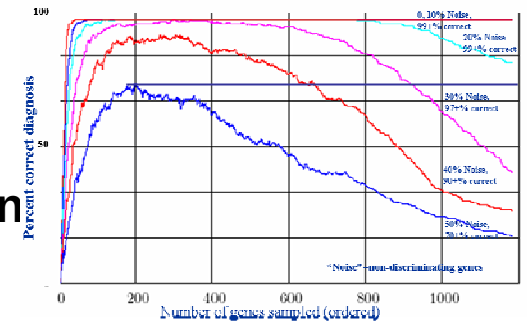
@RAMS



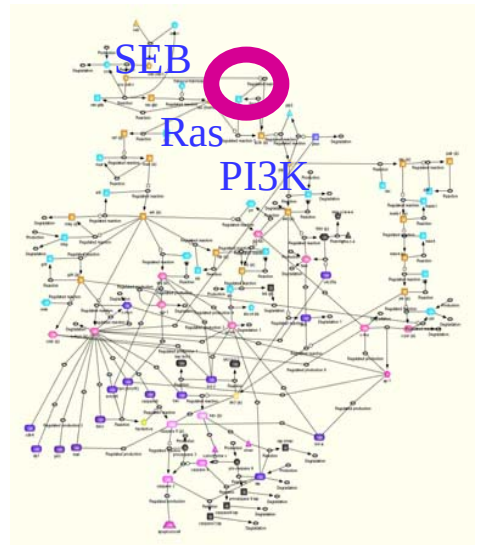
Biomarkers



Feature extraction



Pathway Analysis



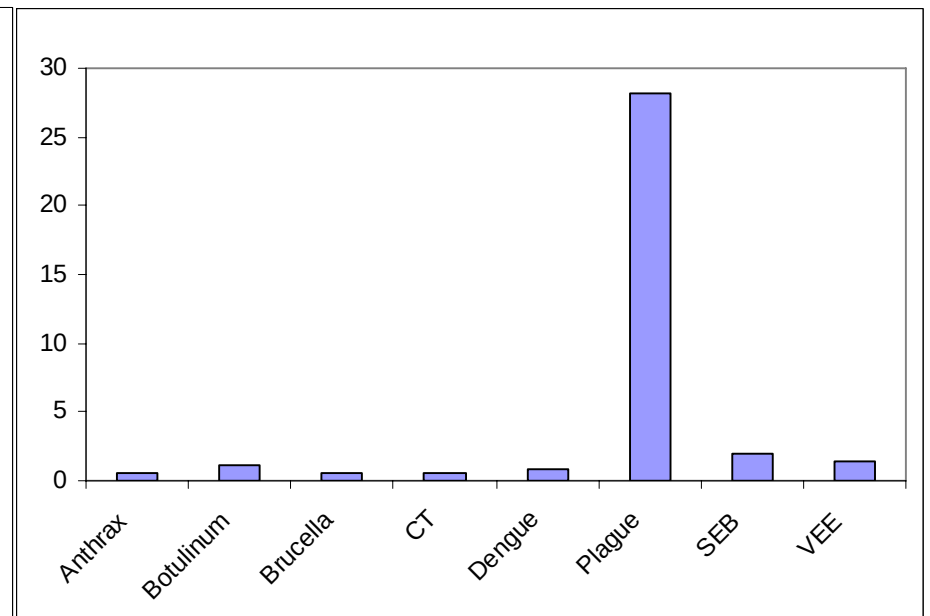
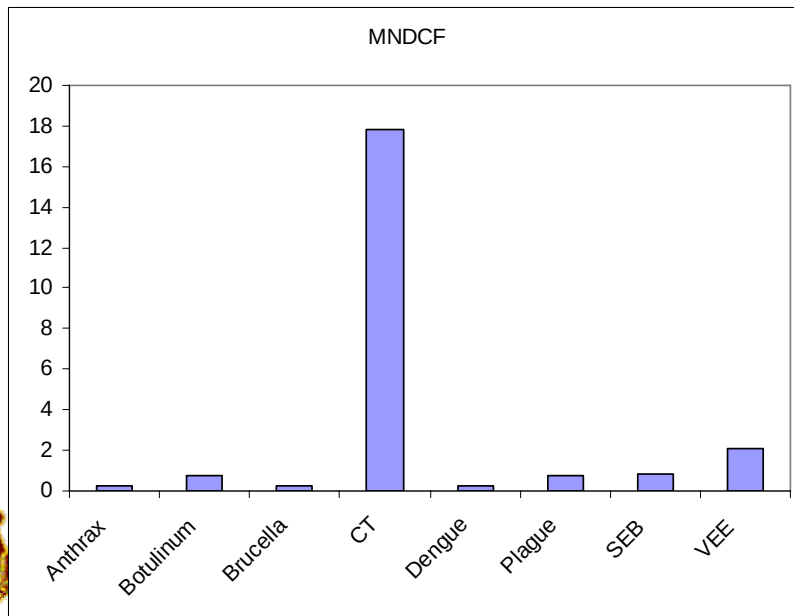
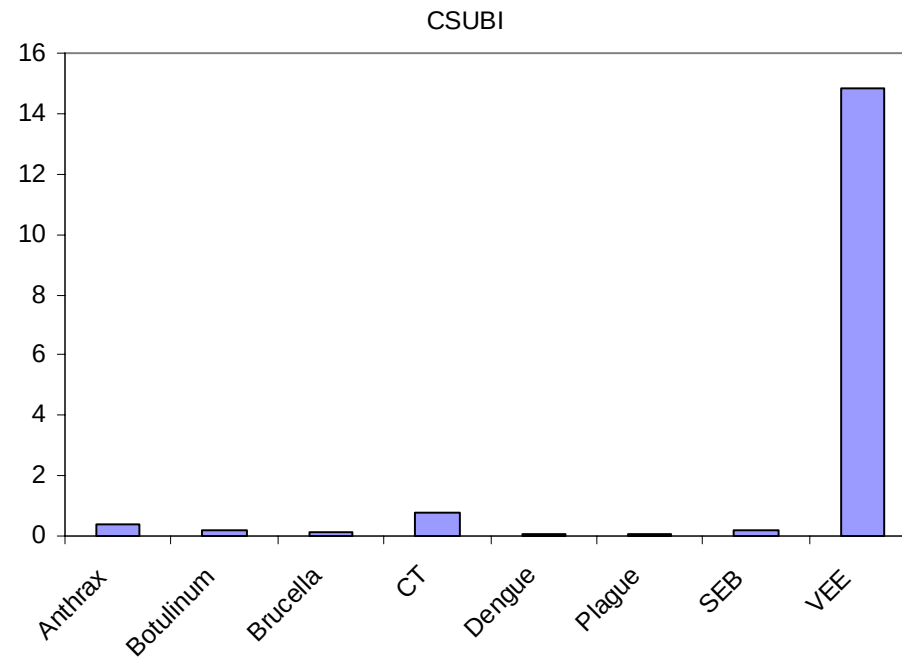
GeneCite/Pathway screen:

High throughput literature and pathway mining

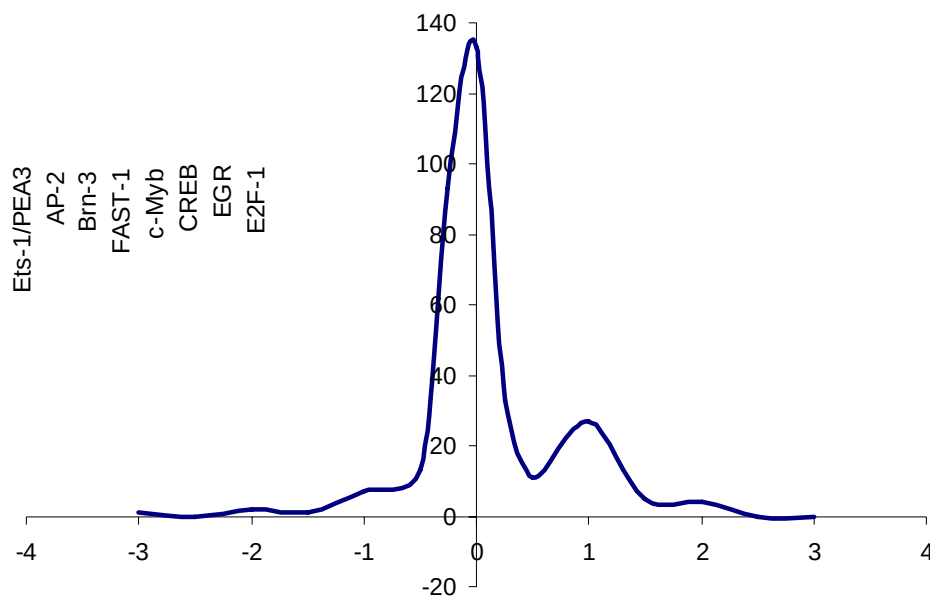
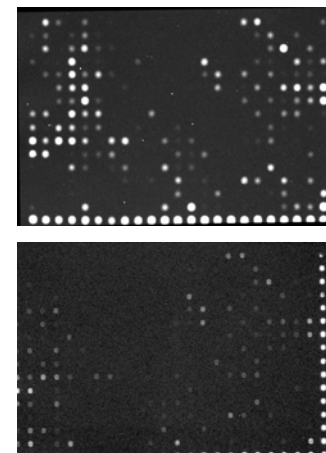
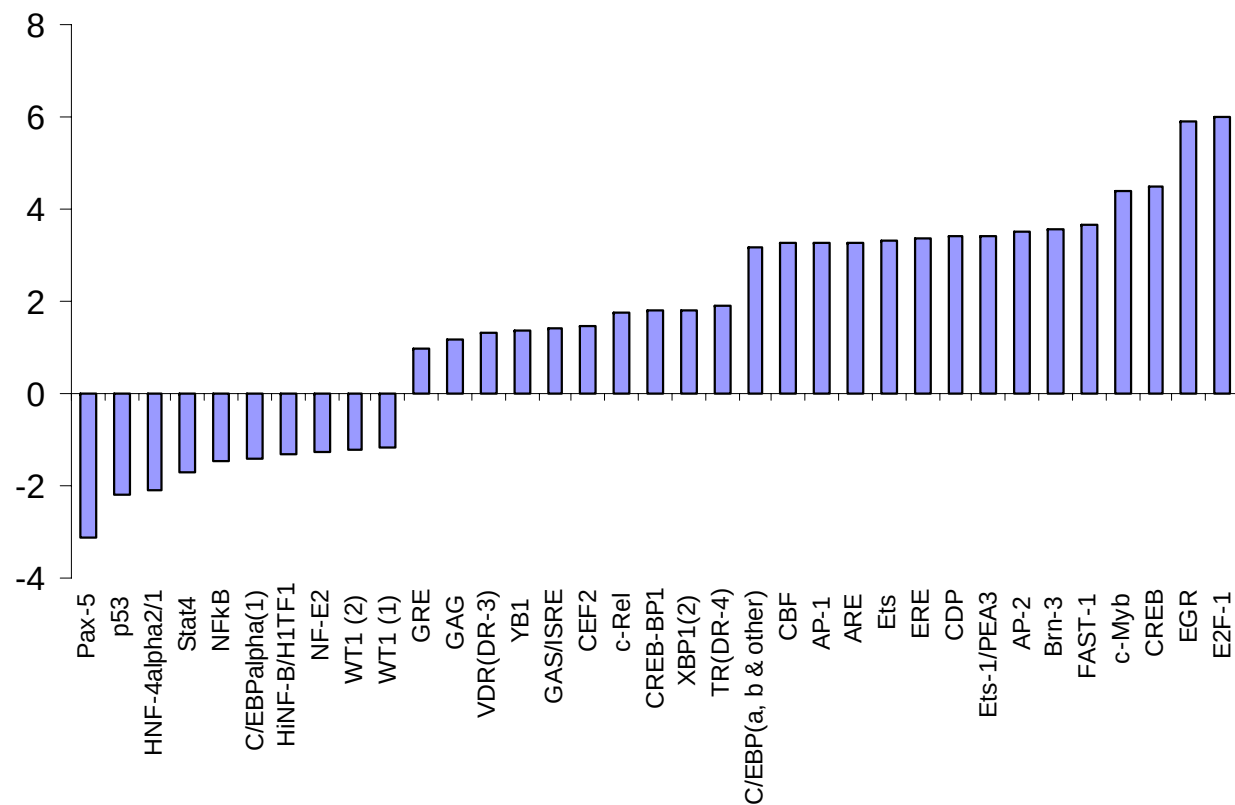
Pathway Name	Genes In Pathway		
Macrophage	singed-lik superoxide dismutase 2, mitochondrial	serine (or cysteine) proteinase inhibitor, clade B, member 1	Nef-associated factor 1
	6624	6648	5055
Cytosolic Pathway	calreticulin		10318
	811	811	
Neutrophil Surface Markers	small inducible cytokine subfamily A (Cys-Cys), member 1		
	6356		
IL-5 Signaling Pathway	small inducible cytokine subfamily A (Cys-Cys), member 1		
	6356		
TNF/Stress Related Signaling	TRAF family member-associated NFKB activator		
	10010		
Eosinophils in Chemokine Attraction	small inducible cytokine subfamily A (Cys-Cys), member 1		
	6356		
TNFR2 Signaling Pathway	TRAF family member-associated NFKB activator		
	10010		
NFAT and Hypertrophy of the Heart	calreticulin		
	811		
Reversal of Insulin Resistance	leptin receptor		
	3953		
Role of Brac1, Brac2 and Atr in the Heart	RAD1 homolog (S. pombe)		
	5810		
CCR3 signaling in Eosinophils	small inducible cytokine subfamily A (Cys-Cys), member 1		
	6356		
Erythropoietin mediated neuroprotection	superoxide dismutase 2, mitochondrial		
	6648		

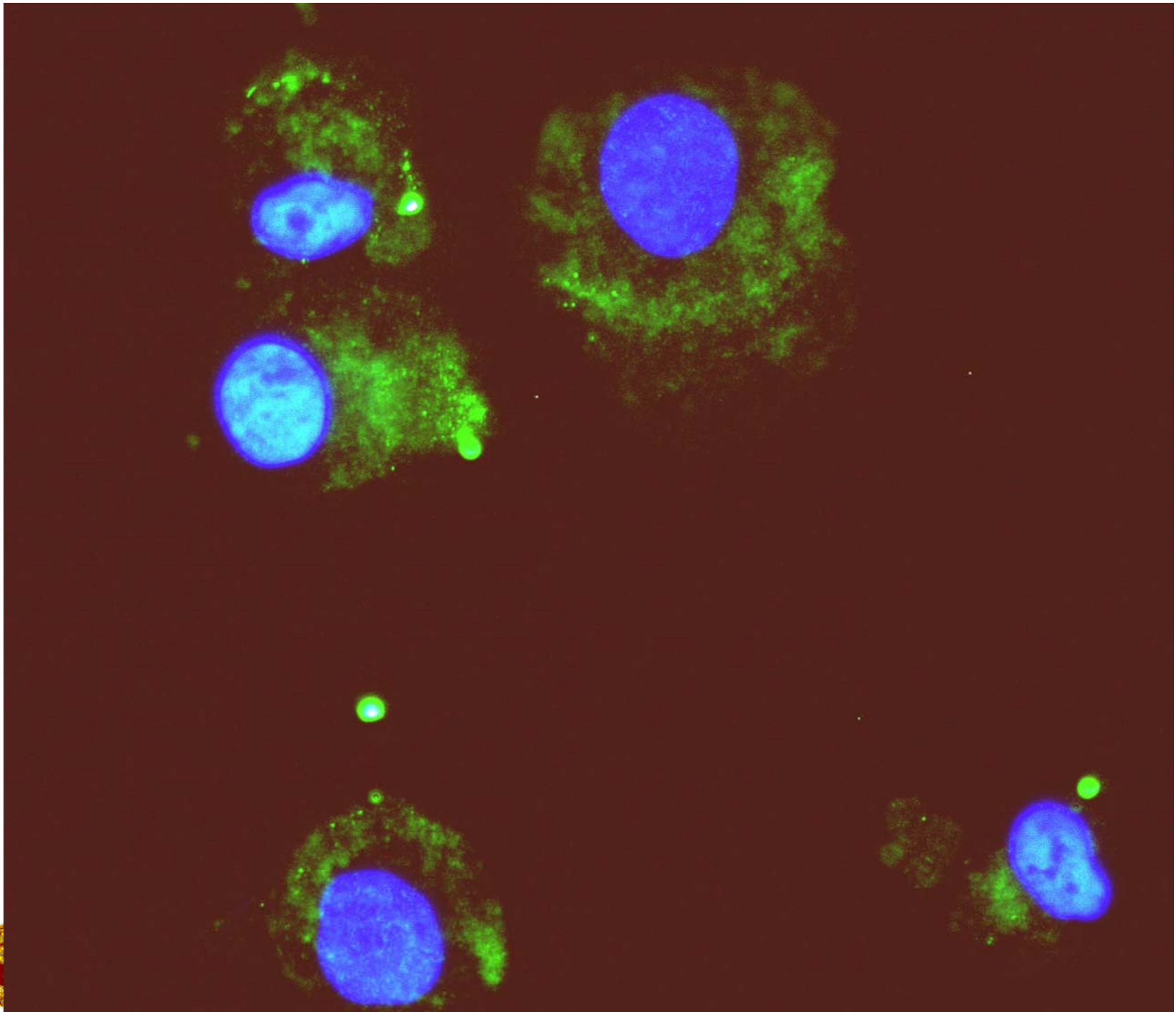


Hammamieh, et al. www.Bioinformatics.org



Transcription Factors binding activity





Genomic Analyses of Host Responses to Venezuelan Equine Encephalitis



- Alphaviruses could theoretically be produced in either wet or dried form and stabilized for weaponization.

- **Incubation Period:** 1-6 days

- **Signs/Symptoms:** VEE may be characterized by sudden onset with malaise, high fever (101-105°F), severe headache, rigors, photophobia, myalgias (especially in the legs and lumbosacral area), cough, sore throat, and vomiting. These symptoms may be followed by a prolonged period of asthenia and lethargy. With natural infection children, and rarely adults, may develop encephalitis.



Diagnosis

- Differential Diagnosis*: Cases with VEE may be difficult to distinguish from those with influenza. Appearance of cases with neurological symptoms may suggest the diagnosis. Other potential bioterrorist agents that cause flu-like illnesses (*Bacillus anthracis*, *Yersinia pestis*, *Coxiella burnetii*) need to be considered.
- Diagnostic Tests*: Virus may be isolated from blood or throat swabs during the acute phase of illness. Blood may be collected for culture in a tiger-top (SST) or red-top tube. Throat swabs may be collected for culture up to 5 days after onset of symptoms. Cerebrospinal fluid (CSF) may also be collected for culture from patients with neurologic symptoms.



Identification of biomarkers for VEE to identify exposures.

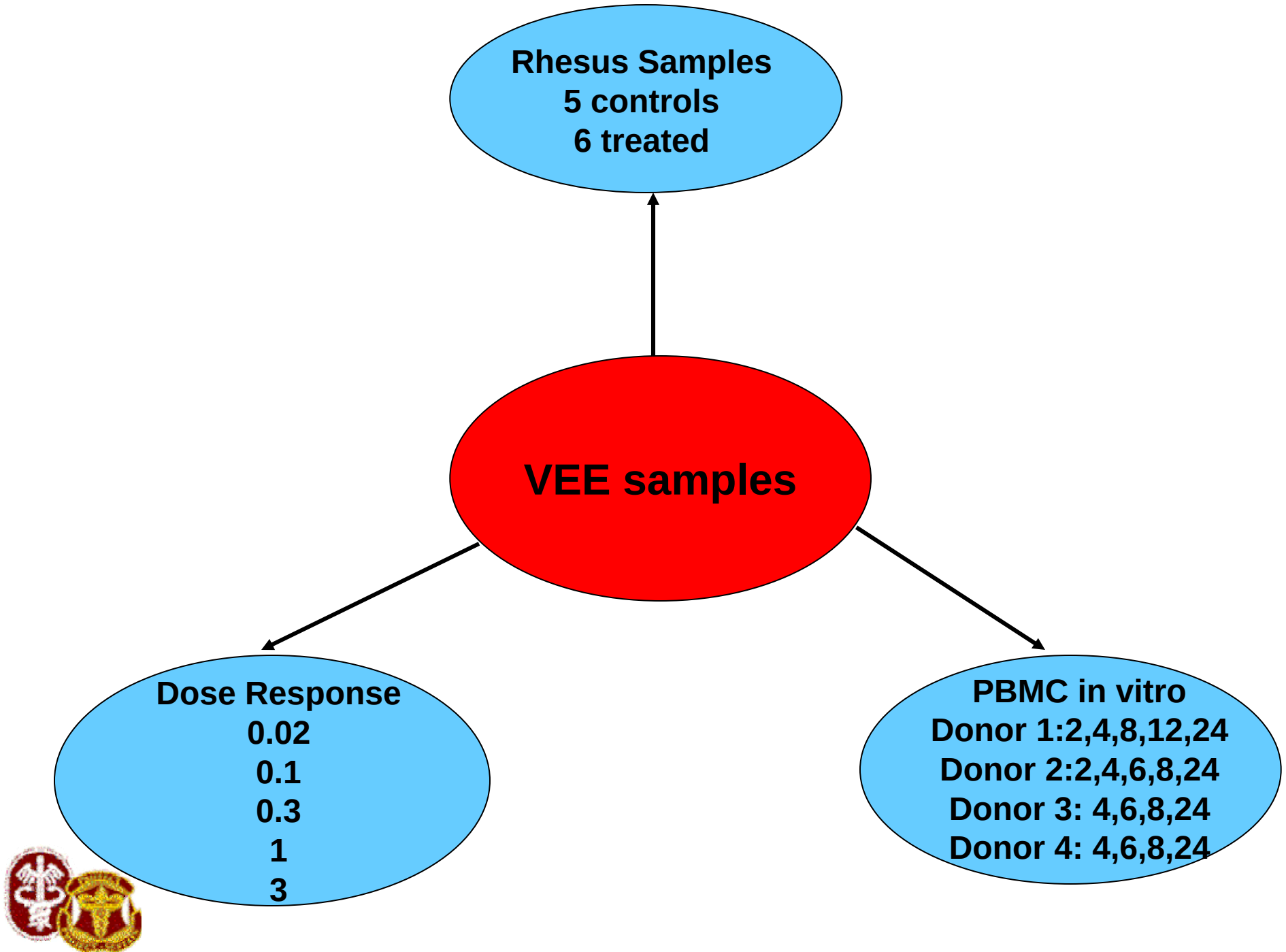
Identification of signatures for VEE compared to common viral exposures.



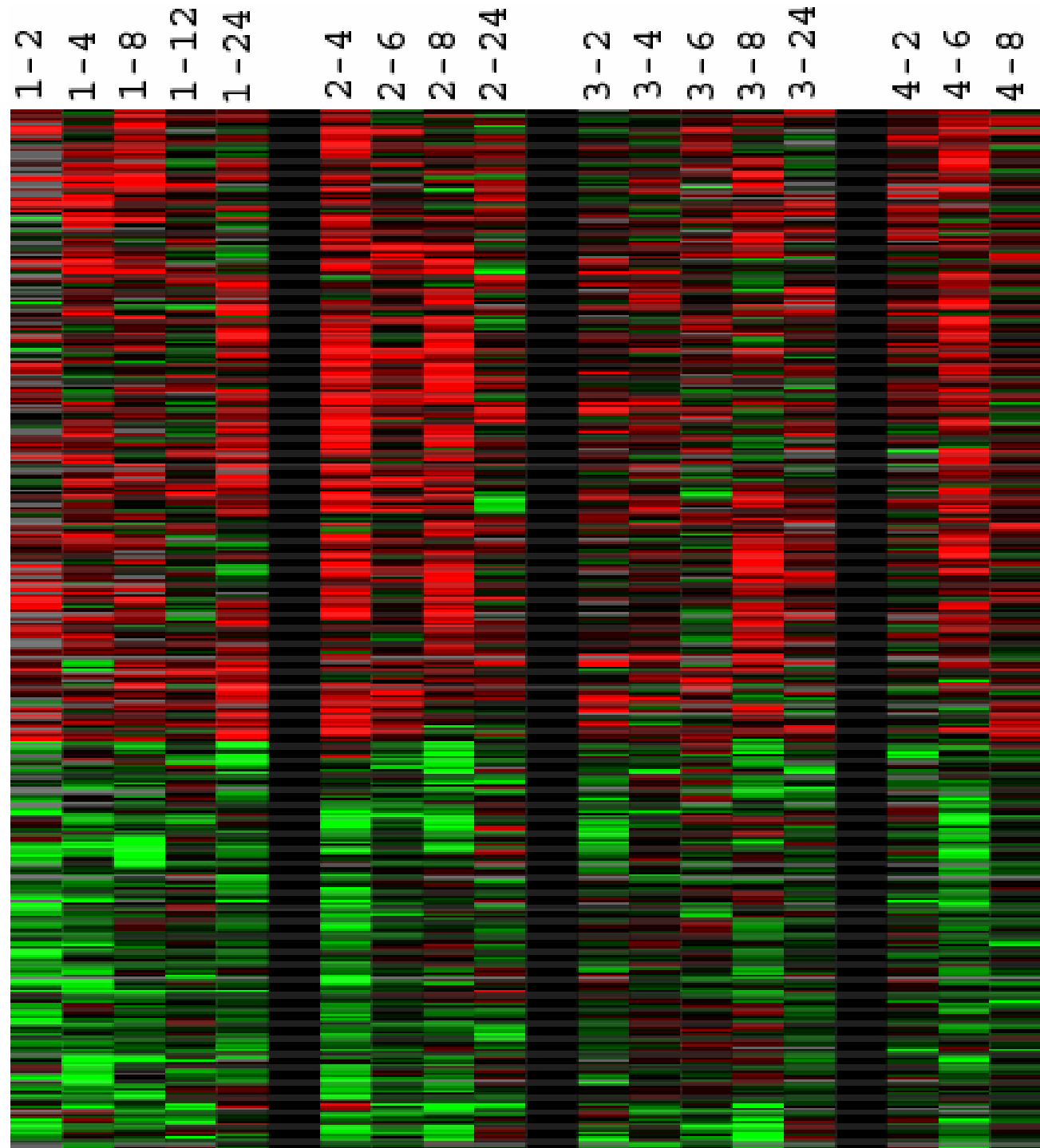
Identification of biomarkers for VEE to identify exposures.

Identification of signatures for VEE compared to common viral exposures.

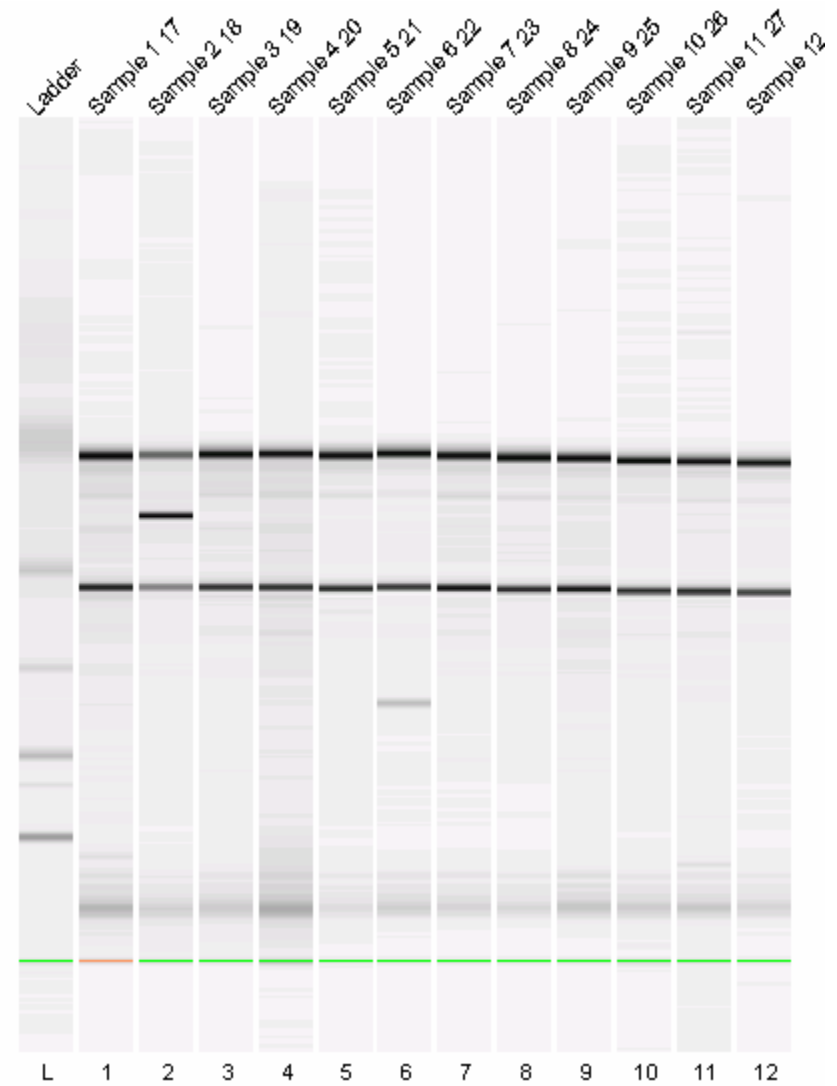
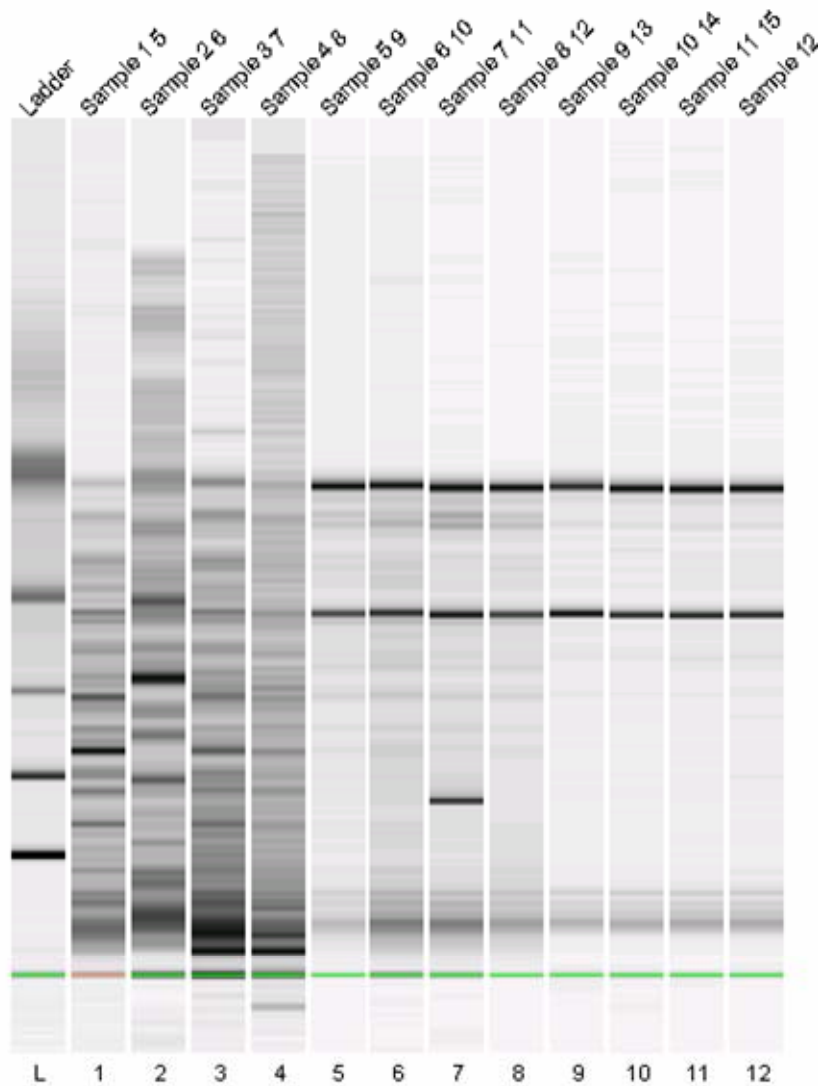




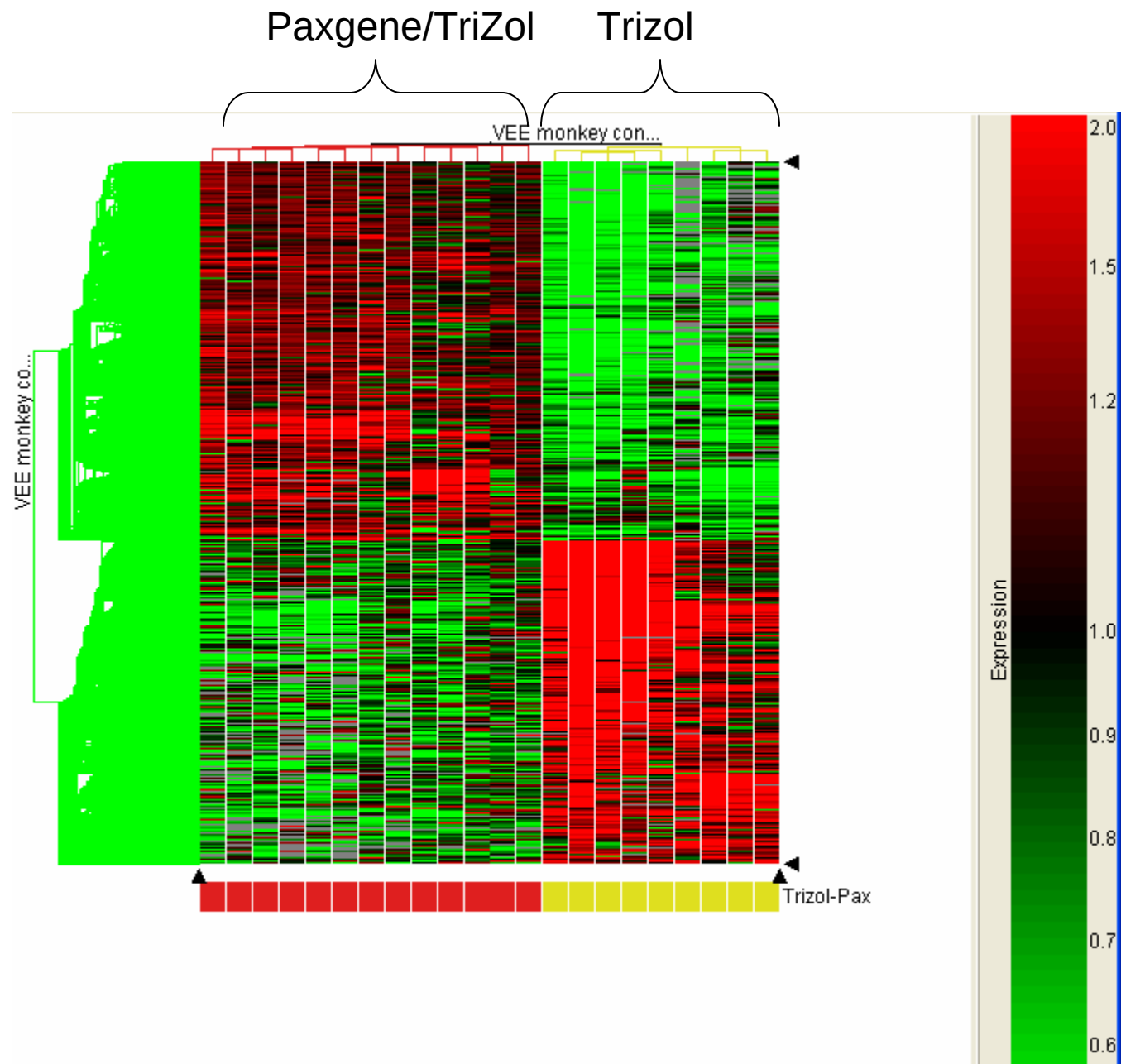
Genes that exhibited highly statistically significant differential expression in PBMC exposed to VEE at various time points when compared to control untreated cells

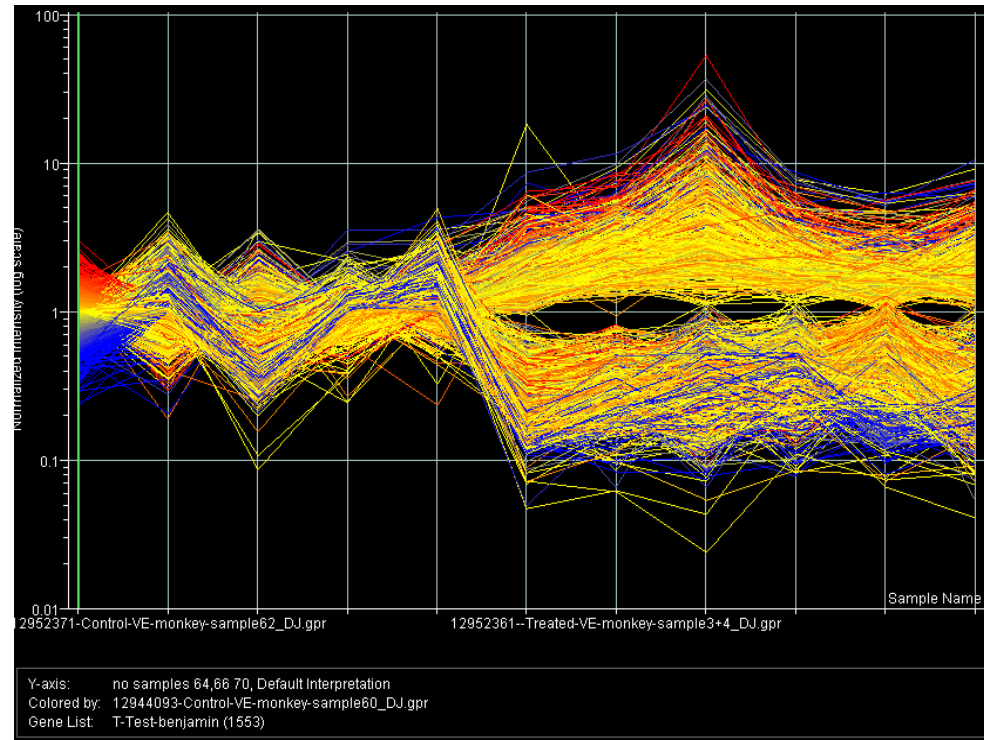
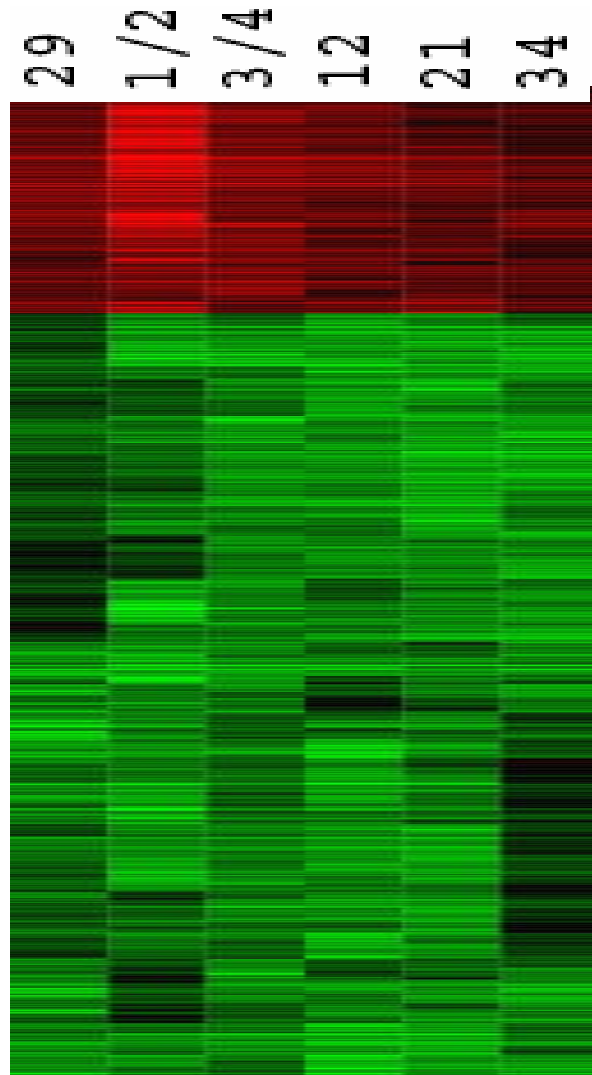


Analysis of RNA samples Isolated using TriZol



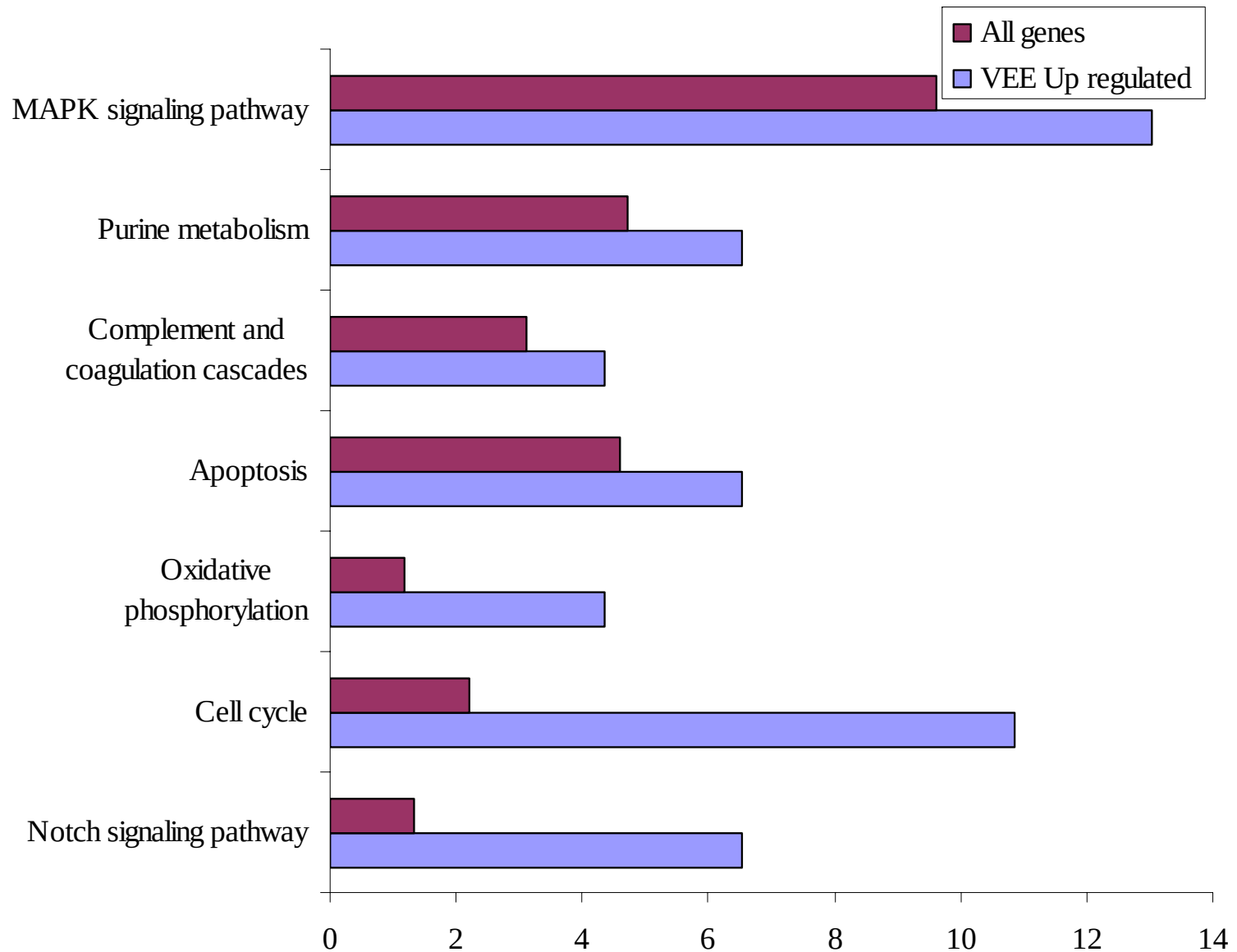
Gene
expression of
monkey PBMC
control RNA
samples
isolated using
TRIZOL and
Paxgene/Trizol



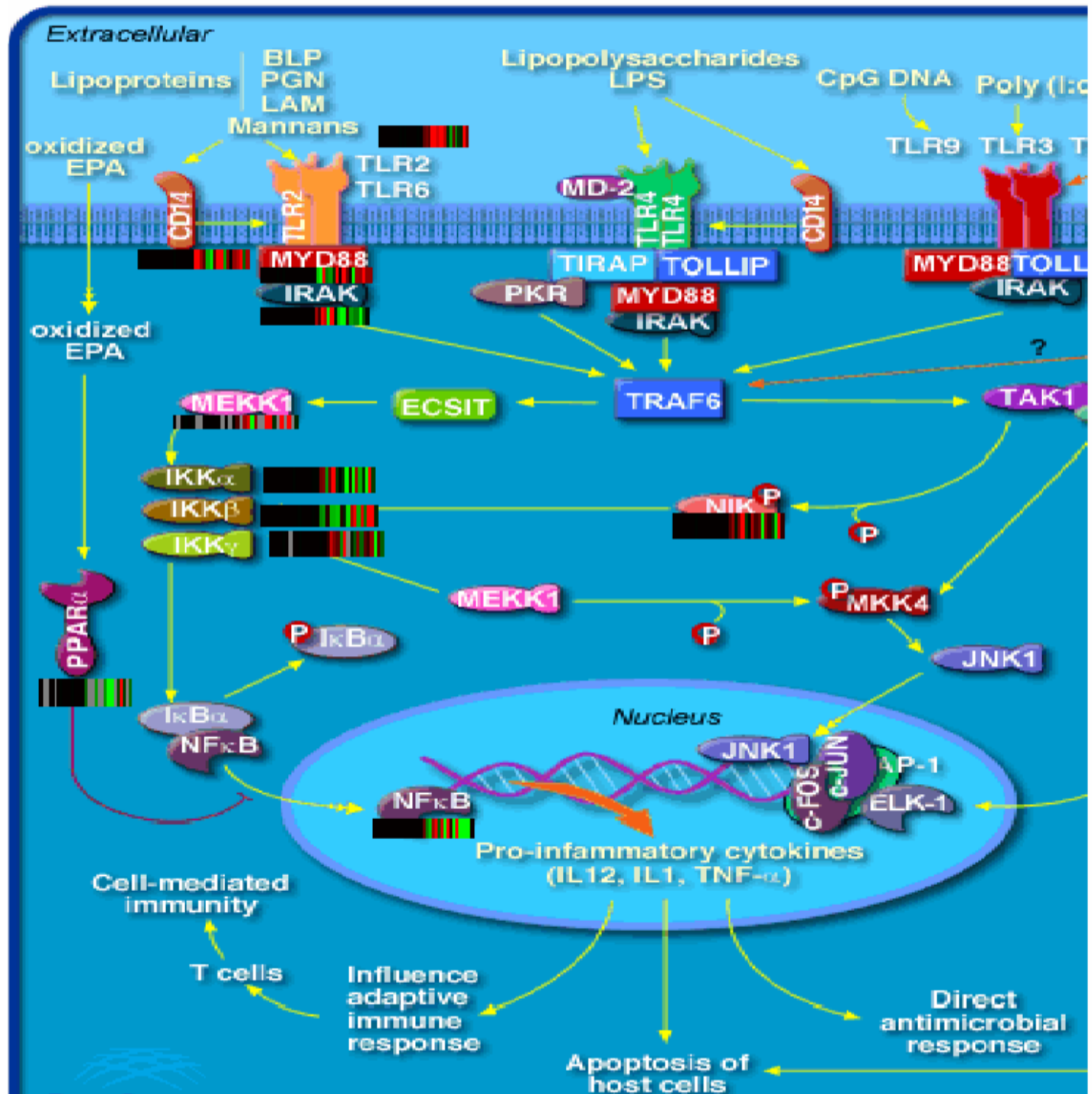


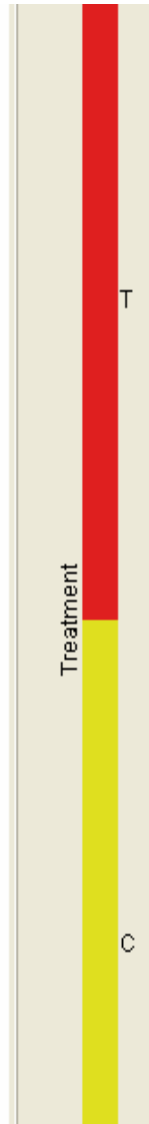
Cluster view and expression profiles of genes that were differentially expressed between NHPs exposed to VEE and control NHPs

Functional analysis of genes that were highly up regulated by VEE in the treated NHPs.



Effect of
VEE on
Toll-like
receptor
pathway

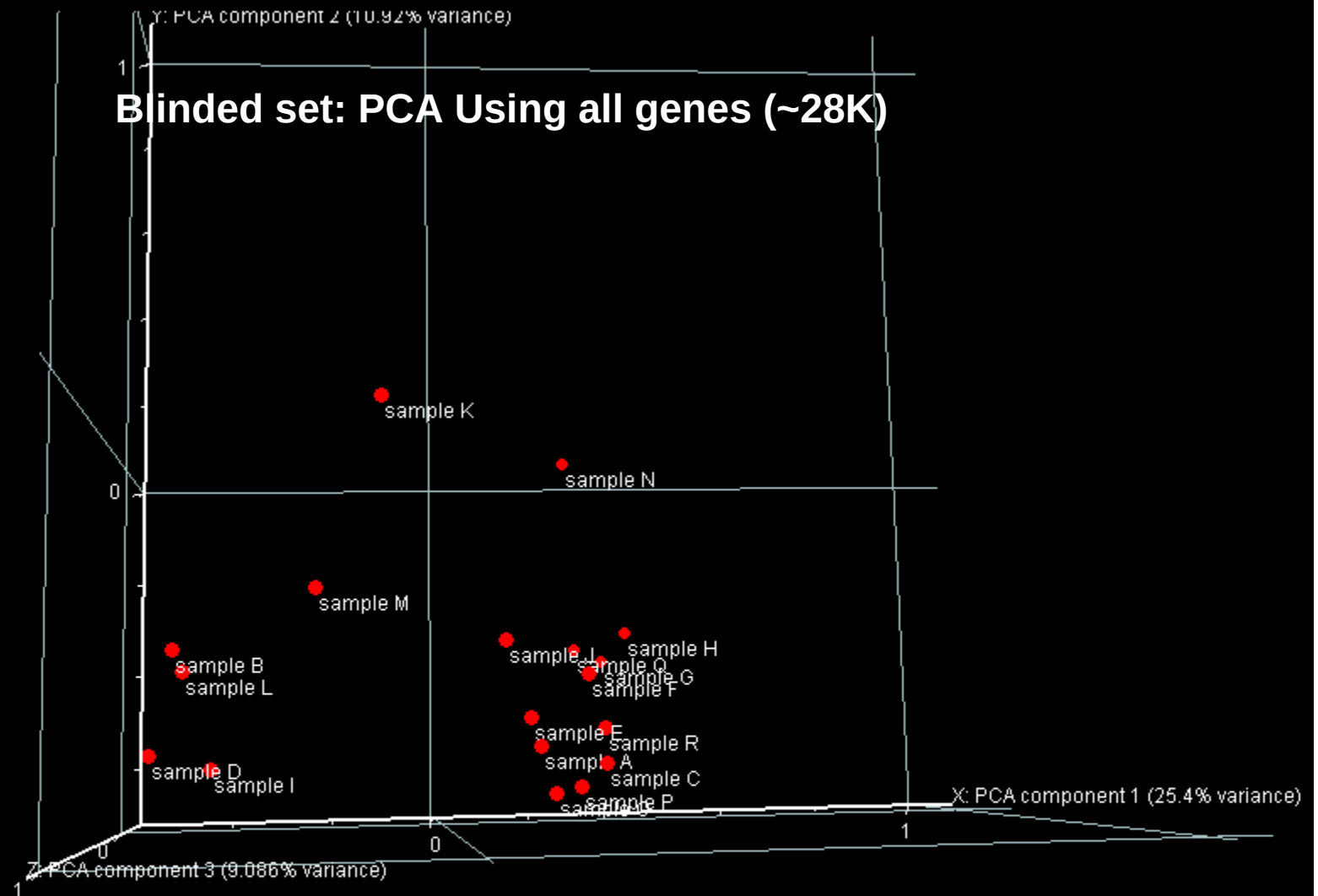




To evaluate our oligo microarray system we were asked to use our data to predict exposures to VEE using a blinded test set. We were given 18 blinded RNA samples from both control and VEE treated NHP. We used known control and VEE samples to train the algorithms. We used our whole genome oligo microarrays (36K) to hybridize the training and test sets. We analyzed the microarray data for both sets. We then applied predictive modeling and feature extraction algorithms to predict and identify exposures to VEE in blinded samples.



Blinded set: PCA Using all genes (~28K)



X-axis: PCA component 1 (25.4% variance)
Y-axis: PCA component 2 (10.92% variance)

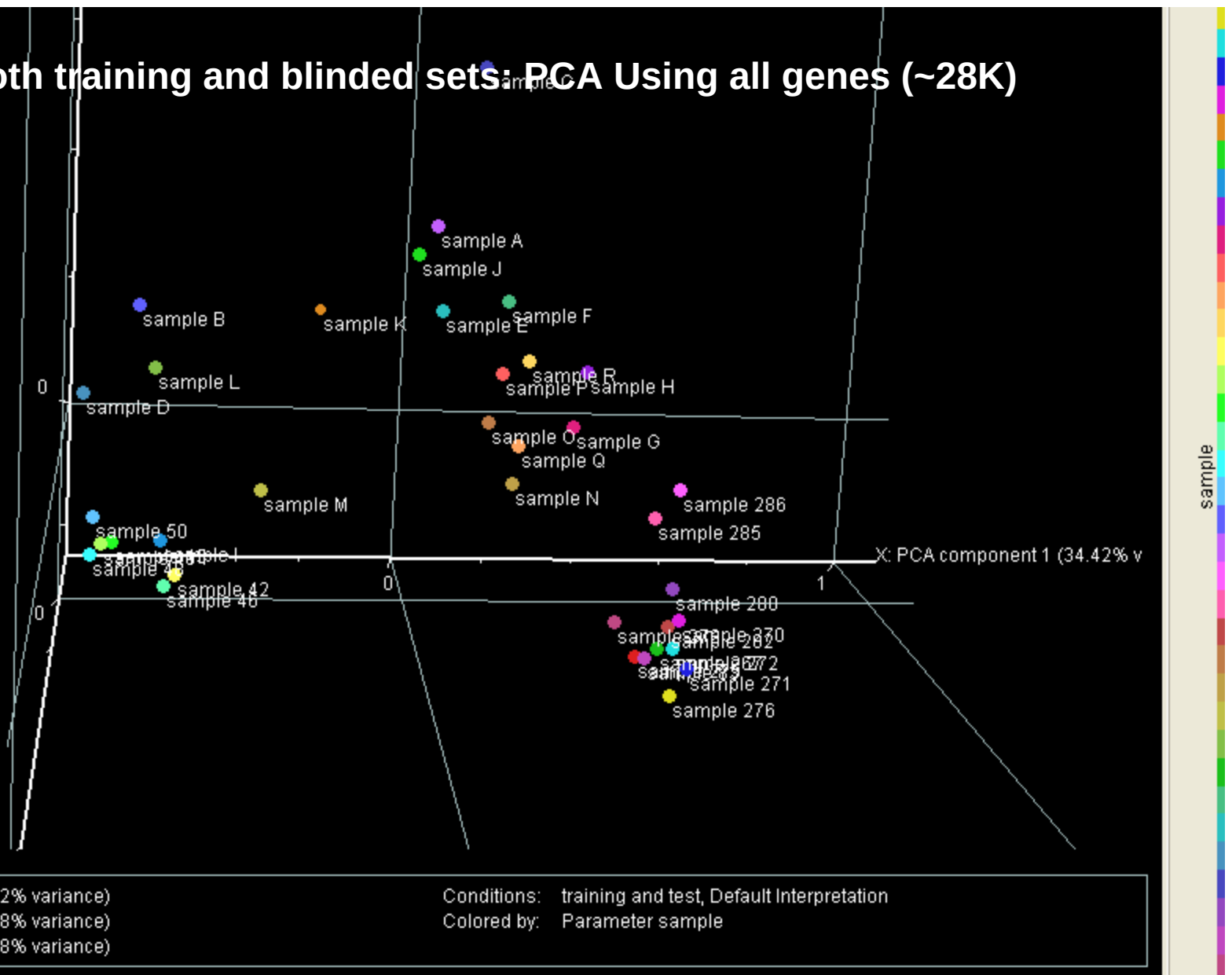
Z-axis: PCA component 3 (9.086% variance)
Conditions: Blinded set, Default Interpretation

Zoom Out Zoom Fully Out Inspect Display Options...

View: 3D Scatter Plot

Close Help

Both training and blinded sets: PCA Using all genes (~28K)

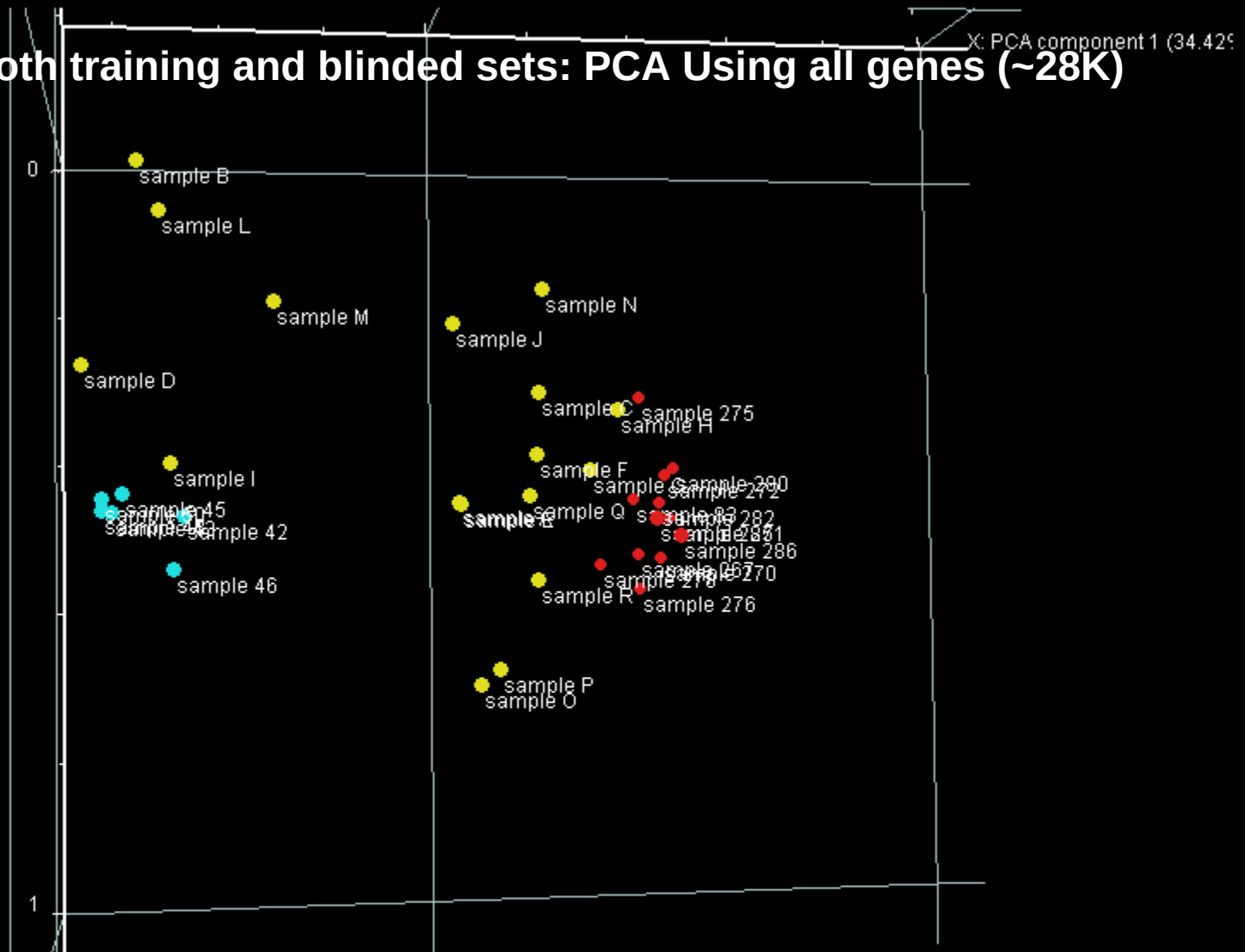


Zoom Out Zoom Fully Out Inspect Display Options...

View: 3D Scatter I

Close Help

Both training and blinded sets: PCA Using all genes (~28K)



axis: PCA component 1 (34.42% variance)
axis: PCA component 2 (9.598% variance)
axis: PCA component 3 (6.638% variance)

Conditions: training and test, Default Interpretation
Colored by: Parameter Treatment

Treatment

Zoom Out Zoom Fully Out Inspect Display Options...

View: 3D Scatter Plot

Close Help

Prediction results

Prediction Results (K-Nearest Neighbors)			
Training Set: Pax-Trizol All Samples (mode: Log) Test Set: Blinded set All Samples (mode: Log) Initial Gene List: all genes (27,746 genes)			
	Condition	Prediction	P value ratio
1	13256433_sampleA.gpr	C	0.00151
2	13256434_sampleB.gpr	T	0.0113
3	13250846_sampleC.gpr	C	0.00151
4	13250847_sampleD.gpr	T	0.0113
5	13250848_sampleE.gpr	C	0.208
6	13250849_sampleF.gpr	C	0.00151
7	13256857_sampleG.gpr	C	0.00151
8	13256858_sampleH.gpr	C	0.00151
9	13256859_sampleI.gpr	T	0.0113
10	13256860_sampleJ.gpr	C	0.0317
11	13256862_sampleK.gpr	C	0.00151
12	13256424_sampleL.gpr	T	0.0113
13	13256425_sampleM.gpr	T	0.0113
14	13256426_sampleN.gpr	C	0.00151
15	13256427_sampleO.gpr	C	0.0317
16	13256856_sampleP.gpr	C	0.00151
17	13256855_sampleQ.gpr	C	0.00151
18	13256854_sampleR.gpr	C	0.00151

Save Predictor Genes

Show Details

Copy to Clipboard

Get Text Description...

Close



Prediction results

Condition	Prediction	P value	Arena vote	C vote	VEE vote	Respiratory vote
Treatment T , ID 495 , Day 14	VEE	0.038	3	0	4	3
Treatment T , ID 573 , Day 14	VEE	0.00241	0	4	6	0
Treatment T , ID 479 , Day 3	VEE	0.00515	0	2	5	3
Treatment T , ID 476 , Day 3	VEE	0.0051	1	1	5	3
Treatment T , ID 494 , Day 4	VEE	0.000448	0	2	6	2
Treatment T , ID 498 , Day 4	VEE	0.111	0	5	5	0
Treatment C , ID 492 , Day C	C	2.70E-08	0	10	0	0
Treatment C , ID 494 , Day C	C	2.70E-08	0	10	0	0
Treatment C , ID 479 , Day C	C	2.70E-08	0	10	0	0
Treatment C , ID 573 , Day C	C	2.70E-08	0	10	0	0
Treatment C , ID 157 , Day C	C	2.70E-08	0	10	0	0
Treatment C , ID 192 , Day C	C	2.70E-08	0	10	0	0
Treatment C , ID 476 , Day C	C	2.70E-08	0	10	0	0
Treatment C , ID 476 , Day C	C	2.70E-08	0	10	0	0
Treatment C , ID 479 , Day C	C	5.31E-05	0	8	1	1
Treatment C , ID 494 , Day C	C	2.70E-08	0	10	0	0
Treatment C , ID 498 , Day C	C	2.70E-08	0	10	0	0
Treatment C , ID 498 , Day C	C	2.70E-08	0	10	0	0
Treatment C , ID 573 , Day C	C	2.70E-08	0	10	0	0



Genomic Analyses of Host Responses to Venezuelan Equine Encephalitis

Identification of biomarkers for VEE to identify exposures
Identification of signatures for VEE compared to common viral exposures



Alphavirus could theoretically be produced in either wet or dried form and stabilized for weaponization.

Incubation Period: 1-6 days

Signs/Symptoms: VEE may be characterized by sudden onset with malaise, high fever (101-105°F), severe headache, rigors, photophobia, myalgias (especially in the legs and lumbosacral area), cough, sore throat, and vomiting. These symptoms may be followed by a prolonged period of asthenia and lethargy. With natural infection children, and rarely adults, may develop encephalitis.

Diagnosis:

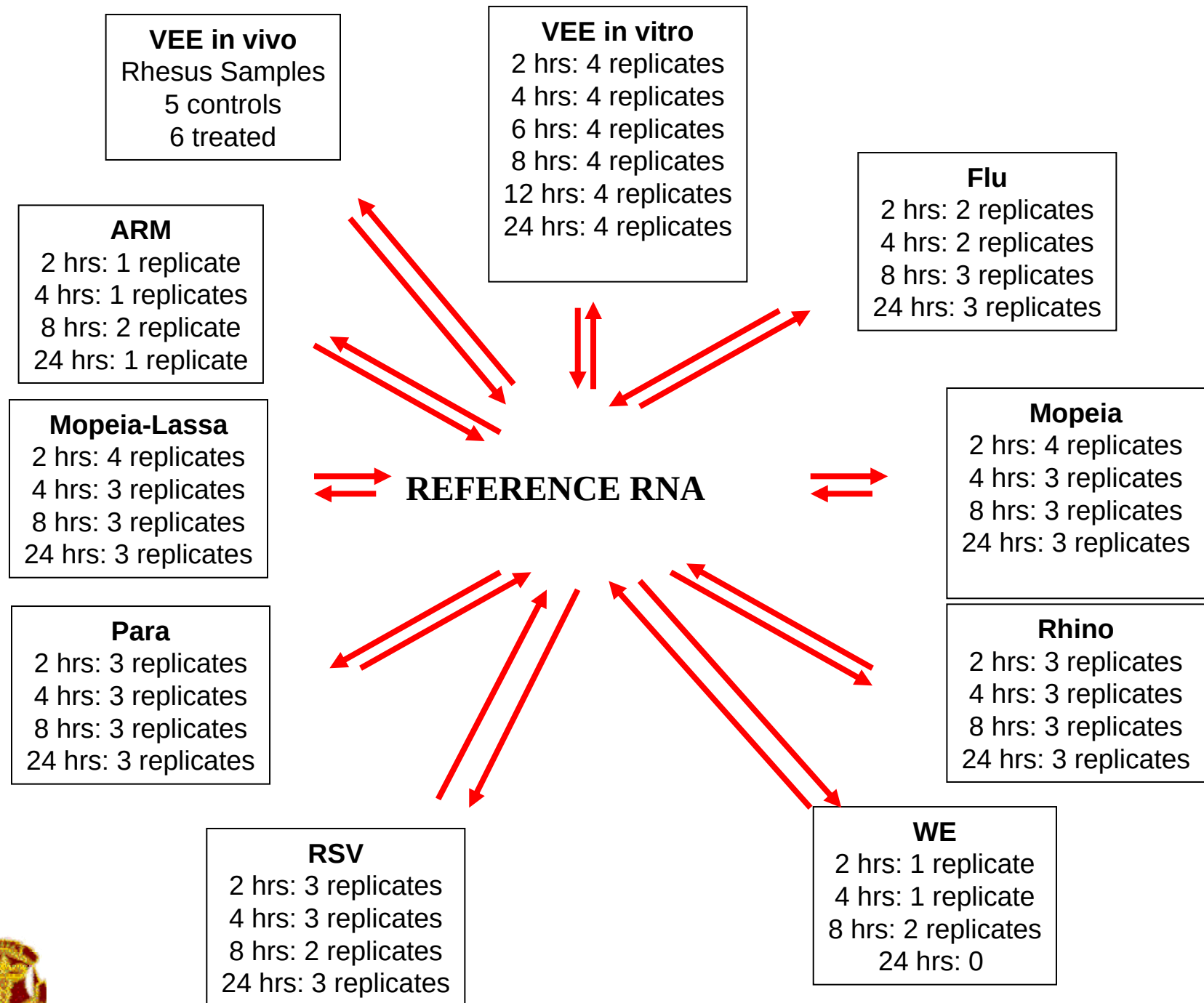
Differential Diagnosis: Cases with VEE may be difficult to distinguish from those with influenza. Appearance of cases with neurological symptoms may suggest the diagnosis.

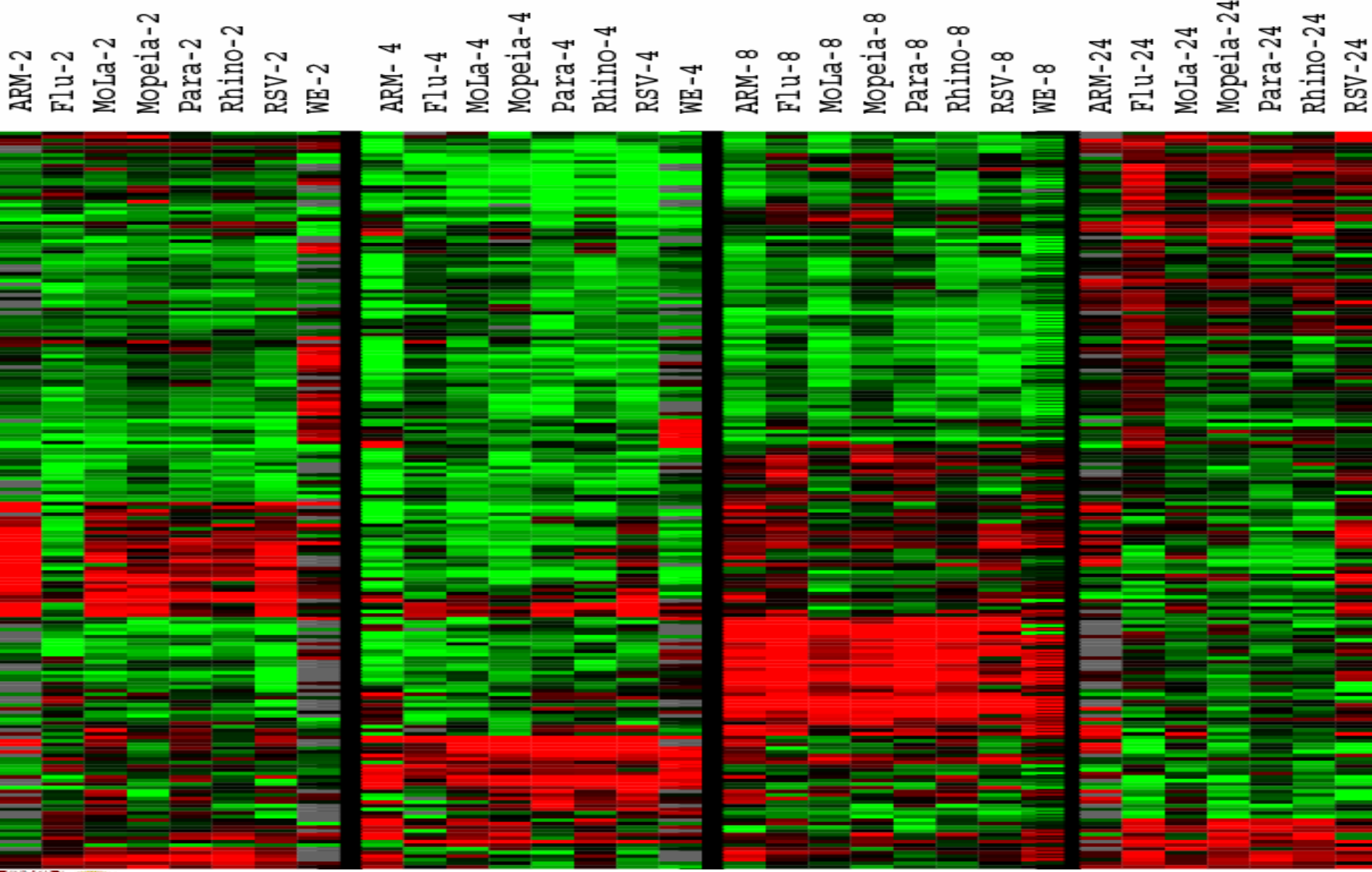
Other potential bioterrorist agents that cause flu-like illnesses (*Bacillus anthracis*, *Yersinia pestis*, *Coxiella burnetii*) need to be considered.

Diagnostic Tests: Virus may be isolated from blood or throat swabs during the acute phase of illness. Blood may be collected for culture in a tiger-top (SST) or red-top tube. Throat swabs may be collected for culture up to 5 days after onset of symptoms. Cerebrospinal fluid (CSF) may also be collected for culture from patients with neurologic symptoms.

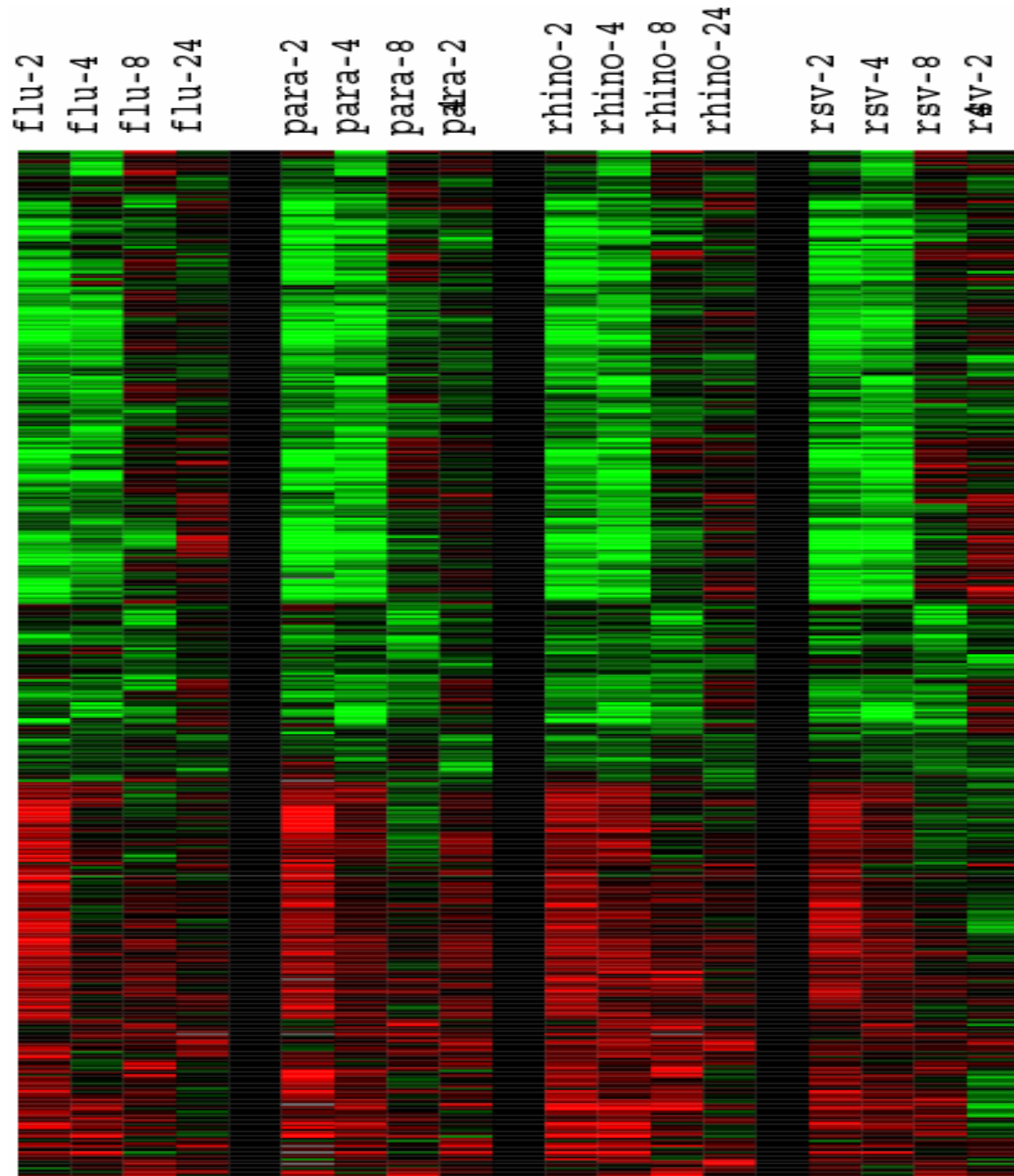
Infection Control: Standard blood precautions and barrier-nursing techniques must be employed. Patients should be isolated. The virus is destroyed by heat (80°C for 30 minutes) and regular disinfectants.

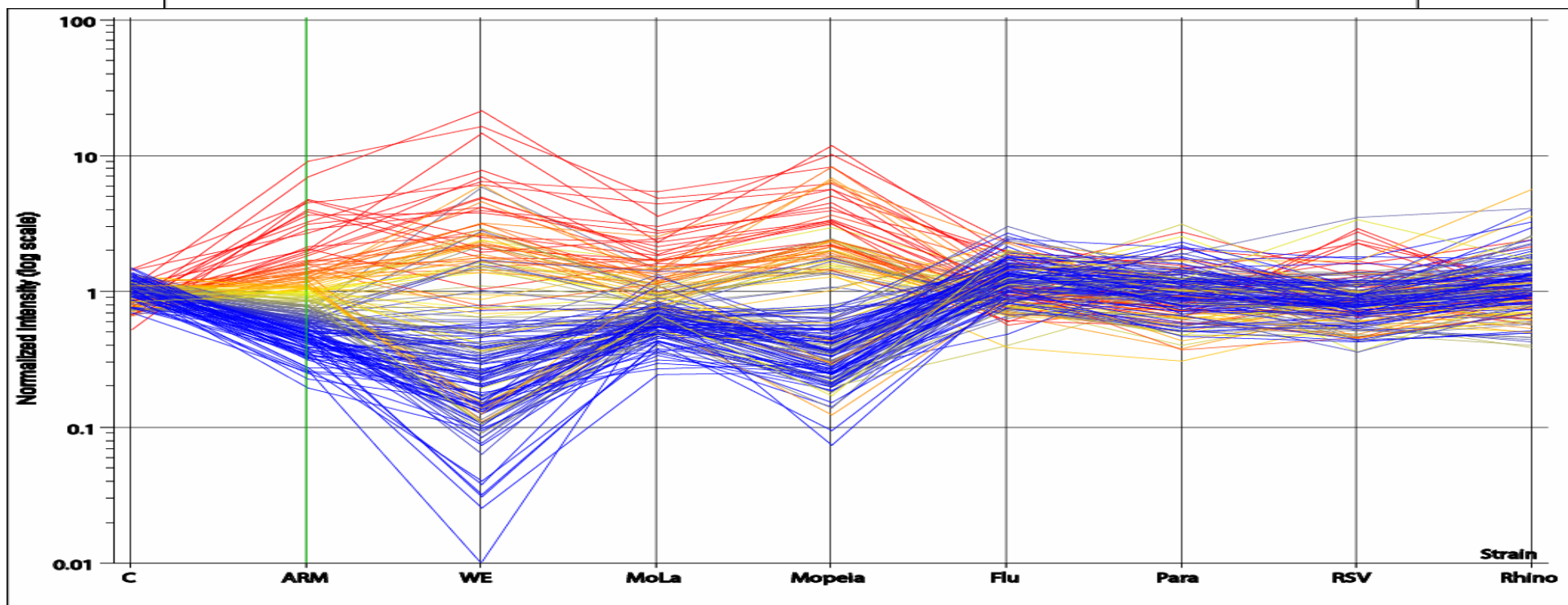
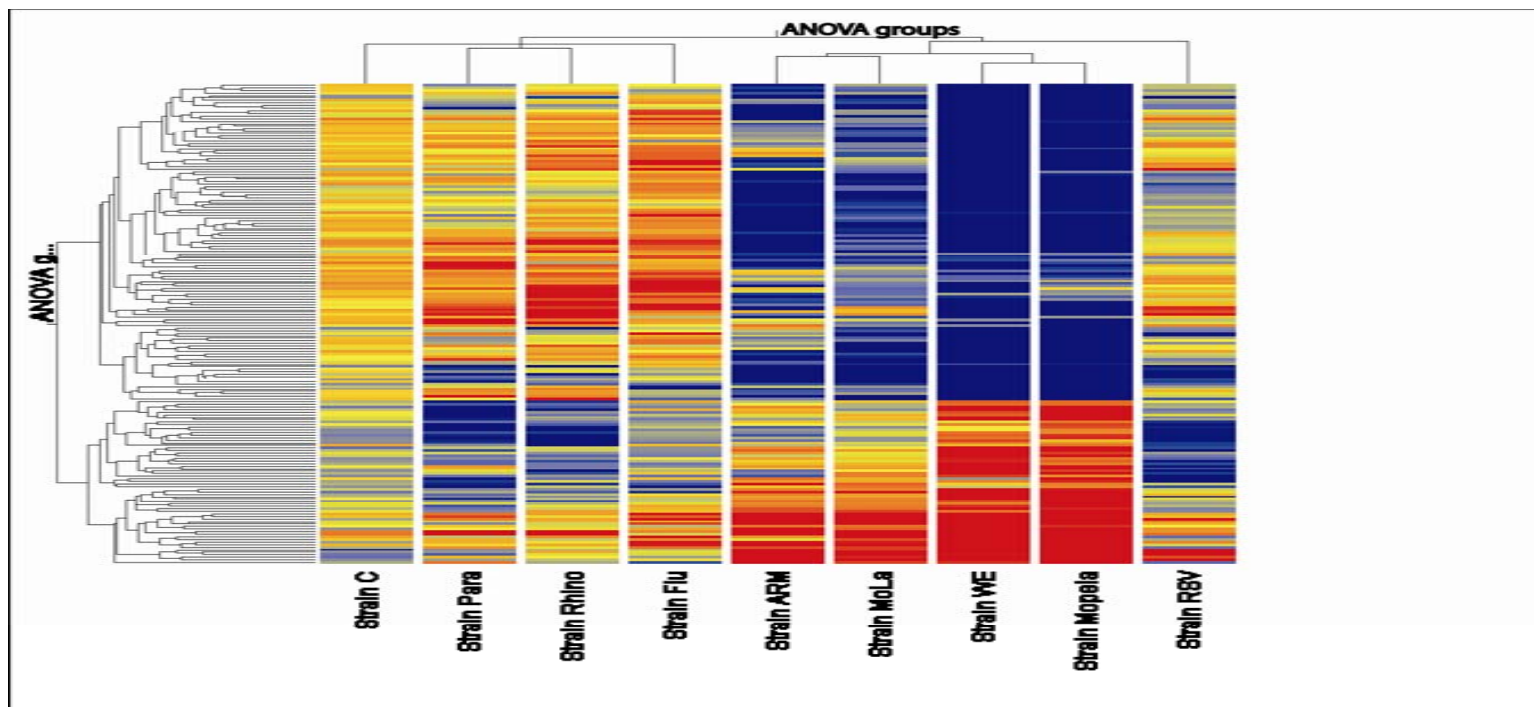


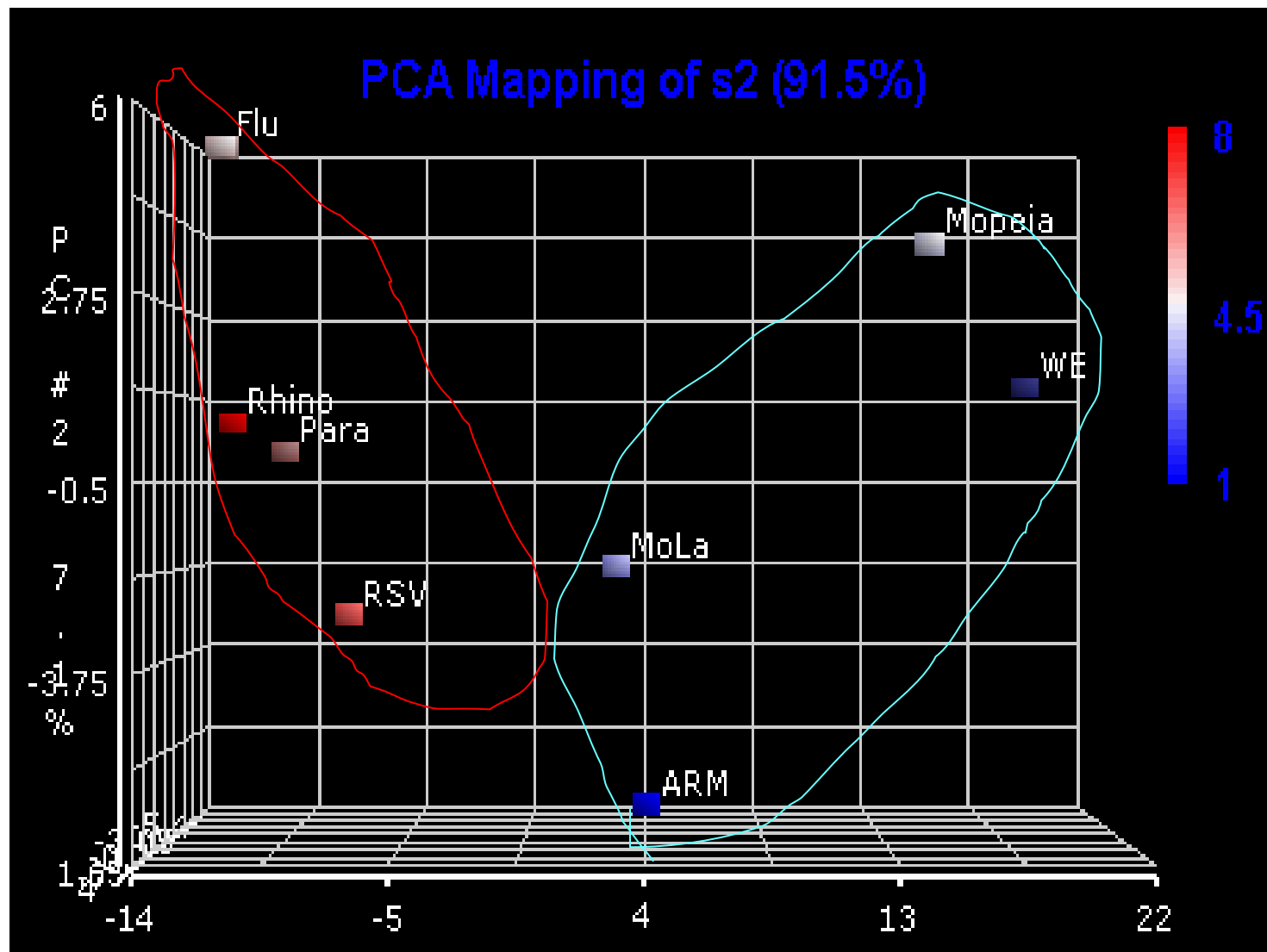


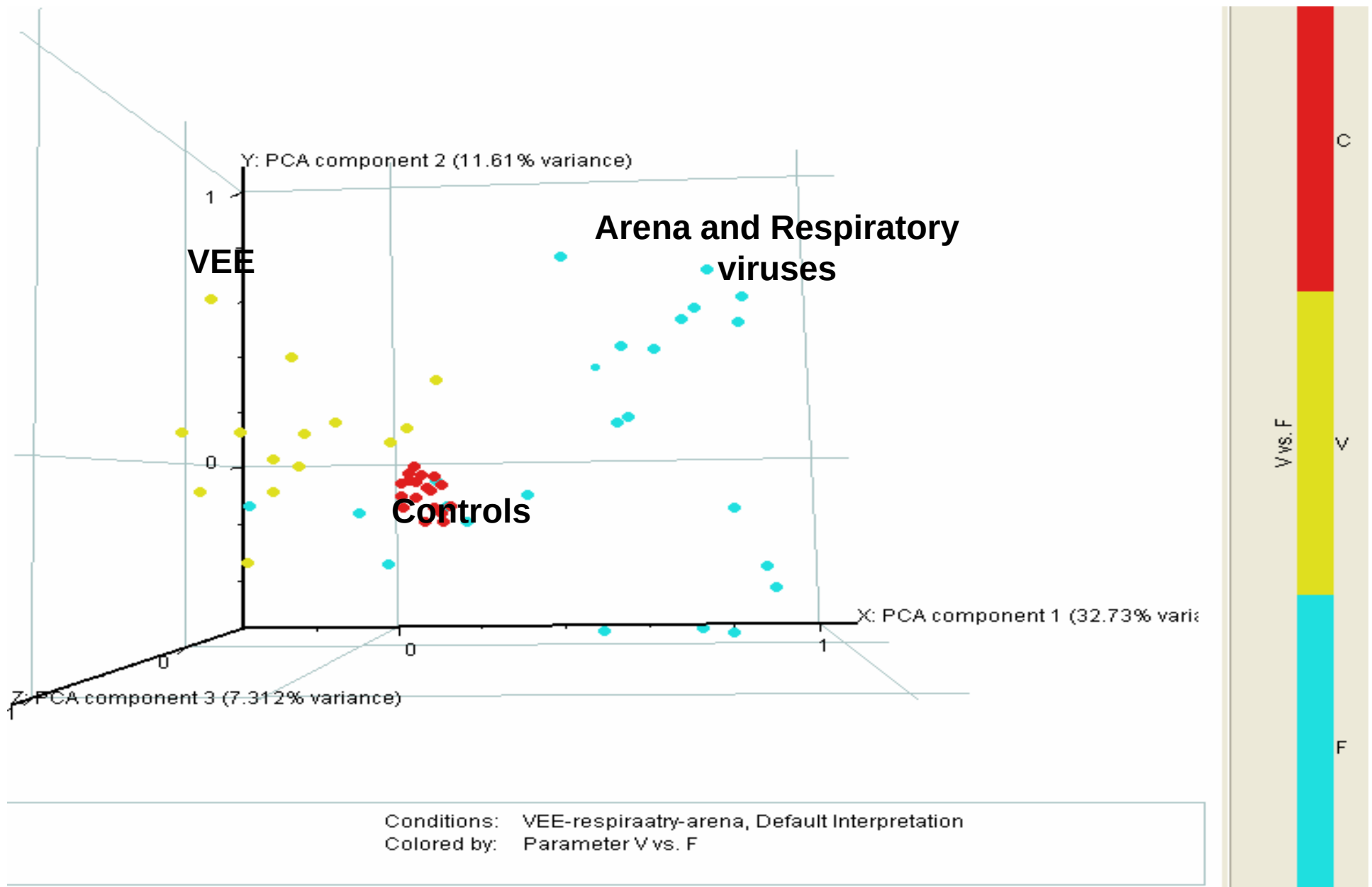


Genes that were
similarly expressed
In PBMC exposed to
Respiratory viruses
in vitro







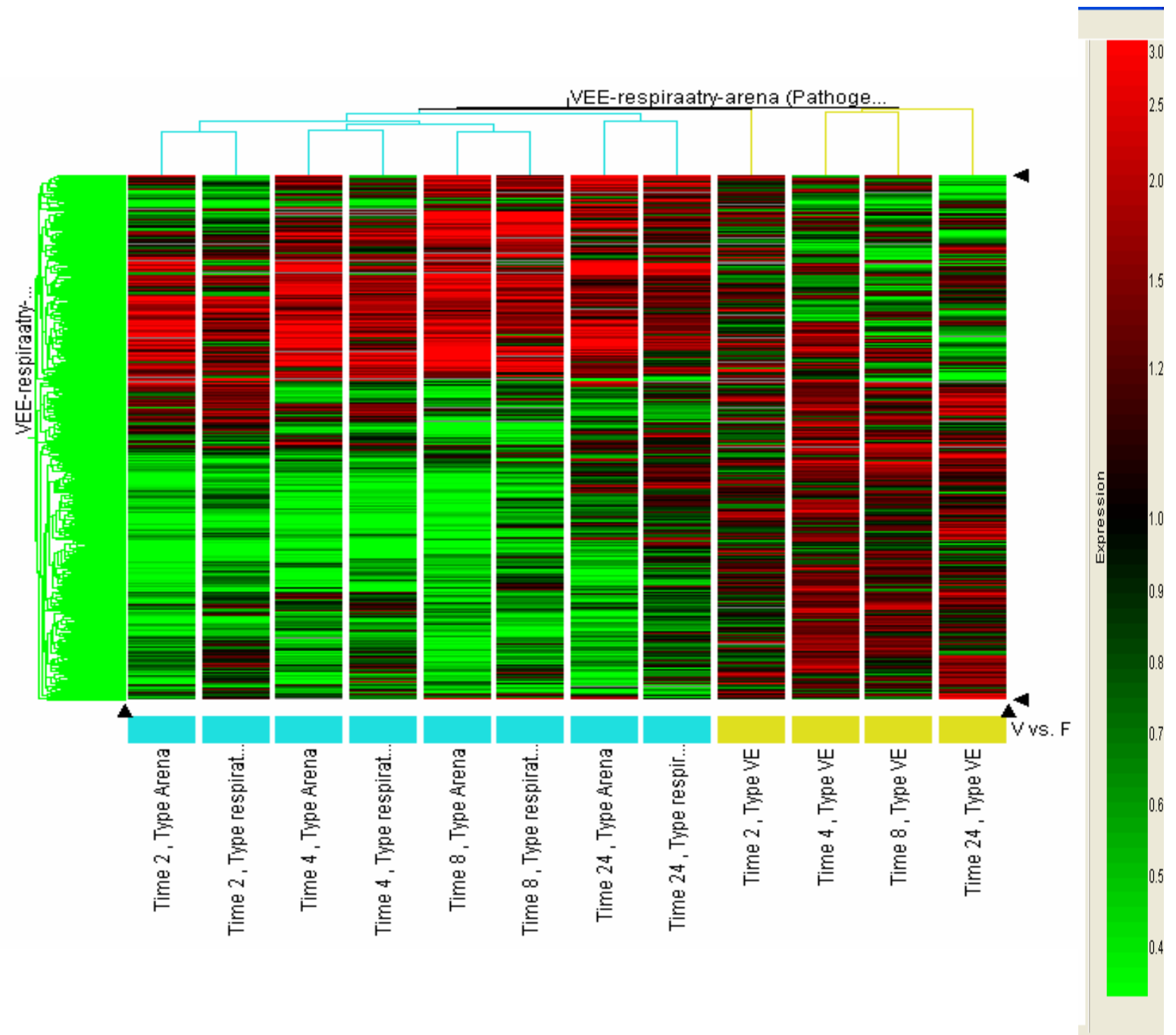


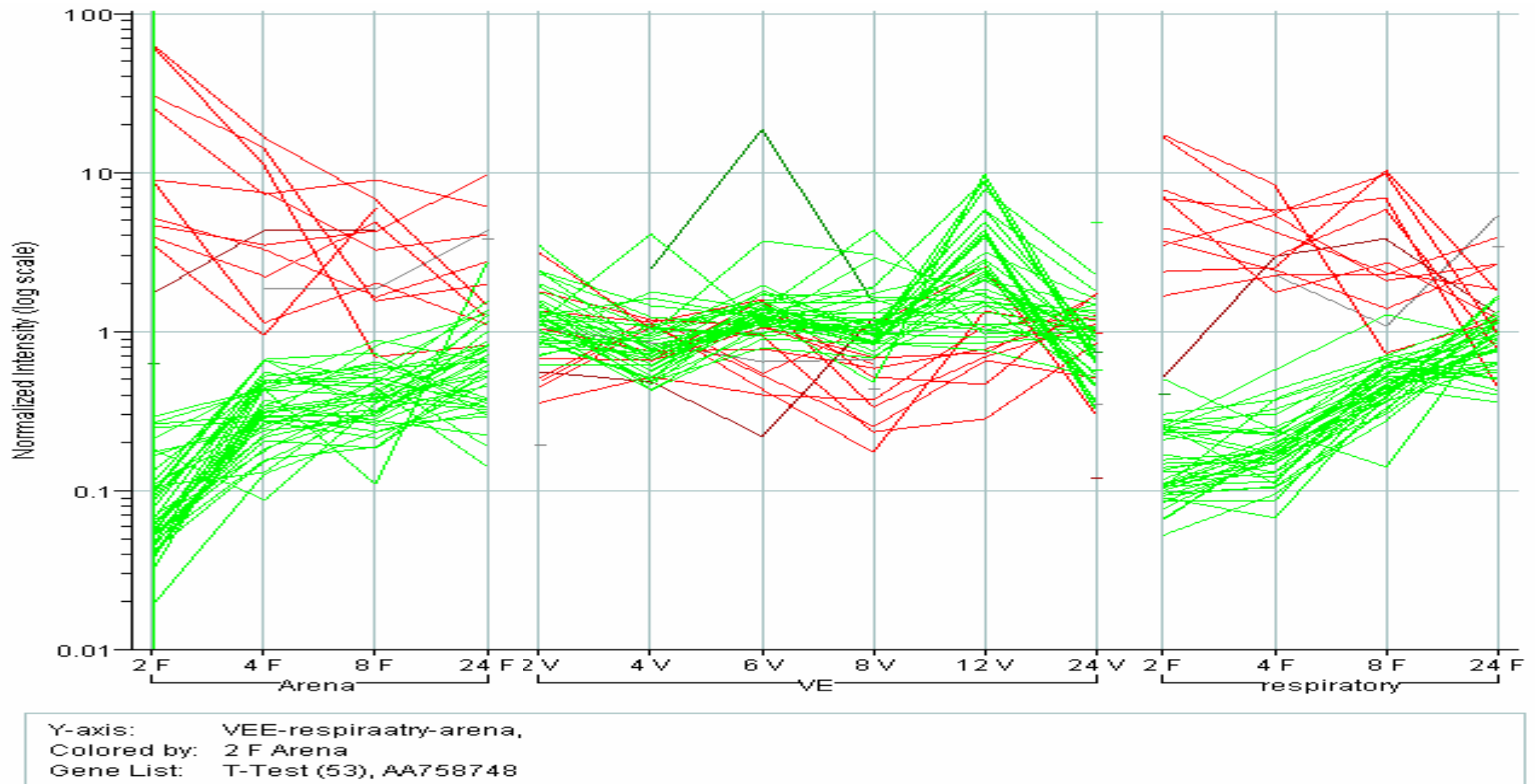
3-D scatter plot of the principal component analysis applied on samples obtained from PBMC exposed to VEE, arena and respiratory viruses.

Yellow: VEE

Red: control samples

Cluster view of genes that were significantly differentially expressed between VEE and arena and respiratory viruses. These genes were obtained using ANOVA with $p < 0.05$ followed by Benjamini correction.





Expression patterns of genes differentially expressed between VEE and Arena and respiratory viruses

Predictive modeling to identify genes signatures of exposure to VEE in vitro.

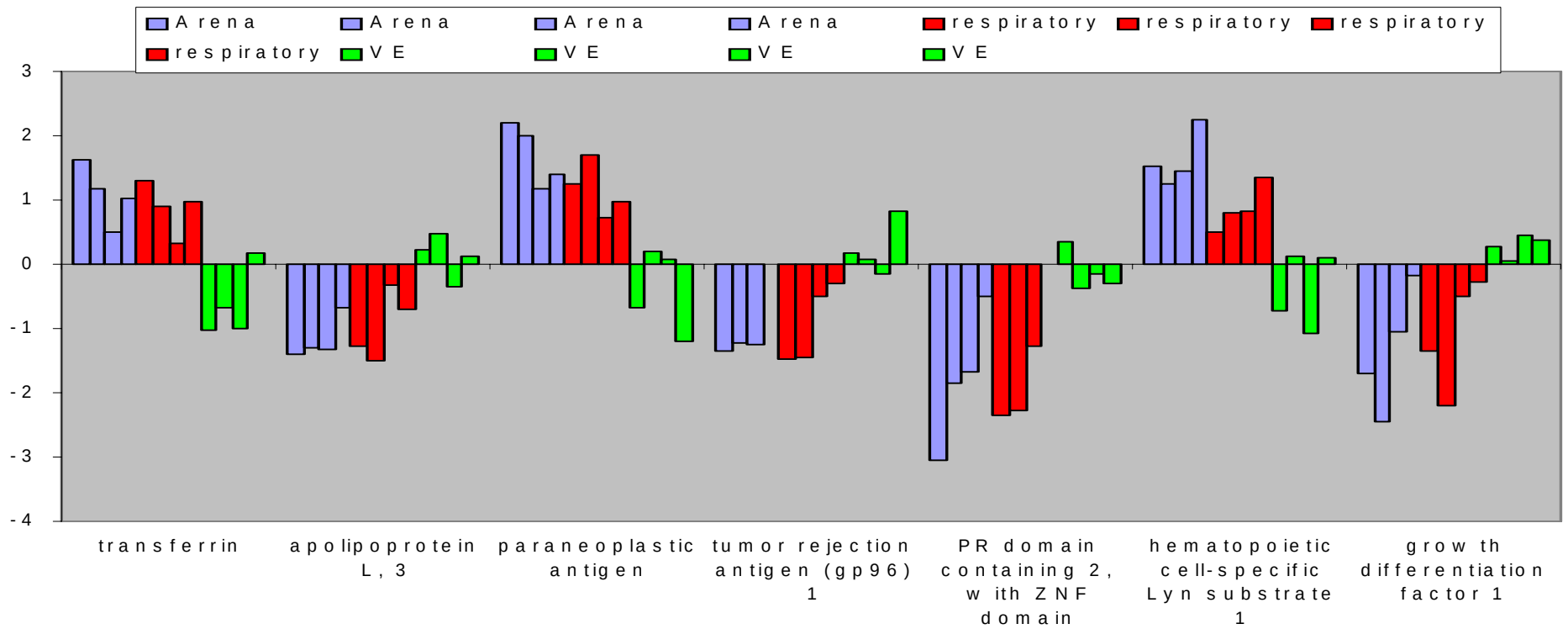
32 out of 34 samples were predicted correctly.

Classification was generated by the 'Support Vector Machines' algorithm using the training set VEE, respiratory and arena viruses at various time points and the VEE in vitro samples as a test set.

The Fisher's Exact Test method was used to select the top 30 genes from the T-Test gene list.



Sample	Predicted value
Agent C , TIme 12 , Donor 1	C
Agent C , TIme 2 , Donor 1	C
Agent C , TIme 2 , Donor 3	C
Agent C , TIme 2 , Donor 4	C
Agent C , TIme 24 , Donor 1	C
Agent C , TIme 24 , Donor 2	C
Agent C , TIme 24 , Donor 3	C
Agent C , TIme 4 , Donor 1	C
Agent C , TIme 4 , Donor 2	C
Agent C , TIme 4 , Donor 3	C
Agent C , TIme 6 , Donor 2	C
Agent C , TIme 6 , Donor 3	C
Agent C , TIme 6 , Donor 4	C
Agent C , TIme 8 , Donor 1	C
Agent C , TIme 8 , Donor 2	C
Agent C , TIme 8 , Donor 3	C
Agent C , TIme 8 , Donor 4	C
Agent VE , TIme 12 , Donor 1	V
Agent VE , TIme 2 , Donor 1	V
Agent VE , TIme 2 , Donor 3	V
Agent VE , TIme 2 , Donor 4	V
Agent VE , TIme 24 , Donor 1	V
Agent VE , TIme 24 , Donor 2	V
Agent VE , TIme 24 , Donor 3	F
Agent VE , TIme 4 , Donor 1	V
Agent VE , TIme 4 , Donor 2	V



Expression patterns of genes selected to be unique signatures for VEE

