In silico Microarray Data analysis of Ebola Hemorrhagic Fever

MALARVIZHI, A.

(10PBF06)

A THESIS SUBMITTED TO

AVINASHILINGAM DEEMED UNIVERSITY FOR WOMEN

COIMBATORE-641043

IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF SCIENCE IN BIOINFORMATICS

APRIL-2012

CERTIFICATE

In silico Microarray Data analysis of Ebola Hemorrhagic Fever

MALARVIZHI, A.

(10PBF06)

A THESIS SUBMITTED TO
AVINASHILINGAM DEEMED UNIVERSITY FOR WOMEN
COIMBATORE

IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF SCIENCE IN BIOINFORMATICS

APRIL-2012

Signature of the 30/4/12
Head of the Department

30/4/2012
Signature of the Supervisor

ACKNOWLEDGEMENT

ACKNOWLEDGEMENT

The author is profusely thankful to the **God Almighty** for his graceful blessings in all her endeavors and mercies showered upon her throughout the period of her study.

The author is grateful to **Dr. T.S.K. Meenakshi Sundarum**, Chancellor, Avinashilingam Deemed University for Women, Coimbatore, for providing her the opportunity to undertake the project.

The author records her deep sense of gratitude to **Dr. Sheela Ramachandran,** Vice Chancellor, Avinashilingam Deemed University for Women, Coimbatore, for providing all the amenities required and immense support in the conduct of the study.

The author expresses her noble indebtedness to **Dr. Saroja Prabhakaran**, Former Vice-Chancellor, Avinashilingam Deemed University for Women, Coimbatore, for her kind help and support provided for the conduct of study.

The author wishes to thanks **Dr. Gowri Ramakrishnan**, Registrar, Avinashilingam Deemed University for Women, Coimbatore, for the administrative support and help in embellishing the study.

The author express her sincere gratitude and thanks to **Dr. R. Parvatham,** Dean, Faculty of Science, Professor and Head , Department of Biochemistry, Biotechnology and Bioinformatics, Avinashilingam Deemed University for Women, Coimbatore, for her guidance and help in providing facilities to carry out the study.

The author has boundless pleasure in expressing her gratitude to **Dr. N. Santhi,** Assistant Professor, Department of Biochemistry, Biotechnology and Bioinformatics, Avinashilingam Deemed University for Women, Coimbatore, for the timely suggestions to complete the work.

The author is extremely thankful and grateful to her guide **Ms, R, Sangeetha**, Lecturer, Department of Biochemistry, Biotechnology and Bioinformatics, Avinashilingam Deemed University for Women, Coimbatore, for her meticulous care, constructive guidance, invaluable suggestions, keen interest, patience, help, encouragement, support and motivation for the good finish of the dissertation.

The author expresses sincere thanks to all **the Staff members** of the Department of Biochemistry, Biotechnology and Bioinformatics and Department of Computer Science, Avinashilingam Deemed University for Women, Coimbatore, for their help and cooperation.

The author expresses her sincere thanks to her friends, **Ms. S. Dhanalakshmi** and **Mrs. C. Jebarani,** Avinashilingam Deemed University for Women, Coimbatore, for the help rendered.

No word can fulfill the author to express her deep indebtedness to her beloved **parents, my relations and my friends** for being the motivating forces behind this dissertation and providing moral support, encouragement in carrying out the work

The author express her special thanks to all other unseen expert hands eased the pains of the present study.

MALARVIZHI, A.

CONTENTS

CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	LIST OF TABLES	
	LIST OF FIGURES	
1.0	INTRODUCTION	1
2.0	REVIEW OF LITERATURE	6
3.0	MATERIALS AND METHODS	20
4.0	RESULTS AND DISCUSSION	36
5.0	SUMMARY AND CONCLUSION	55
	BIBLIOGRAPHY	56
	WEBOGRAPHY	62

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
2.1	EBOLA HEMORRHAGIC FEVER: STATISTICS IN DIFFERENT COUNTRIES	9
4.1	DELTA TABLE FOR SIGNIFICANT GENES	44
4.2	LIST OF SIGNIFICANT GENES USING t-TEST AND SAM	45
4.3	LIST OF GENES THAT ARE UPREGULATED AND DOWNREGULATED USING SAM	47

LIST OF FIGURES

FIGURE NO.	TITLE OF THE FIGURE	PAGE NO.
2.1	STRUCTURE OF EBOLA GENOME	7
2.2	EBOLA VIRUS STRUCTURE	10
2.3	MICROSCOPIC VIEW OF EBOLA VIRION	10
2.4	ZARIE EBOLA VIRUAL REPLICATION	11
2.5	SIGNAL TRANSDUCTION PATHWAY OF THE HOST IMMUNE SYSTEM IN EBOLA HEMORRHAGIC FEVER CONDITION	12
2.6	LIFE CYCLE OF EBOLA VIRUS	14
2.7	MODES OF TRANSMISSION	16
3.1	GENE EXPRESSION OMNIBUS HOMEPAGE	21
3.2	INNATEDB HOME PAGE	23
3.3	MULTIEXPERIMENT VIEWER (MeV) HOME PAGE	26
3.4	MATLAB 7.0.4 HOME PAGE	27
3.5	GSE24943 DATASET RETRIVED FOR CLUSTERING	30
3.6	HIERARCHICAL CLUSTERING	32
3.7	SELF ORGANISING MAP	33
3.8	PRINCIPAL COMPONENT ANALYSIS	34
4.1	DATASET CHOOSEN FOR ANALYSIS USING MATLAB 7.0.4	36

FIGURE NO.	TITLE OF THE FIGURE	PAGE NO.
42	PSEUDOCOLOUR IMAGE OF EBOLA HEMORRHAGIC FEVER GENE WITH FOREGROUND AND BACKGROUND EFFECT	37
4.3	COLORMAP HOT IMAGE OF DISEASED GENE	38
4.4	SPATIAL EFFECTS IN THE FORGROUND AND BACKGROUND INTENSITIES OF PRE INFECTION OF EBOLA VIRUS	39
4.5	SPATIAL EFFECTS IN THE FORGROUND AND BACKGROUND INTENSITIES OF EBOLA HEMORRHAGIC FEVER	40
4.6	SCATTER PLOT OF EBOLA HEMORRHAGIC FEVER VS CONTROL	41
4.7	SCATTER PLOT OF EBOLA HEMORRHAGIC FEVER VS CONTROL AFTER NORMALIZATION	42
4.8	SCATTER PLOT FOR INTENSITY VS RATIO	43
4.9	SOTA DENDOGRAM	44
4.10	GENE CLUSTERS GRAPH USING KMC	45
4.11	METAGENE GRAPGH USING NMF	46
4.12	VOLCANO PLOT USING t-TEST	46
4.13	HIERARCHICAL CLUSTERING OF SIGNIFICANT GENES USING t-TEST	47
4.14	GRAPH SHOWING UPREGULATED AND DOWNREGULATED GENES IN EBOLA HEMORRHAGIC FEVER USING SAM	48
4.15	LIST OF SIGNIFICANT GENES IN EBOLA HEMORRHAGIC FEVER USING t-TEST	49

INTRODUCTION

1.0 INTRODUCTION

Ebola hemorrhagic fever is a viral hemorrhagic fever and one of the most virulent viral diseases known to humankind. The Ebola virus was first identified in the western equatorial province of Sudan and in a nearby region of Zaire in 1976 after significant epidemics in Nzara, southern Sudan and Yambuku, northern Zaire. There are five distinct species of the Ebola virus: Four of the five have caused disease in humans: Ebola-Zaire, Ebola-Sudan, Ebola-Ivory Coast and Ebola-Bundibugyo. The fifth, Ebola-Reston, has caused disease in nonhuman primates, but not in humans (http://www.who.int/csr/disease/ebola/en/).

Ebola disease is also known as hemorrhagic fever. It is a disease with hemorrhage bleedings and high temperature. It is seen in humans and nonhuman primates. Nonhuman primates include monkeys, gorillas, and chimpanzees. The term Ebola originates from the Ebola River in the Democratic Republic of Congo, formerly known as Zaire. In 1976 the first human outbreak of this virus infection occurred in the region of the Ebola River. The Ebola infection is a very severe one. It causes death in up to 90% of infected patients. Few patients can survive this dangerous infection. These are similar for the affected animals (Colebunders *et al*., 2000 and Bruce *et al.,* 2002). Therefore, Ebola infection is a significant threat to humans and animals (Ustun and Ozgurler, 2005).

The ebolaviruses and marburgviruses are enveloped, nonsegmented, negative-strand RNA viruses that belong to the family Filoviridae. There are five antigenically distinct ebolaviruses that are, 40% different in amino acid sequence, and are each named after the location of the outbreak during which they were first identified: Zaire (now known simply as Ebola virus or EBOV), Sudan virus (SUDV), Tai Forest virus (TAFV), Reston virus (RESTV) and Bundibugyo virus (BDBV) (Retuya *et al*., 1997).

The RESTV, which appears to be non-pathogenic in humans, although it remains pathogenic to non-human primates (Sanchez *et al*., 2001). Reasons why RESTV has not caused disease in humans are unclear. However, microarray analyses have shown that RESTV has a reduced ability to suppress host immune responses (Zhang *et al*., 2012).

The Ebola virus is transmitted by direct contact with the blood, body fluids and tissues of infected persons. Transmission of the Ebola virus has also occurred by handling sick or dead infected wild animals such as chimpanzees, gorillas, monkeys, forest antelope, fruit bats (http://www.who.int/csr/disease/ebola/en/).

Large quantities of ebolavirus have been found in blood as soon as two days after the onset of symptoms. The incubation period for filoviruses can be as short as two days or as long as 21 day; in most cases, symptoms appear in 4 to 10 days. Filoviruses have been reported to survive for some time in blood and tissues at room temperature (CFSPH, 2009).

These agents are a potential biological threat agent of deliberate use because the viruses have low infectious doses and a clear potential for dissemination by the aerosol route. The recent development of several candidate therapeutics and vaccines for EBOV has been promising; however, there are no approved preventive vaccines or post-exposure treatments to date (Geisbert *et al.*, 2010).

The primary difficulty for patients with EBOV infection is the failure of the immune system to react to this fastmoving disease. Patients who die from Ebola hemorraghic fever are unable to develop an adequate immune response as a result of immune dysregulation, which leads to uncontrolled virus replication and multiorgan infection and failure (Phoolcharoena *et al*., 2011).

There is no standard treatment for Ebola hemorrhagic fever. Treatment is primarily supportive and includes minimizing invasive procedures, balancing electrolytes, and, since patients are frequently dehydrated, replacing lost coagulation factors to help stop bleeding, maintaining oxygen and blood levels, and treating any complicating infections (Peters *et al.*, 2005).

The completion of whole genome sequencing projects has led to a rapid increase in the availability of genetic information. In the field of transcriptomics, the emergence of

Microarray based technologies and the design of DNA biochips allow high-throughput studies of RNA expression in cell and tissue at a given moment (Garden and Littlejohn, 2001). Today, microarray technology is one of the popular tools in molecular biology

with its main advantage being that, unlike other traditional methods, it is not limited to investigating 'one gene at a time'(Georgi *et al.,* 2005).

DNA microarray technology has proved to be a fundamental tool in studying Gene Expression. The accumulation of data sets from this technology that measure the relative abundance of mRNA of thousands of genes across tens or hundreds of samples has underscored the need for quantitative analytical tools to examine such data. Due to the large number of genes and complex gene regulation networks, clustering is a useful exploratory technique for analyzing these data (Chen *et al,* 2002).

Microarray technique can be used to measure the expression levels of tens of thousands of genes in different conditions. These conditions may be time series during a biological process. With the substantial growth of biological data, it greatly increased the challenges of understanding the results of huge data. To meet the challenges, data mining techniques specially referred to clustering algorithms were used to reveal the structure of the data and identify the interesting patterns (Valarmathie *et al*., 2012).

Gene expression omnibus is a tool in NCBI where the data was taken. Each sample has its own accession number. The microarray data is downloaded from http http://www.ncbi.nlm.nih.gov/geo/ from the database-wide expression profile for individual genes, and it is possible to standardize individual gene expression intensities in a specific assay by using their unique database-wide means and standard deviations. This consideration of gene's behavior in a wide variety of biological conditions gives new insight into interpreting the expressional difference between given samples (Guttula *et al*., 2011).

In a biological point of a view, the expressional difference of a gene with a small database wide expressional variation should have more attention than those with large database wide variations (Guttula *et al*., 2012). Cluster analysis is a one of the primary data analysis tools in data mining (Chandrasekhar *et al*., 2011).

Clustering, also called unsupervised learning, algorithms are frequently used to group genes with similar expression profiles (Svrakic *et al*., 2003). This facilitates our visualization of coexpressions of genes and also allows us to cluster arrays with similar

expression patterns. An important component of clustering algorithms, such as the hierarchical and K-mean algorithms, is to define appropriate metrics in an input space (Tibshirani *et al*., 2002).

Hierarchical clustering is a multivariate tool often used in phylogenetics and comparative genomics to relate the evolution of species (Eisen *et al*., 1998**).** K-means clustering is a typical partition based clustering method. It is an iterative process of assigning cluster memberships and re-estimating cluster parameters (Chen *et al*., 2005).

The input vectors can be either the expression profiles across different arrays for grouping genes with similar expression patterns across different subjects or the expression profiles across different genes for clustering subjects with similar microarray data. Commonly used metrics include the Euclidean distance and Pearson correlation. With these metrics, the hierarchical and K-mean algorithms can be applied for clustering. This is often presented as dendrograms or color-coded representation of similarly expressed genes (Fan and Ren, 2006).

Clustering techniques which support microarray expressed data in order to find the gene functions and regulations are

- Hierarchical clustering
- K-means clustering
- K-medoid
- Self organizing map
- Two way clustering
- Density based clustering
- Graph theoretic based clustering
- Genetic K-means
- Fuzzy clustering

(Valarmathie *et al*., 2012).

The objective of the current study includes

- Retrieval of gene expression data of Ebola Hemorrhagic Fever GSE24943 from GEO database.
- Visualization of microarray, Scatter plot distribution and interpretation Microarray Data using MATLAB.
- Clustering the co expressed gene and analysis cluster pattern of GSE24943 data sets using Multiple Array Viewer Tool.
- Indentifying significant genes using t-test and SAM in Multiple Array Viewer Tool.
- Gene Ontology study of significant genes.

REVIEW OF LITERATURE

2.0 REVIEW OF LITERATURE

Ebolavirus (EBOV) is a filovirus causing severe viral hemorrhagic fever in humans and non-human primates (Sanchez *et al.,* 2007). There are five species of EBOV: *Zaire ebolavirus* (ZEBOV), *Sudan ebolavirus* (SEBOV), *Cote d'Ivoire ebolavirus* (CIEBOV), *Reston ebolavirus* (REBOV), and *Bundibugyo ebolavirus* (BEBOV). ZEBOV has the highest virulence with a case fatality rate of 60–90% (Feldmann and Geisbert 2010). The Ebolavirus cause hemorrhagic fevers with high mortality in humans, and no effective treatments are currently approved (Richardson *et al.,* 2010), although candidate vaccines are promising (Falzarano *et al.,* 2011).

Ebola hemorrhagic fever (EHF) is one of the most severe viral infections of humans. In outbreaks in central Africa caused by the *Zaire* species of *ebolavirus* (ZEBOV), the mortality rate among identified cases has reached 80–90%, while fatalities in epidemics caused by the Sudan species have been in the range of 50–60% (Bwaka *et al.,* 1999; Sanchez *et al.,* 2006)

ZEBOV, first identified in 1976, seems to be the most virulent, killing approximately up to 90% of infected individuals, whereas REBOV, which was initially isolated from Cynomolgus monkeys imported from the Philippines into the USA in 1989, is less pathogenic in experimentally infected non-human primates (Hoch and McCormick, 1999).

Ebolavirus are often associated with limited outbreaks characterized by impressive case fatality (25% to 90%) in remote regions of Africa; they are also of significant concern from a biodefense perspective. These agents are a potential biological threat agent of deliberate use because the viruses have low infectious doses and a clear potential for dissemination by the aerosol route. The recent development of several candidate therapeutics and vaccines for EBOV has been promising (Yen *et al.,* 2011).

2.1 SCIENTIFIC CLASSIFICATION

Order : Mononegavirales

Family : Filoviridae

Genus : Ebolavirus and Marburgvirus

Species : Ebola

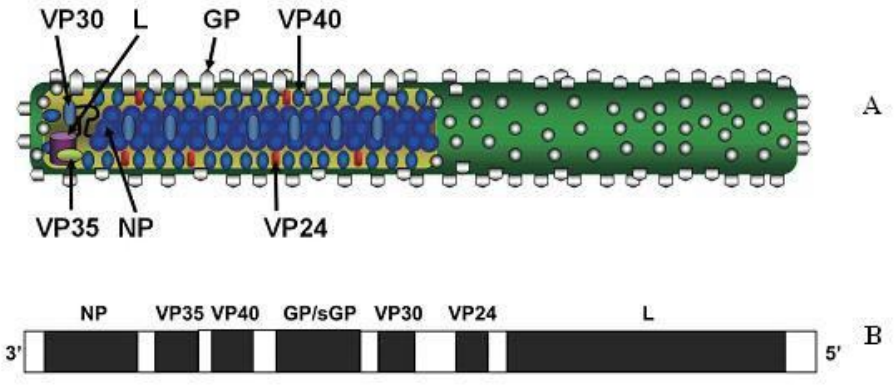
Subtypes

- Ebola-Zaire, Ebola-Sudan, Ebola-Ivory Coast, Ebola Bundibugyo - disease in humans
- Ebola-Reston-disease in nonhuman primates (Feldmann and Geisbert, 2010)

2.2 MORPHOLOGY

Ebolavirus size is enveloped, helical, cross-striated nucleocapsid, filamentous or pleomorphic virions that are flexible with extensive branching, 80 nm in diameter and 970-1200 nm in length. Nucleic acid is linear, negative-sense, single-stranded RNA, approximately 18,900 kb in length. Physicochemical properties are stable at room temperature and can resist desiccation and inactivated at 60°C for 30 minutes and its infectivity greatly reduced or destroyed by UV light and gamma irradiation, lipid solvents, b-propiolactone, formaldehyde, sodium hypochlorite, and phenolic disinfectants (Transfusion, 2009**).**

Figure 2.1
STRUCTURE OF EBOLA GENOME



A Diagrammatic representation of Ebola viral particle, **B** Negative sense genome organization

(Source: Takada, 2012)

2.3 EBOLA HEMORRHGIC FEVER MORTALITY RATE

Zaire ebolavirus (ZEBOV) outbreaks have generally been associated with the highest fatality rates, often approaching 90 percent (Sanchez *et al.,* 2006). The severity and rapid onset of disease, characterized by fever, shock and coagulation defects, are correlated with the suppression of the host innate immune system and uncontrolled viral replication (Mohamadzadeh *et al.,* 2007).

Only the few survivors of EBOV infections show detectable amounts of viral specific antibodies early in infection, suggesting that the rapid progression of disease effectively shuts down early immune responses, which also prevents the development of adaptive immune responses (Bray and Geisbert, 2005).

Because of the severity of disease, high mortality rates, and the potential use as a bioterrorist agent, filoviruses remain a significant threat to global human health (Bray and Murphy, 2007). Therefore, understanding the determinants that contribute to EBOV virulence and pathogenesis is of critical importance in order to facilitate the development of countermeasures (Reed and Mohamadzadeh, 2007),

The virus is especially dangerous for pregnant women, as it can cause heavy vaginal bleeding leading to miscarriage. The spread of Ebola can occur through direct contact with an infected person, for instance at a burial ceremony where mourners touch recently-deceased Ebola victims. Contact with blood or secretions of an infected person can also cause transmit the virus (http://focus.rw/wp/2011/05/ministry-of-health-takes-measures-to-prevent-ebola-outbreak/).

2.4 SYMPTOMS

The incubation period for Ebola Hemorrhagic Fever ranges from 2 to 21 days. The onset of illness is abrupt and is characterized by fever, headache, joint and muscle aches, sore throat, and weakness, followed by diarrhea, vomiting, and stomach pain. A rash, red eyes, hiccups and internal and external bleeding may be seen in some patients. However, it is known that patients who die usually have not developed a significant immune response to the virus at the time of death (CDC, 2010).

Table 2.1

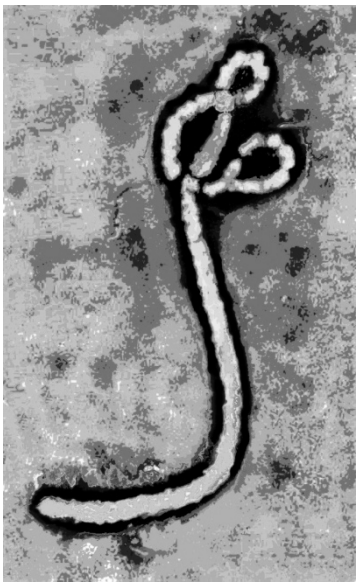
EBOLA HEMORRHAGIC FEVER: STATISTICS IN DIFFERENT COUNTRIES

Year(s)	Country	Ebola Subtype	Number of Pupil affected	Number of deaths	Situation
1967	Zaire [Democratic Republic of Congo]	Ebola-Zaire	318	280	Occurred in Yambuku and Surrounding area. Disease was spread by close personal contact and by use of contaminated needles and syringes in hospitals/clinics. This outbreak was the first recognition of the disease.
1994	Gabon	Ebola-Zaire	52	31	Occurred in Mekouka and other gold-mining camps deep in the rain forest. Initially thought to be yellow fever; identified as Ebola hemorrhagic fever in 1995.
1995	Democratic Republic of the Congo	Ebola-Zaire	315	250	Occurred in Kikwit and Surrounding area. Epidemic spread through families and hospitals.
2001-2002	Gabon	Ebola-Zaire	65	53	Outbreak occurred over the order of Gabon and the Republic of the Congo.
	Republic of Congo	Ebola-Zaire	57	43	Outbreak occurred over the border of Gabon and the Republic of the Congo. This was the first time that Ebola hemorrhagic of the Congo.
2002-2003	Republic of Congo	Ebola-Zaire	143	129	Outbreak occurred in the districts of Mbomo and Kelle in Cuvette Ouest Department.
2007	Democratic Republic of Congo	Ebola-Zaire	264	187	Outbreak occurred in Kasai Occidental Province.

Source: (Tyagi *et al.*, 2010)

Figure 2.2

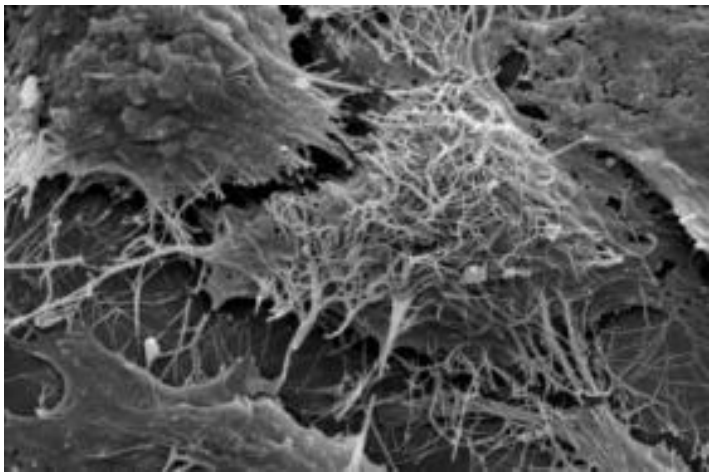
EBOLA VIRUS STRUCTURE



(Source: http://www.utmb.edu/virusimages/)

Figure 2.3

MICROSCOPIC VIEW OF EBOLA VIRION

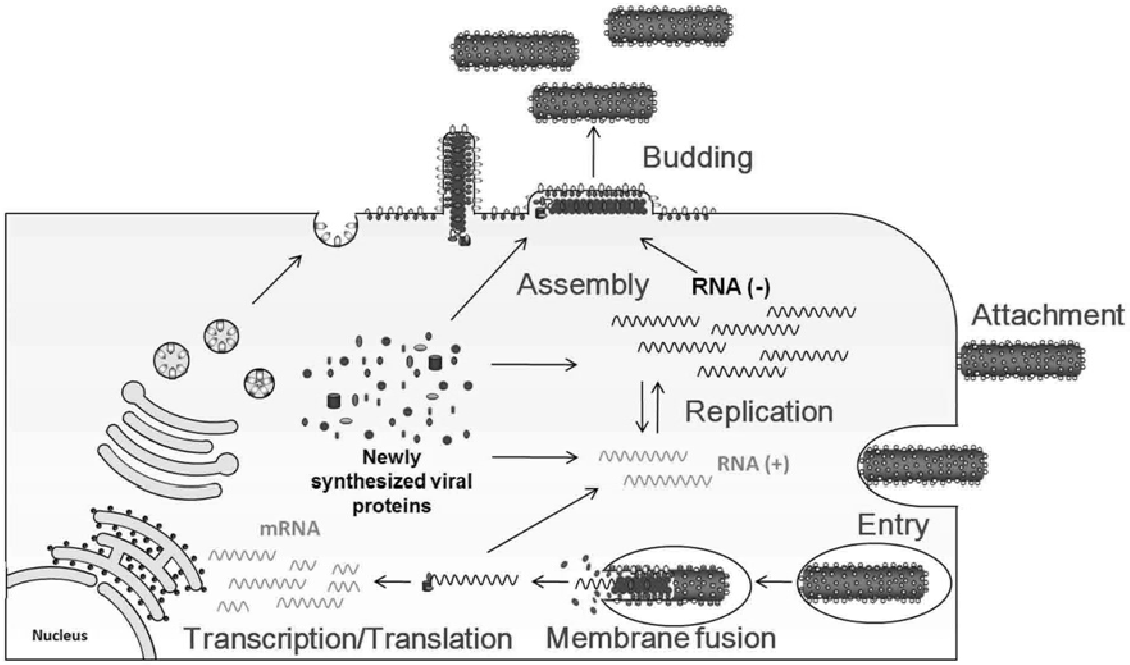


(Source: http://phys.org/news152290263.html)

Scanning electron microscope image of Ebola virions (spaghetti-like filaments) on the surface of a tetherin-expressing cell (center). The other three cells seen in this image (upper right and upper and lower left) do not have the filamentous virus on their surfaces.

Figure 2.4

ZARIE EBOLA VIRUAL REPLICATION



(Source: Takada, 2012)

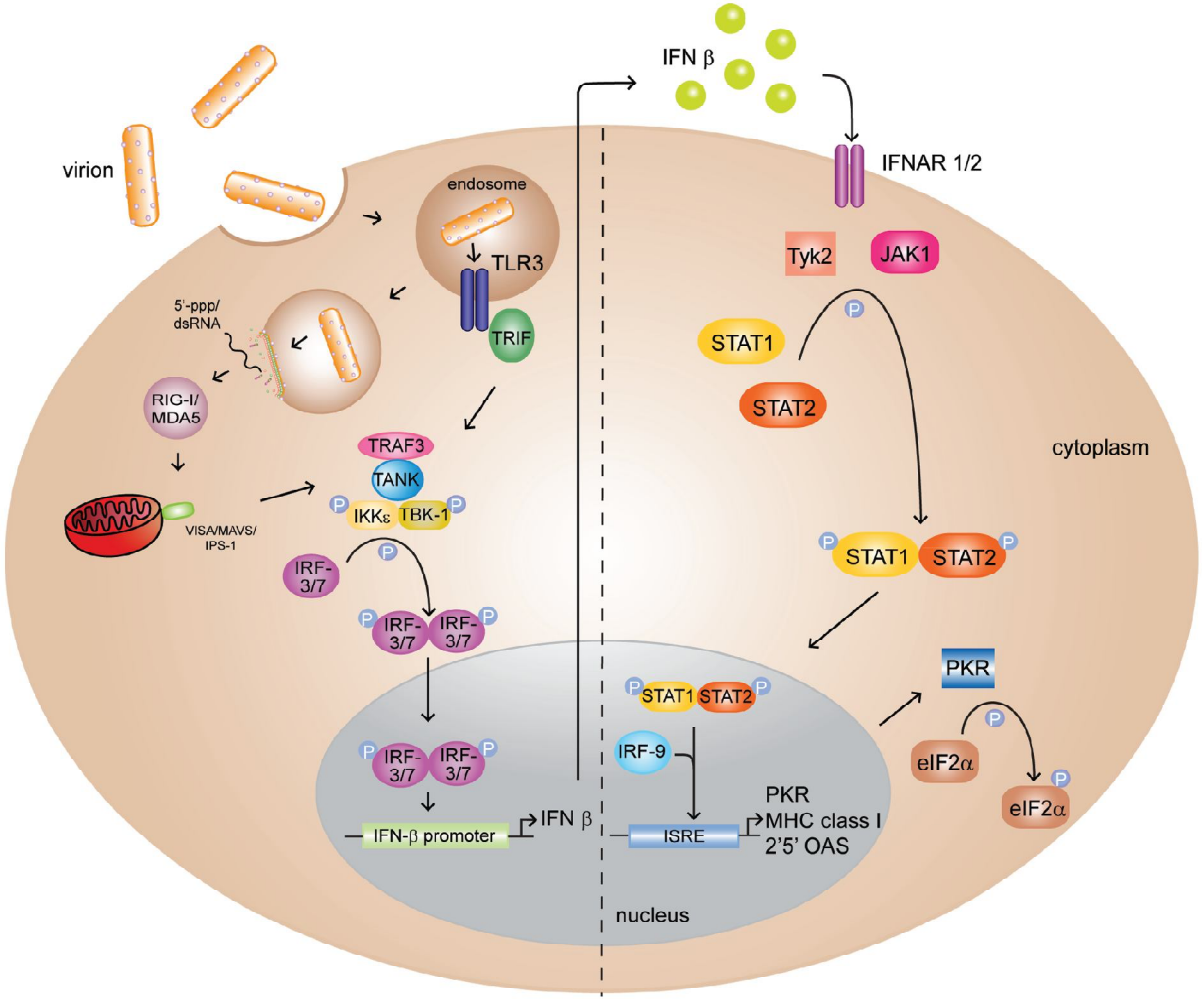
The fourth gene from the 3'end of the filovirus genome encodes the viral envelope Glycoprotein (GP), which is responsible for both receptor binding and fusion of the virus envelope with the host cell membrane (Wool and Bates, 1998). GP is highly glycosylated with large amounts of N- and O-linked glycans, most of which are uniformly located in the middle one-third of the GP, designated the mucin-like region (MLR) (Yang *et al.*, 2000 and Manicassamy *et al.*, 2007). The amino acid sequences of the MLR are highly variable among filovirus species (Sanchez *et al.*, 1998).

GP undergoes proteolytic cleavage by host proteases such as furin, which produces two subunits, GP1 and GP2, linked by a disulfide bond. The GP1 subunit mediates viral attachment, most likely through the MLR or the putative receptor binding region (Dube *et al.*, 2009). The GP2 subunit has the heptad repeat regions required for assembling GP as a trimer. The hydrophobic fusion loop on GP2 is thought to catalyze fusion of the viral envelope and host cell membrane (Ito *et al.*, 1999).

Although the trigger to promote the conformational change leading to membrane fusion is not fully understood, it was recently suggested that endosomal proteolysis of EBOV and MARV GPs by cysteine proteases such as cathepsins B and L plays an important role in inducing membrane fusion (Matsuno *et al.*, 2010). Since GP is the only viral surface GP, it is believed to have an important role in controlling the tropism and pathogenesis of filovirus infection (Sanchez *et al.*, 2007).

Figure 2.5

SIGNAL TRANSDUCTION PATHWAY OF THE HOST IMMUNE SYSTEM IN EBOLA HEMORRHAGIC FEVER CONDITION



(Source: Leung *et al.*, 2010)

2.5 HOST IMMUNE RESPONSE

The host innate immune system serves as the first line of defense against viral infections, which are mediated primarily through macrophages and dendritic cells (Basler Amarasinghe, 2009). Early suppression of host interferon (IFN) production and signaling plays a decisive factor in disease outcome (Wauquier *et al.,* 2010). The protein VP35 blocks production of IFN-α/β by binding dsRNA, a key hallmark of viral infection, and shielding it from recognition by host immune sensors such as RIG-I and MDA-5 (Kimberlin *et al.,* 2010 and Leung *et al.,* 2010).

By contrast, the protein VP24 inhibits signaling downstream of both IFN-α/β and IFN-γ. Binding to these proteins prevents them from shuttling otherwise activated, phosphorylated STAT1 to the nucleus (Reid *et al.,* 2006). STAT1 belongs to the STAT family of transcription factors, is a key mediator of the IFN response pathway and plays an essential role in the immune response to viruses. Numerous immune factors like type I and type II interferon, interleukins like IL-6 and IL-10, growth factors, angiotensin, and TNFα cause STAT1 to be phosphorylated (P-STAT1) by the Janus family kinases (JAKs) (Zhang *et al*., 2012).

Upon phosphorylation, P-STAT1 either dimerizes or forms a complex with IFN α/β-stimulated gene factor 3 (ISGF3), and is subsequently transported to the nucleus via karyopherin a proteins where it regulates genes involved in the immune response (Levy and Darnell, 2002).

2.6 INCUBATION PERIOD

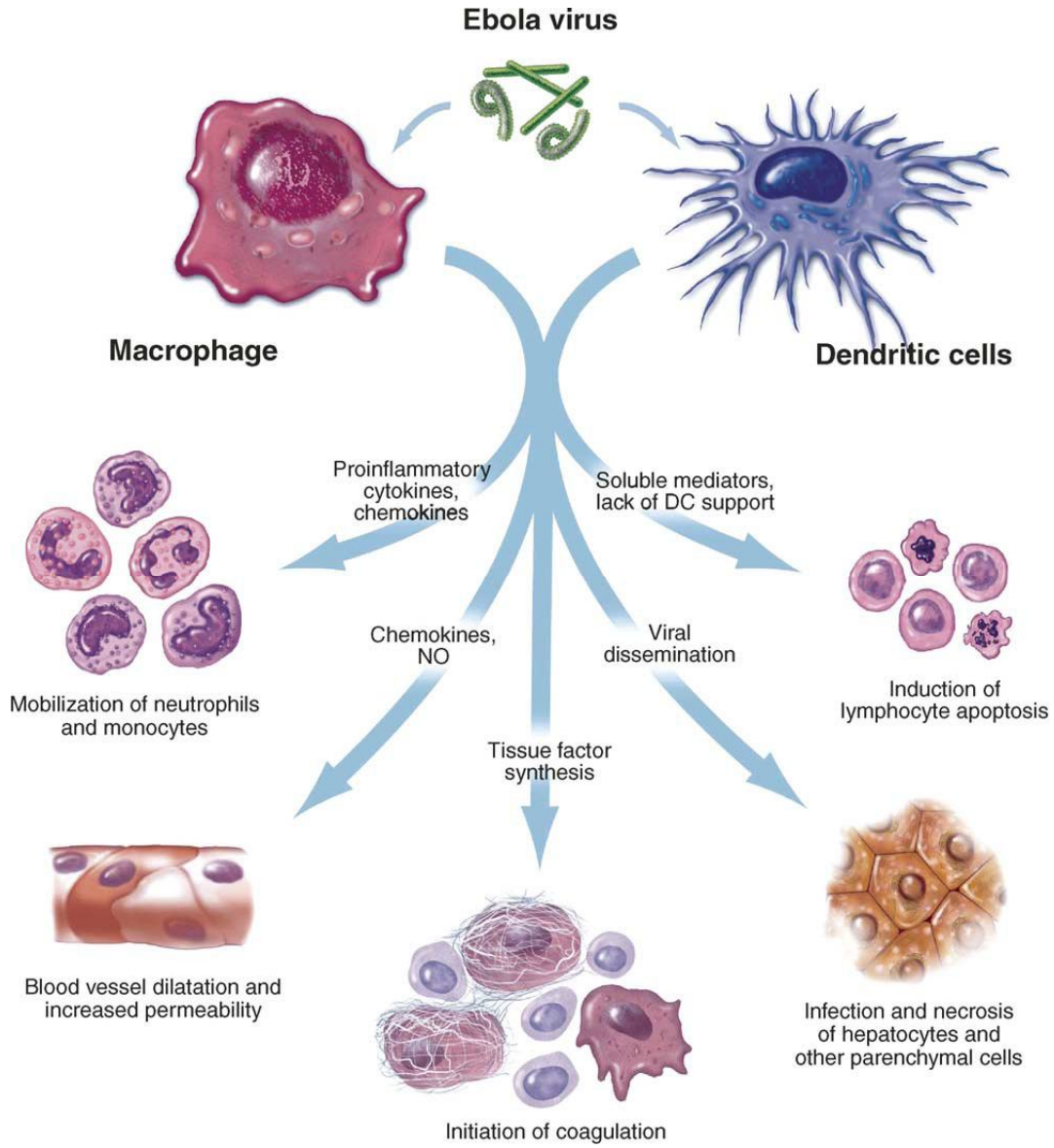
The incubation period for filoviruses can be as short as two days or as long as 21 day; in most cases, symptoms appear in 4 to 10 days. Morbidity and Mortality in human population: The African filoviruses usually have high mortality rates. *Zaire ebolavirus* is the most pathogenic virus. The case fatality rate for this infection was 59% to 88% in all outbreaks to 2008; in five of seven epidemics, it was at least 78% (CFSPH, 2009).

2.7 LIFE CYCLE OF EBOLA VIRUS

Ebola virions are able to infect a broad range of primate cells, perhaps because the heavily glycosylated surface glycoprotein (GP) can bind to a variety of target molecules, including cell surface lectins (Takada *et al.*, 2004).

Figure 2.6

LIFE CYCLE OF EBOLA VIRUS



(Source: Bray and Geisbert, 2005)

2.7.1 EBOLA INFECTION IN MACROPHAGE

Macrophages play a central role in inducing the hypotension and shock of Ebola Hemorrhagic Fever. Virus-infected macrophages also play an important role in initiating DIC by synthesizing cell-surface tissue factor (TF), which interacts with circulating factors VIIa and X to trigger the extrinsic coagulation pathway, leading to deposition of fibrin on the surface of infected cells and on membrane microparticles released into the bloodstream (Geisbert *et al.,* 2003).

Binding of coagulation factors to cell-surface TF also alters macrophage function by exciting intracellular signalling pathways, through the phosphorylation of the cytoplasmic tails of TF and associated membrane-bound protease-activated receptors (PARs) (Ruf, 2004).

The ability of ZEBOV to disseminate rapidly from its site of entry suggests that infected cells are unable to produce sufficient amounts of interferon (IFN) or respond adequately to exogenous types I or II IFN. There is good evidence that the ZEBOV VP35 protein blocks IFN production by virus-infected cells in a manner similar to the influenza virus NS1 protein, by preventing the recognition of dsRNA that normally leads to phosphorylation of IRF-3 and its translocation to the nucleus (Hartman *et al.,* 2004).

Preliminary findings suggest that a second viral protein, VP24, contributes to this process by blocking responses to exogenous IFN. Such inhibition would profoundly impair anti-viral defenses, since types I and II IFN are needed to activate NK cells, assist adaptive immunity through upregulation of major histocompatibility complex (MHC) molecules and activate macrophages and DC for effective anti-microbial function (Bray and Geisbert, 2005).

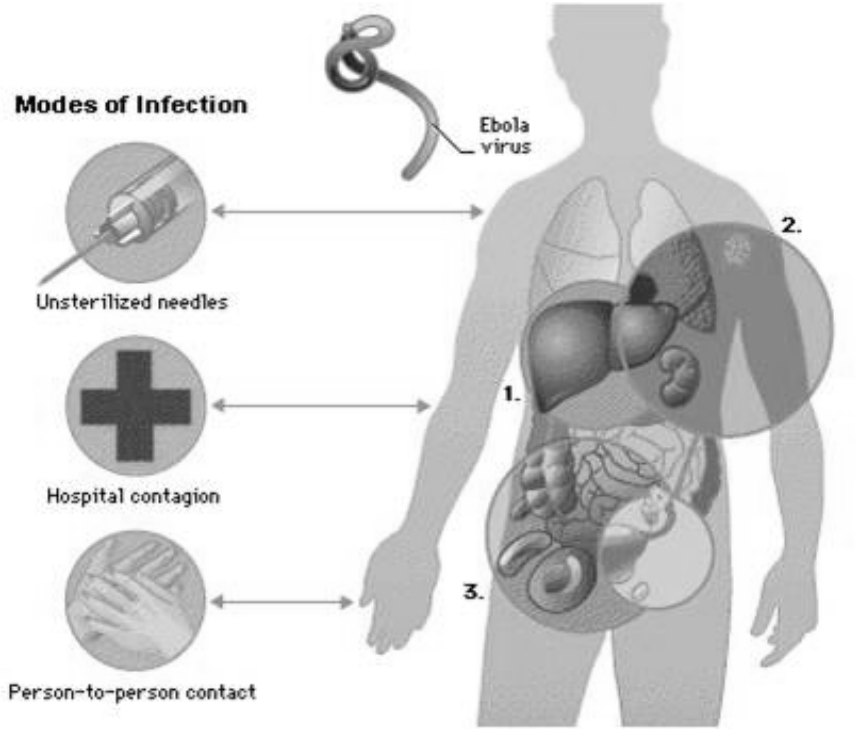
2.7.2 EBOLA INFECTION IN DENDRITIC CELL

Dendritic cells play a critical role as “gatekeepers” in the induction of antigen-specific immunity, their response to ZEBOV infection may be crucial in determining the outcome of infection. Inhibition of DC function by ZEBOV has been demonstrated by comparing the responses of human myeloid DC to noninfectious virus-like particles (VLP) or to live virus. Exposure of immature DC to VLP triggered a strong inflammatory

response, with release of TNF-α, IL-6, IL-8, and MIP-1α and induced their transformation to an antigen presenting phenotype by upregulating costimulatory molecules CD40, CD80, and CD86,MHCclasses I and II surface antigens, downregulating chemokine receptor CCR5 and upregulating CCR7 (Bosio *et al.*, 2003).

Figure 2.7

MODES OF TRANSMISSION



(Source: http://microbewiki.kenyon.edu/images/9/93/Transfer.jpg)

The Ebola virus is transmitted by direct contact with the blood, secretions, organs or other body fluids of infected persons. Burial ceremonies where mourners have direct contact with the body of the deceased person can play a significant role in the transmission of Ebola (Bell, 2007).

Health care workers have frequently been infected while treating Ebola patients, through close contact without correct infection control precautions and adequate barrier nursing procedures. The virus has been confirmed to be transmitted through body fluids. Transmission through oral exposure and through conjunctiva exposure is likely, which have been confirmed in non-human primates (Johnson *et al.*, 1995). Filoviruses are not

naturally transmitted by aerosol. They are, however, highly infectious as breathable 0.8-1.2 micron droplets in laboratory conditions (Tyagi *et al.*, 2010).

2.8 PATHOGENICITY

The major early targets of ZEBOV infection are two types of cells: macrophages, which employ a battery of innate immune mechanisms for initial anti-viral defense, and dendritic cells (DC), which have innate immune functions, but also specialize in initiating adaptive immune responses by presenting antigens to naive T cells. ZEBOV infection partially impairs the function of both cells, so that they are able to initiate inflammation and coagulation, but cannot prevent the systemic spread of virus. In consequence, additional target cells are attracted to sites of infection, and virus disseminates to resident macrophages and DC in tissues throughout the body, causing massive release of proinflammatory mediators and vasoactive substances (Hensley *et al.*, 2002).

These host responses produce a syndrome of refractory hypotension and DIC resembling septic shock, which results from the response of the same cell populations to endotoxin and other bacterial products. The extensive tissue injury caused by replication of ZEBOV in macrophages and DC and in parenchymal cells of the liver and other organs also plays a major role in fatal disease. Natural killer (NK) cells and T lymphocytes remain uninfected, but undergo apoptosis, further impairing immune function (Bray and Murphy, 2007).

2.9 DIAGONOSIS

Diagnosing Ebola HF in an individual who has been infected only a few days is difficult because early symptoms, such as red eyes and a skin rash, are nonspecific to the virus and are seen in other patients with diseases that occur much more frequently. However, if a person has the constellation of symptoms described above, and infection with Ebola virus is suspected, isolate the patient and notify local and state health departments and the Centers for Disease Control and Prevention, (CDC, 2010).

There is no FDA-approved treatment or vaccine for ebolavirus infections. Ebolaviruses have been categorized by NIH/NIAID as Category **A** Priority Pathogens because they could be misused for the development of biological weapons (Ou *et al.*, 2012).

2.10 INSILICO GENE EXPRESSION STUDIES USING MICROARRAY DATA

Functional genomics involves the analysis of large datasets of information derived from various biological experiments. One such type of large-scale experiment involves monitoring the expression levels of thousands of genes simultaneously under a particular condition, called gene expression analysis (Babu, 2004).

2.10.1 IMAGE ANALYSIS OF MICROARRAY DATA

Microarray image quantitation involves various steps, including addressing or gridding, segmentation or background separation, and normalization. The analysis of DNA microarray gene expression data involves two main steps and they are i) image quantitation, the extraction of gene expression data, ii) gene expression data analysis, in which after the ratios of the intensities are obtained, various methods can be applied to cluster the genes into different functional groups based on the ratios retrieved in the image quantification (Rueda and Qin, 2005).

2.10.2 CLUSTER ANALYSIS OF MICROARRAY DATA

Most techniques for analyzing microarray data can be thought of as either 'supervised' or 'unsupervised'. Supervised methods require the genes or conditions to be associated with labels that provide information about a pre-existing classification (Raychaudhuri *et al*., 2001). Cluster analysis partitions gene data into meaningful clusters which capture the natural structure of the data to find genes that have similar functionality (Pati and Das, 2011).

Hierarchical clustering strategies are easy to implement but suffer because the decision about where to create branches and in what order the branches should be presented can be arbitrary. K-means clustering requires a parameter, k, the number of expected clusters; initially cluster centers are selected randomly. In each iteration of the algorithm, all of the profiles are assigned to the cluster whose center they are nearest to (using the distance metric), and then the cluster center is recalculated based on the profiles within the cluster (Sherlock, 2000).

Instead of simply partitioning data into disjoint clusters, self-organizing maps organize the clusters into a 'map' where similar clusters are close to each other (Toronen

et al., 1999). The computational details are similar to k-means clustering except that cluster centers are recalculated at each iteration using both the profiles within the cluster as well as the profiles in adjacent clusters (Raychaudhuri *et al*., 2001).

2.10.3 STATISTICAL METHODS USED FOR ANALYSIS OF MICROARRAY DATA

The t-test is probably the most commonly used Statistical Data Analysis procedure for hypothesis testing. Actually, there are several kinds of t-tests, but the most common is the "two-sample t-test" also known as the Student's t-test or the independent samples t-test (http://www.statisticallysignificantconsulting.com/Ttest.htm).

The t-test is used to test differences in means between two groups. The t-test is used when the dependent variable is a continuous interval/ratio scale variable and the independent variable is a two-level categorical variable. The t-test can be used even if sample sizes are very small, as long as the variables within each group are normally distributed and the variation of scores within the two groups is equal (http://allnurses.com/general-nursing-articles/t-tests-one-378029.html).

SAM identifies genes with statistically significant changes in expression by assimilating a set of gene-specific t tests. Each gene is assigned a score on the basis of its change in gene expression relative to the standard deviation of repeated measurements for that gene. Genes with scores greater than a threshold are deemed potentially significant. The percentage of such genes identified by chance is the false discovery rate (FDR). To estimate the FDR, nonsense genes are identified by analyzing permutations of the measurements. The threshold can be adjusted to identify smaller or larger sets of genes, and FDRs are calculated for each set (Tusher *et al*., 2001).

MATERIALS & METHODS

3.0 MATERIALS AND METHODS

The following tools and databases are used for gene expression study on Ebola Hemorrhagic fever. Microarray techniques provide a platform where one can measure the expression levels of thousands of genes in hundreds of different conditions whereas using traditional methods in molecular biology can only report the expression levels of single genes. Many analysis tools help biologists to discover co-expressed genes (Babu, 2004). Among these, microarray which is the most familiar tool helps in these aspects. Microarray can also be used to determine which genes are expressed in which tissues and at which times during embryonic development (Butte, 2002).

3.1 DATABASES

Databases are repositories of information, generally store on computers. They contain an organized series of records that can be searched through and displayed in the computer screen, downloaded or emailed to a specific address.

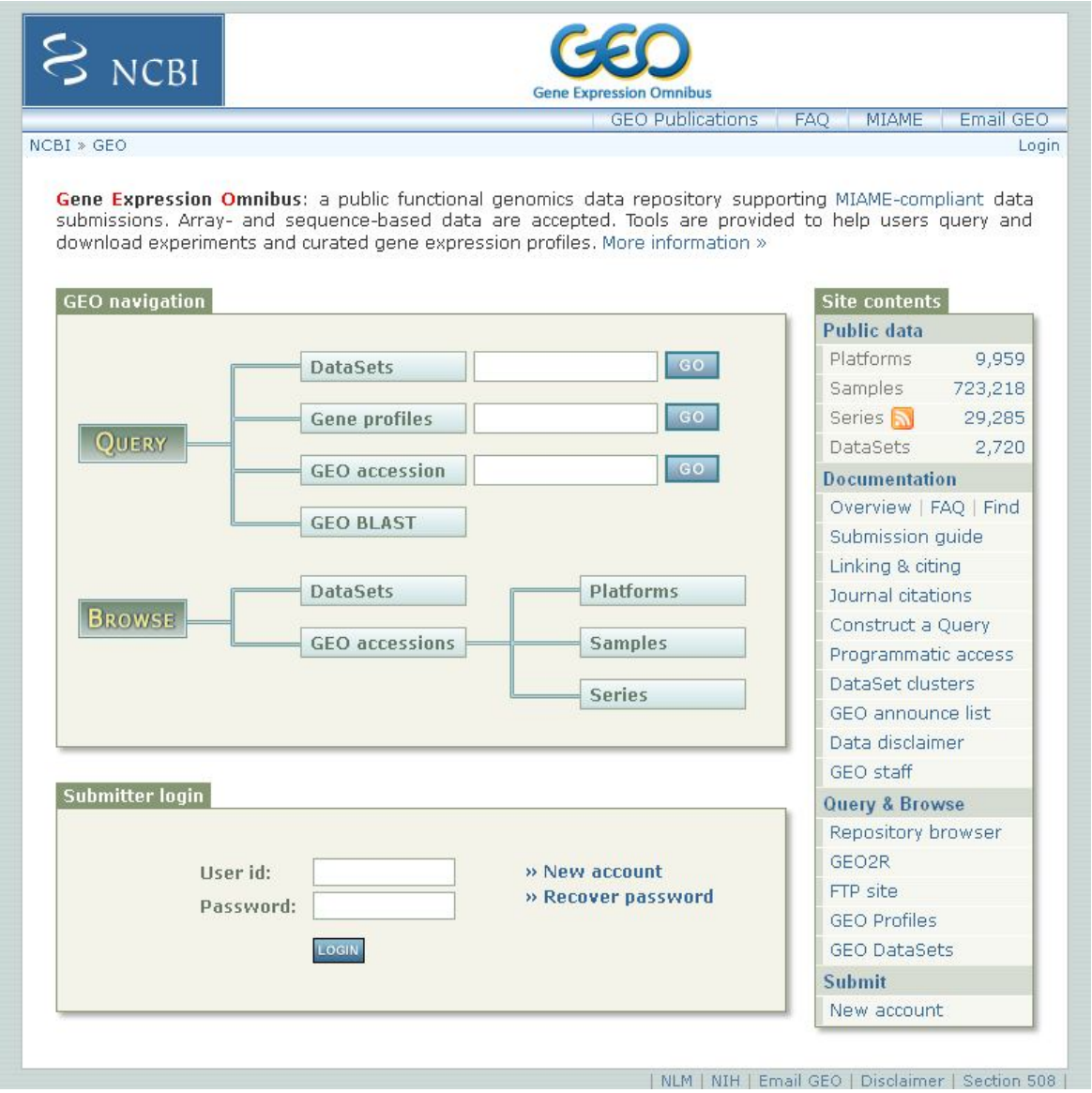
3.1.1 GENE EXPRESSION OMNIBUS (GEO)

Gene expression omnibus is a tool in NCBI where the data was taken. Each sample has its own accession number. The accession number can either start with GSD or GSM followed by the number. The other format is not accepted by the tool. Gene expression omnibus consist of samples in tab delimited form which either opens in note pad or Microsoft excel (Guttula *et al.*, 2012).

The Gene Expression omnibus maintained by NCBI-National Center for Biotechnology Information (http://www.ncbi.nih.gov/geo/), is one of the largest and most widely used database. It is a public functional genomic data repository and freely distributes high-throughput gene expression data submitted by the scientific community (Edger *et al.*, 2002).

Figure 3.1

GENE EXPRESSION OMNIBUS HOMEPAGE



3.1.1.1 GEO DATASETS

The GEO Datasets database stores original submitter-supplied records such as Series, Samples and Platforms as well as curated Datasets. The GEO Datasets database can be searched using many different attributes including keywords, organism, Dataset type and authors. Dataset records contain additional resources, including cluster tools and differential expression queries (http://www.ncbi.nlm.nih.gov/geo/info/datasets.html).

3.1.1.2 GEO PROFILES

The GEO Profiles database stores gene expression profiles derived from curated GEO Datasets. Each Profile is presented as a chart that displays the expression level of one gene across all Samples within a Dataset. GEO Profiles can be searched using many different attributes including keywords, gene symbols, gene names, GenBank accession numbers, or Profiles flagged as being differentially expressed (http:// www.ncbi.nlm.nih.gov /geo/info/profiles.html).

3.1.2 GENE ONTOLOGY (GO)

Gene ontology is a major public annotation effort, which provides descriptions of the molecular functions, biological processes and sub-cellular locations attributed to gene products from all organisms. GO links primarily biological knowledge to information provided in highly-controlled, structured vocabularies, and is designed to improve the accessibility of scientific knowledge to search engines and algorithmic processing. Consequently, GO has proved to be highly beneficial to investigators who need to understand and analysis large amount of data produced from a range of high-throughput investigative techniques (http://www.geneontology.org/).

3.1.3 INNATEDB

InnateDB is a publicly available database of the genes, proteins, experimentally-verified interactions and signaling pathways involved in the innate immune response of humans, mice and bovines to microbial infection. The database captures an improved coverage of the innate immunity interactome by integrating known interactions and pathways from major public databases together with manually-curated data into a centralized resource. Over 17,738 interactions have been manually curated to date by the InnateDB curation team. The database can be mined as a knowledgebase or used with our integrated bioinformatics and visualization tools for the systems level analysis of the innate immune response (http://www.innatedb.ca/).

Figure 3.2

INNATEDB HOME PAGE

3.2 TOOLS USED FOR ANALYSIS OF DIFFERENTIALLY

EXPRESSED GENES

Human Ebola virus (EBOV) infection causes hemorrhagic fever and death within a few days. EBOV forms the *Filoviridae* family, a group of enveloped, nonsegmented, Negative-strand RNA viruses. The genome is 19,000 nucleotides long and bears linearly arranged genes that encode seven structural proteins and one nonstructural protein (Leroy *et al*., 2004).

The *in silico* microarray data analysis was carried out with the online available data set; GSE24943, the dataset was collected from GEO database In this *in silico* analysis study, the GenePix Results of GSE24943 array was used. 44 raw data samples from GSE24943 were used for the further analysis.

3.2.1 TM4 SOFTWARE

MultiExperiment Viewer is one member of a suite of microarray data management and analysis applications originally developed at The Institute for Genomic Research (TIGR). Within the suite, known as TM4, there are four programs: MADAM, Spotfinder, MIDAS and MeV. Together, they provide functions for managing microarray experimental conditions and data, converting scanned slide images into numerical data, normalizing the data and finally analyzing that normalized data. These tools are all OSI certified open-source and are freely available through the TM4 website, http://www.tm4.org.

The Microarray Data Manager (MADAM) is a data management tool used to upload, download, and display a plethora of microarray data to and in a database management system (MySQL). An interface to MySQL, Madam allows scientists and researchers to electronically record, capture, and administrate annotated gene expression and experiment data to be shared with and ultimately used by others within the scientific community.

TIGR Spotfinder is image-processing software created for analysis of the image files generated in microarray expression studies. TIGR Spotfinder uses a fast and reproducible algorithm to identify the spots in the array and provide quantification of expression levels.

Microarray Data Analysis System (MIDAS**)** is an application that allows the user to perform normalization and data analysis by applying statistical means and trim the raw experimental data, and create output for MeV.

MultiExperiment Viewer (MeV) is an application that allows the user to view processed microarray slide representations and identifies genes and expression patterns of interest. Slides can be viewed one at a time in detail or in groups for comparison purposes. A variety of normalization algorithms and clustering analyses allow the user flexibility in creating meaningful views of the expression data (User manual MeV Version 4.6.2)

3.2.1.1 MULTIEXPERIMENT VIEWER (MEV)

TIGR Multiexperiment Viewer is software for microarray data analysis. MeV is developed by a group of people at TIGR (The Institute For Genomic Research).It can handle several input file formats. The installation is fairly simple and running MeV requires (JRE) Java Runtime environment using a well defined module API (Application Programming Interface).

MeV is a desktop application for the analysis, visualization and data mining of large scale genomic data. It is a versatile microarray tool, incorporating sophisticated algorithms for clustering, visualization, classification, statistical analysis and biological theme discovery.

MeV generates informative and interrelated displays of expression and annotation data from single or multiple experiments. A huge array of algorithms are included in MeV modules, and are available at a button-click, such as K-mean clustering, Hierarchical clustering, self organizing Maps, t-Tests, Significance Analysis of Microarrays and Principal component Analysis (http://www.tm4.org/mev/about).

3.2.1.1.1 MODULES IN MEV

KNNC - K-NEAREST NEIGHBORS CLASSIFICATION)

KNN Classification is a supervised classification scheme. A subset of the entire data set (called the training set), for which the user specifies class assignments, is used as input to classify the remaining members of the data set. The user specifies the number of expected classes, and the training set should contain examples of each class.

SAM - SIGNIFICANCE ANALYSIS OF MICROARRAYS

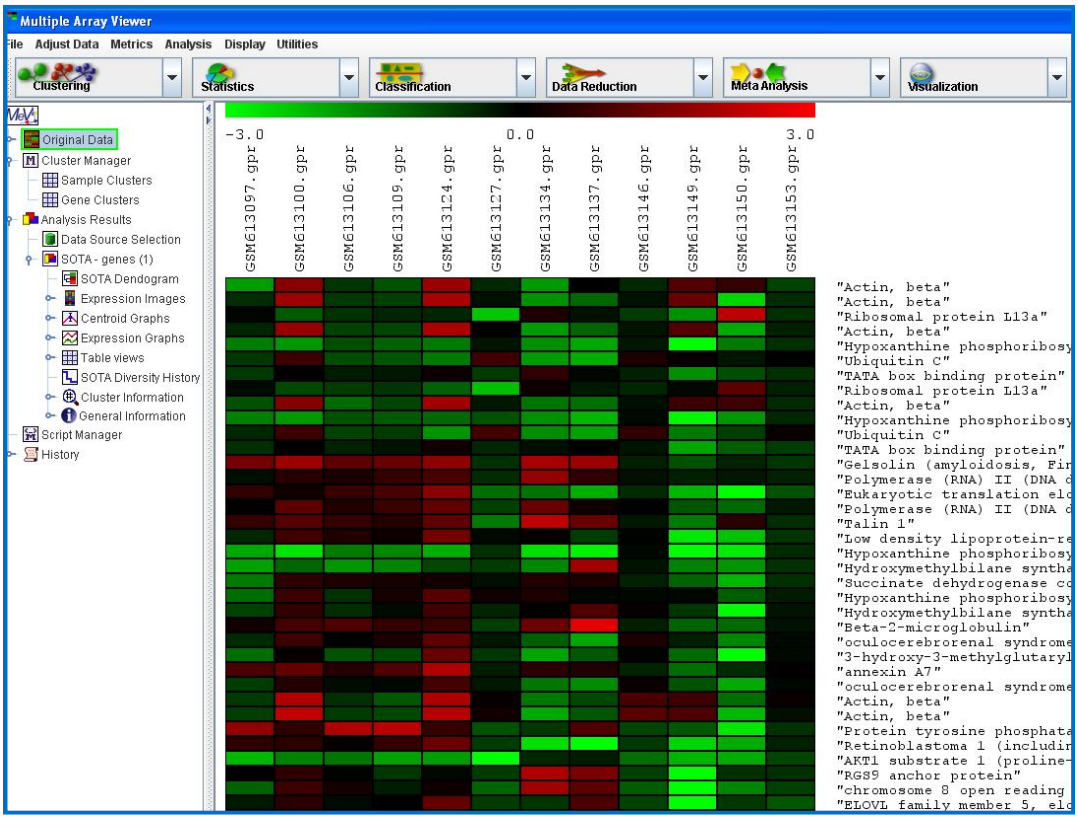
The SAM module is an implementation of the Tusher *et al*, 2001 paper describing the method of determining significance of gene expression changes between samples. SAM is useful when there is an a-priori hypothesis that some genes will have significantly different mean expression levels between different sets of samples.

GSEA - GENE SET ENRICHMENT ANALYSIS

GSEA is a computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (e.g. phenotypes) (http://www.tm4.org/mev/features).

Figure 3.3

MULTIEXPERIMENT VIEWER (MeV) HOME PAGE



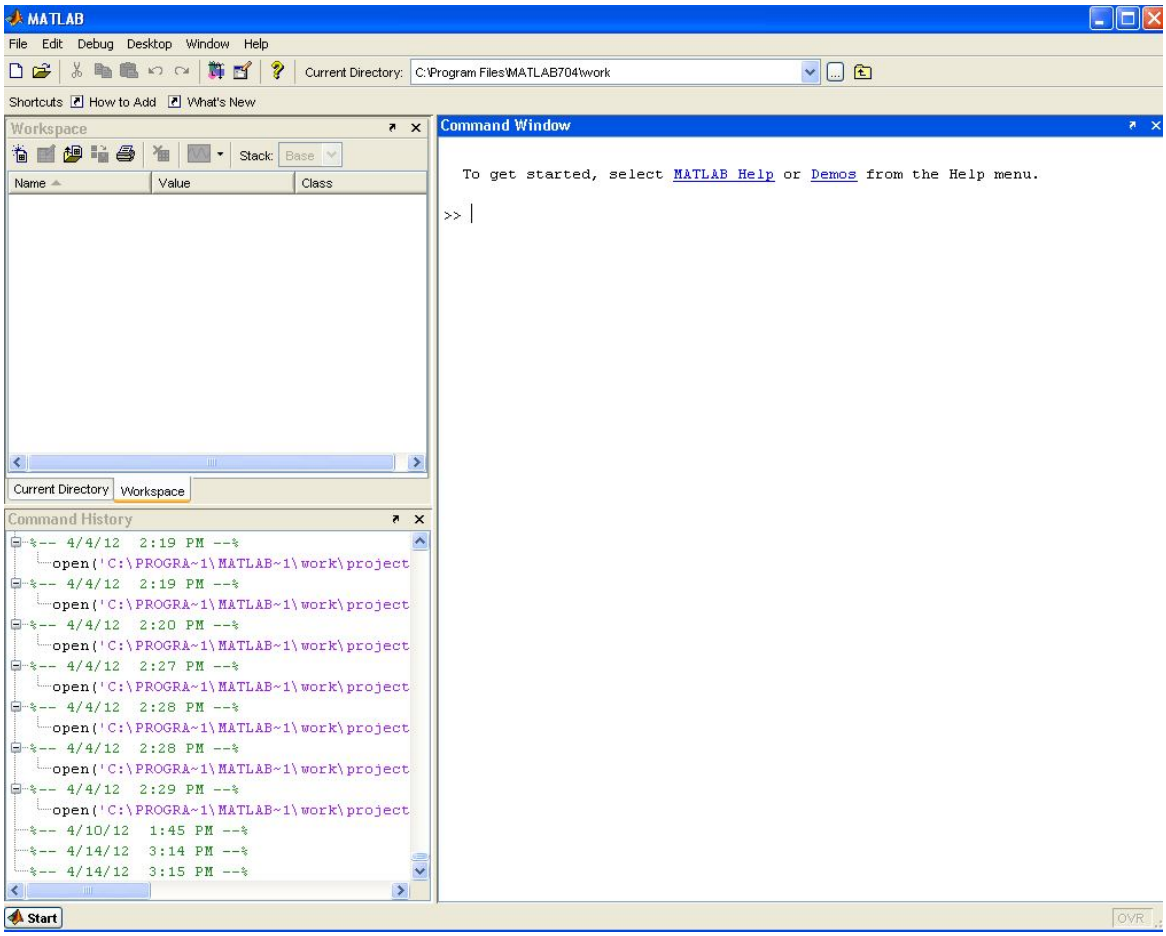
3.2.2 MATLAB VERSION 7.0.4

The Matlab environment presents many advantages for the analysis of microarray data: it offers an efficient and natural way of dealing with large data sets, it provides a comprehensive set of functions, and some of the complex tools used in microarrays, like principal components analysis, are built-in. Moreover, many toolboxes have been developed and are often freely available for more specialized needs, like for instance neural networks or support vector machines. However, quite surprisingly, more basic tasks like hierarchical clustering or nonlinear normalization are not available, or only in a very inefficient implementation. The main tools present are: normalization procedures,

including the loess-type non-linear normalization, a fast implementation of the hierarchical and K-means clustering (Venet, D 2003).

FIGURE 3.4

MATLAB 7.0.4 HOME PAGE



3.3 METHOLOGY

3.3.1 DATA SETS

Gene Expression Omnibus is a public functional genomics data repository supporting MIAME-compliant data submissions. Array and sequence based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles (http://www.ncbi.nlm.nih.gov/geo/). The microarray gene expression data of "GSE24943" Whole-genome transcriptional analysis of Ebola hemorrhagic fever was downloaded from GEO database. The datasets consists of 90

replicates from that 44 datasets were chosen selectively based experiment conducted in microarray studies.

3.3.2 VISUALIZING MICROARRAY DATA USING MATLAB 7.0.4

The GenePix GPR format files retrieved from"GSE24943**"** are GSM613097 and GSM613100 microarray datas are used in the study. The GSM613097 is used as control. It was taken from blood sample of day 0 post-infection of Ebola hemorrhagic fever, rNAPc2 treatment, Recombinant nematode anticoagulant protein-2 (rNAPc2) is an Antithrombotic protein and GSM613100 was taken from blood sample of day 10 post-infection of Ebola hemorrhagic fever, rNAPc2 treatment.

Source File I:

GO Accession : GSM613097

Organism : *Macaca mulatta*

Source name : Blood sample taken day 0 post-infection, rNAPc2 treatment is a Recombinant nematode anticoagulant protein-2 (rNAPc2) is an Antithrombotic protein

Source File II:

GO Accession : GSM613100

Organism : *Macaca mulatta*

Source name : Blood sample taken day 10 post-infection, rNAPc2 treatment.

GPR format files provide a large amount of information about the array including the mean, median and standard deviation of the foreground and background intensities of each spot at the 635nm wavelength (the red, Cy5 channel) and the 532nm wavelength (the green, Cy3 channel) (Matlab, 2012).

3.3.2.1 SPATIAL IMAGES ANALYSIS OF MICROARRAY DATA

This method will illustrates how to visualize microarray data by plotting image maps. The function maimage to show pseudocolor images of the foreground and background. You can use the function subplot to put all the plots. The function maimage

can take a microarray data structure and create a pseudocolor image of the data arranged in the same order as the spots on the array (Matlab, 2012).

3.3.2.2 STATISTICS OF THE MICROARRAY DATA

These methods will illustrate how to visualize distributions in microarray data. The function maboxplot used to look at the distribution of data in each of the blocks. From the box plots you can clearly see the spatial effects in the background intensities (Matlab, 2012).

3.3.2.3 SCATTER PLOTS OF MICROARRAY DATA

This method illustrates how to visualize expression levels in microarray data. There are two columns in the microarray data structure labeled 'F635 Median - B635' and 'F532 Median - B532'. These columns are the differences between the median foreground and the median background for the 635 nm channel and 532 nm channel respectively. These give a measure of the actual expression levels, although since the data must first be normalized to remove spatial bias in the background. The function mairplot is used to create Intensity vs. Ratio plot for the normalized data (Matlab, 2012).

These are various methods to visualize, distribute and analyze gene expression of Ebola hemorrhagic fever. By clicking on the plot we can read the description of gene associated with that plot.

3.3.3 CLUSTERING ANALYSIS

The microarray gene expression data of “GSE24943**“**Whole-genome transcriptional analysis of Ebola hemorrhagic fever was downloaded from GEO database. The 44 datasets were selectively chosen from “GSE24943 based on experiment conducted in microarray studies.

Clustering is one of the unsupervised methods; each cluster is a collection of objects which are similar to each other and are dissimilar to the objects belonging to other clusters. The similarity mostly is measured with distance: two or more objects belong to the same cluster if they are close according to a given distance (Bashar and Myaeng, 2009).

Figure 3.5

GSE24943 DATASET RETRIVED FOR CLUSTERING

3.3.3.1 SUPERVISED AND UNSUPERVISED METHOD

Supervised methods are generally used for two purposes are finding genes with expression levels that are significantly different between groups of samples, and finding genes that accurately predict a characteristic of the sample. Unsupervised methods try to find internal structure or relationships in a data set instead of trying to determine how best to predict a 'correct answer'.

Within unsupervised learning, there are three classes of techniques are feature determination, or determining genes with interesting properties without specifically looking for a particular a priori pattern, such as principal-components analysis; cluster determination, or determining groups of genes or samples with similar patterns of gene

expression, such as nearest neighbour clustering, self-organizing maps, k-means clustering and hierarchical clustering (Butte, 2002).

3.3.4 CLUSTERING METHODS

- HIERARCHIAL METHOD
- SELF ORGANISING MAPS
- PRINCIPAL COMPONENT ANALYSIS
- K-MEAN CLUSTERING

3.3.4.1 HIERARCHIAL METHOD

Hierarchical clustering is a commonly used unsupervised technique that builds clusters of genes with similar patterns of expression. This is done by iteratively grouping together genes that are highly correlated in terms of their expression measurements, then continuing the process on the groups themselves. Dendrograms are used to visualize the resultant hierarchical clustering.

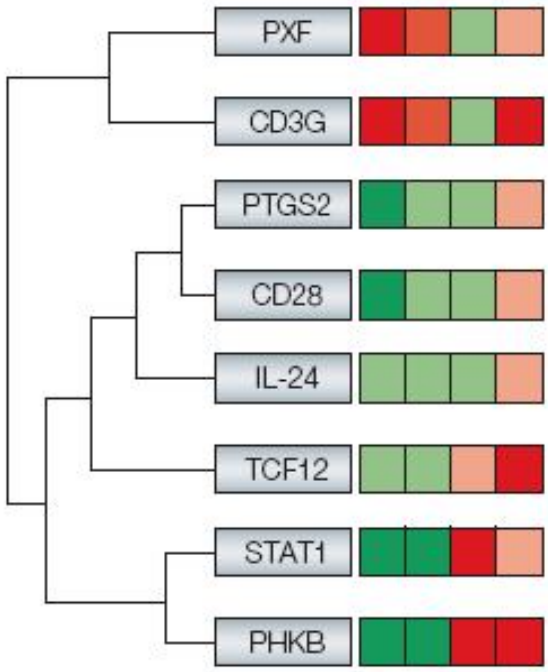
A dendrogram represents all genes as leaves of a large, branching tree. Each branch of the tree links two genes, two branches or one of each. Although construction of the tree is initiated by connecting genes that are most similar to each other, genes added later are connected to the branches that they most resemble. Although each branch links two elements, the overall shape of the tree can sometimes be asymmetric. In visually interpreting dendrograms, it is important to pay attention to the length of the branches. Branches connecting genes or other branches that are similar are drawn with shorter branch lengths. Longer branches represent increasing dissimilarity (Butte, 2002).

A dendrogram represents all genes as leaves of a large, branching tree. The number and size of expression patterns within a data set can be estimated quickly, although the division of the tree into actual clusters is often performed visually. It generally falls into two categories (i.e.) agglomerative and divisive.

Agglomerative is a bottom up approach where each observation starts in its own cluster and pairs of clusters are merged as one moves up the hierarchy. Divisive is a top down approach i.e., all observations start in one cluster and splits are performed recursively as one moves down the hierarchy (Saravanakumar, 2011).

Figure 3.6

HIERARCHICAL CLUSTERING



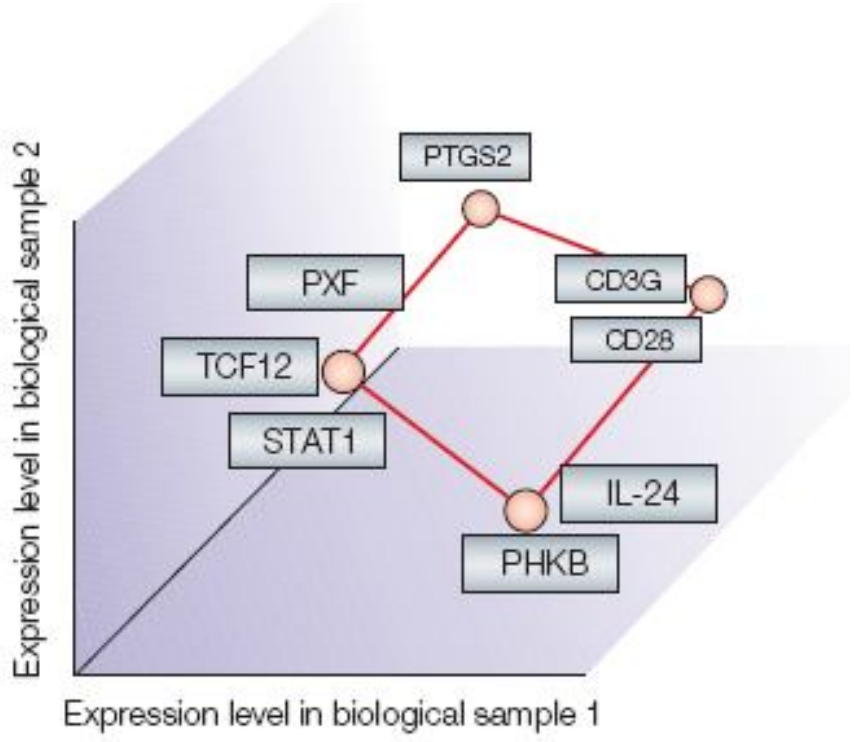
(Source: Butte, 2002)

3.3.4.2 SELF-ORGANIZING MAPS

Self-organizing maps are similar to hierarchical clustering, in that they also provide a survey of expression patterns within a data set, but the approach is quite different (Toronen *et al*., 1999). Genes are first represented as points in multidimensional space. In other words, each biological sample is considered a separate dimension or axis of this space, and after the axes are defined, genes are plotted using expression levels as coordinates. This is easiest to visualize with three or less microarrays, but extends to a larger number of experiments or dimensions. Nearness can be defined using any of the dissimilarity measures described above, although Euclidean distance is most commonly used (Butte, 2002).

Figure 3.7

SELF ORGANISING MAP



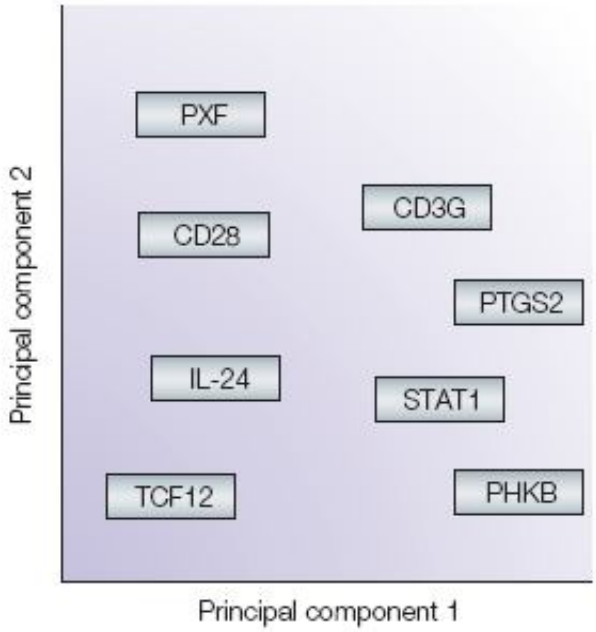
(Source: Butte, 2002)

3.3.4.3 PRINCIPAL COMPONENT ANALYSIS

Principal-components analysis is more useful as a visualization technique than as an analytical method (Alter, 2000). It can be applied to either genes or samples, which are represented as points in multidimensional space, similar to self-organizing maps. Principal components are a set of vectors in this space that decreasingly capture the variation seen in the points. The principal components are linear combinations that include every gene or sample, and the biological significance of these combinations is not directly intuitive (Butte, 2002)

Figure 3.8

PRINCIPAL COMPONENT ANALYSIS



(Source: Butte, 2002)

3.3.4.4 K-MEANS CLUSTERING

K-means clustering is a data mining/machine learning algorithm used to cluster observations into groups of related observations without any prior knowledge of those relationships (Tavazoie *et al.*, 1999). It is one of the simplest clustering techniques and it is commonly used in medical imaging and biometrics.

The K-means clustering algorithm typically uses the Euclidean properties of the vector space. After the initial partitioning of the vector space into K parts, the algorithm calculates the center points in each subspace and adjusts the partition so that each vector is assigned to the cluster the center of which is the closest. This is repeated iteratively until either the partitioning stabilizes or the given number of iterations is exceeded (Brazma and Vilo 2000).

3.3.5 STATISTICAL ANALYSIS TEST

3.3.5.1 SIGNIFICANACE ANALYSIS OF MICROARRAYS

DNA microarray studies may target discovery of individual genes whose expressions are associated with a phenotype. Useful statistical approaches have been proposed for such individual-gene analyses, for example, Significance Analysis of Microarray (SAM) (Tusher *et al*., 2001).

Significance Analysis of Microarrays (SAM) as a statistical technique for finding significant genes in microarray experiments. This techniques aims to control the False Discovery Rate (FDR), which is the proportion falsely rejected null hypotheses among all rejected null hypotheses. The software was developed at Stanford University and is free to academic users. SAM is a convenient MS Excel add-in. The input to SAM is gene expression measurements (the normalized gene expression ratio for each gene) from a set of microarray hybridizations and a response variable from each experiment. SAM works with data from both cDNA and oligo arrays and can also be applied to protein expression and SNP array data (http://www-stat.stanford.edu/~tibs/SAM/).

3.3.5.2 t-TEST

Three t-test designs implemented are one-sample, paired and between-subjects. In the one-sample design, the user specifies a mean. Each gene whose mean log2 expression ratio over all included samples is significantly different from the user-specified mean is assigned to one cluster, while those genes whose means are not significantly different from the user-specified mean are assigned to another cluster. To exclude a sample from the analysis, uncheck the box next to that sample's name in the left pane of the one-sample screen (http://www.tm4.org/mev_manual/test.html).

RESULTS & DISCUSSION

4.0 RESULTS AND DISCUSSION

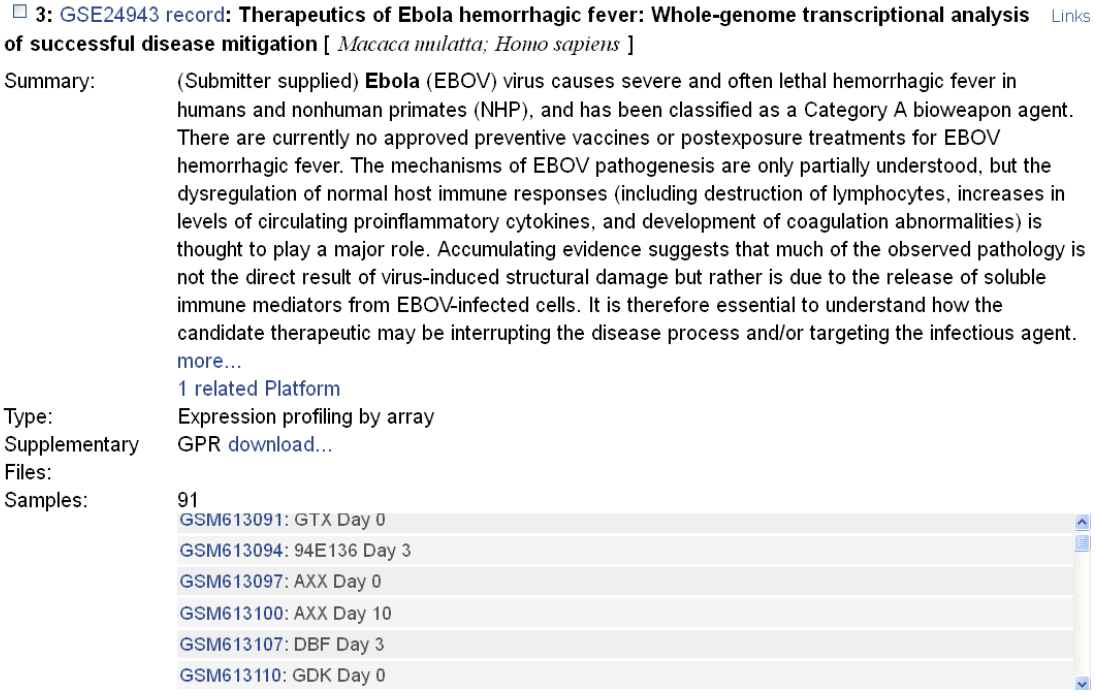
Gene expression microarray is a highly popular technology that allows genome wide measurement of RNA expression levels in a highly quantitative manner. Gene expression data is typically arranged as an $m \times n$ data matrix, with rows corresponding to genes and columns corresponding to experimental conditions (Dharan and Nair, 2009). The *in silico* microarray data analysis was carried out with the online available data set: GSE24943, the dataset was collected from GEO database.

4.1 MICROARRAY DATA VISUALIZATION USING MATLAB

The GenePix GPR format files GO: GSM613097 and GSM613100 contain the data for image analysis. The data is obtained from Blood sample taken day 0 post-infection and day 10 post infection with rNAPc2 treatment of Ebola Hemorrhagic fever was induced using Recombinant nematode anticoagulant protein-2 (rNAPc2). GSM613097 is used as control and GSM613100 is used as Ebola diseased gene.

Figure 4.1

DATASET CHOOSEN FOR ANALYSIS USING MATLAB 7.0.4



☐ 3: GSE24943 record: **Therapeutics of Ebola hemorrhagic fever: Whole-genome transcriptional analysis of successful disease mitigation** [*Macaca mulatta; Homo sapiens*] Links

Summary: (Submitter supplied) **Ebola** (EBOV) virus causes severe and often lethal hemorrhagic fever in humans and nonhuman primates (NHP), and has been classified as a Category A bioweapon agent. There are currently no approved preventive vaccines or postexposure treatments for EBOV hemorrhagic fever. The mechanisms of EBOV pathogenesis are only partially understood, but the dysregulation of normal host immune responses (including destruction of lymphocytes, increases in levels of circulating proinflammatory cytokines, and development of coagulation abnormalities) is thought to play a major role. Accumulating evidence suggests that much of the observed pathology is not the direct result of virus-induced structural damage but rather is due to the release of soluble immune mediators from EBOV-infected cells. It is therefore essential to understand how the candidate therapeutic may be interrupting the disease process and/or targeting the infectious agent. more...

1 related Platform

Type: Expression profiling by array

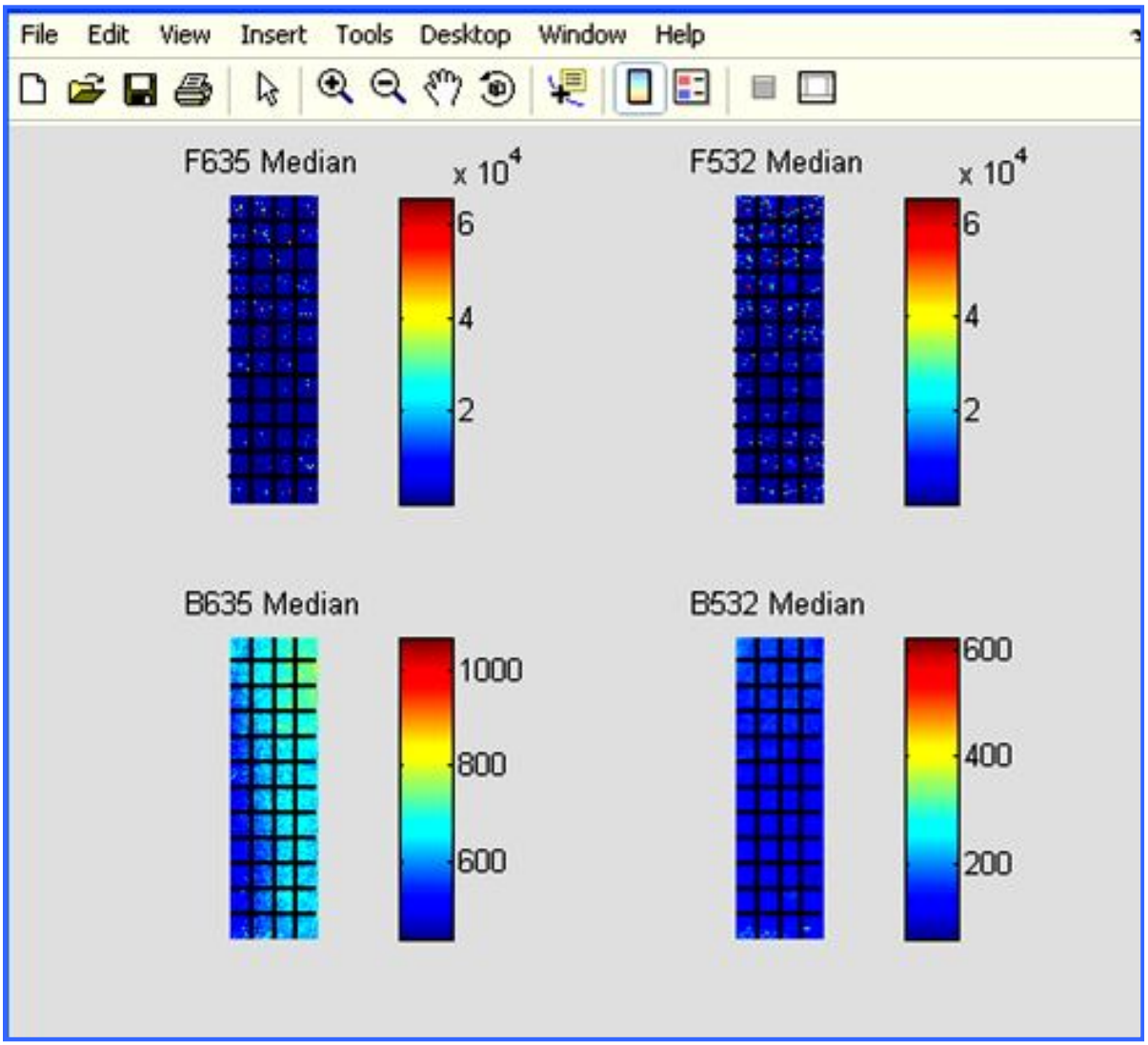
Supplementary Files: GPR download...

Samples: 91

GSM613091: GTX Day 0
GSM613094: 94E136 Day 3
GSM613097: AXX Day 0
GSM613100: AXX Day 10
GSM613107: DBF Day 3
GSM613110: GDK Day 0

FIGURE 4.2

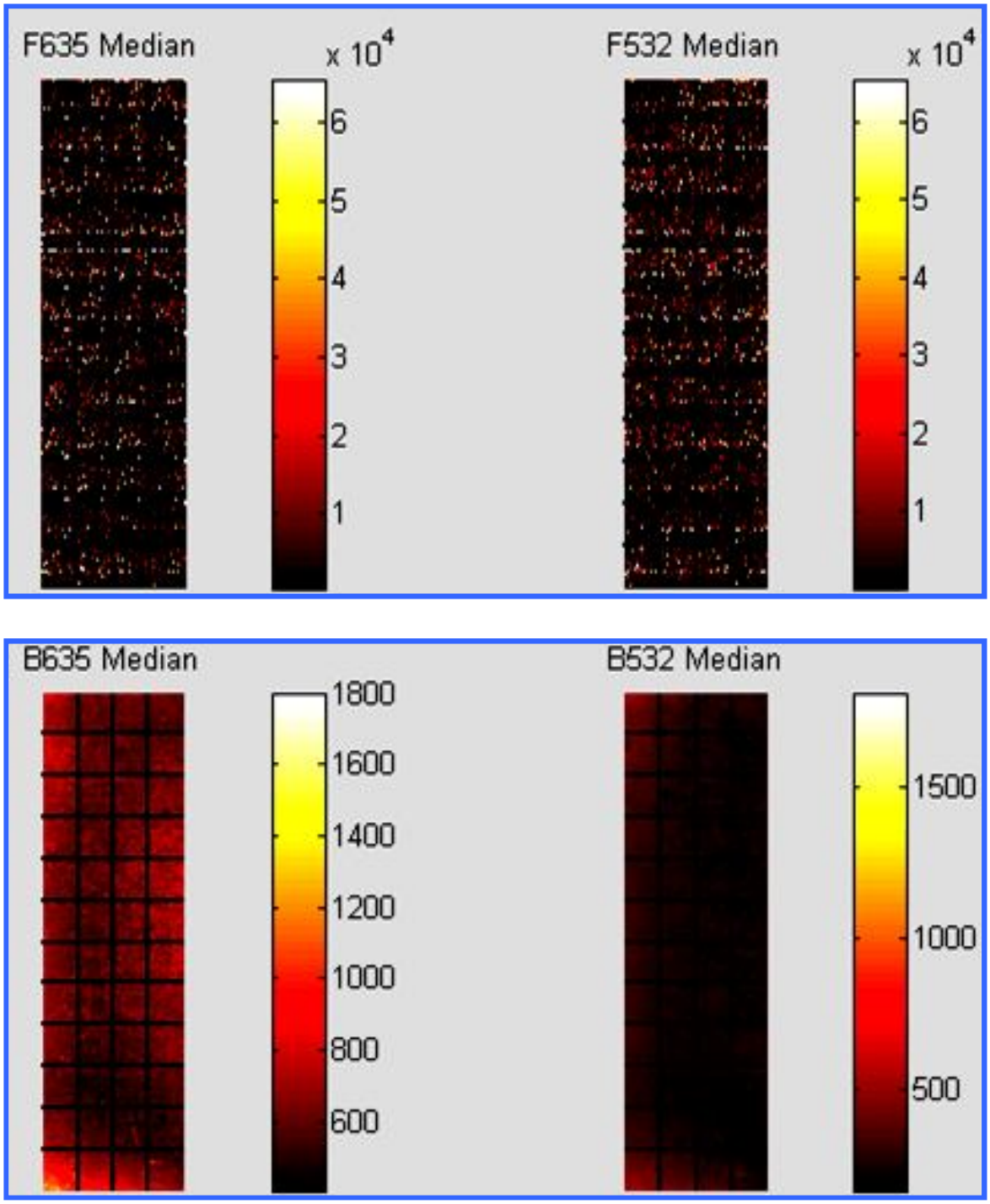
PSEUDOCOLOUR IMAGE OF EBOLA HEMORRHAGIC FEVER GENE WITH FOREGROUND AND BACKGROUND EFFECT



The Ebola Hemorrhagic fever gene expression foreground and Background effect were displayed in Figure 4.2. The "F635 Median" field shows the median pixel values for the foreground of the red (Cy5) channel. The "F532 Median" field corresponds to the foreground of the green (Cy3) channel. The “B635 Median" shows the median values for the background of the red channel. "B532 Median" shows the median values for the background of the green channel (Matlab, 2012).

Figure 4.3

COLORMAP HOT IMAGE OF DISEASED GENE

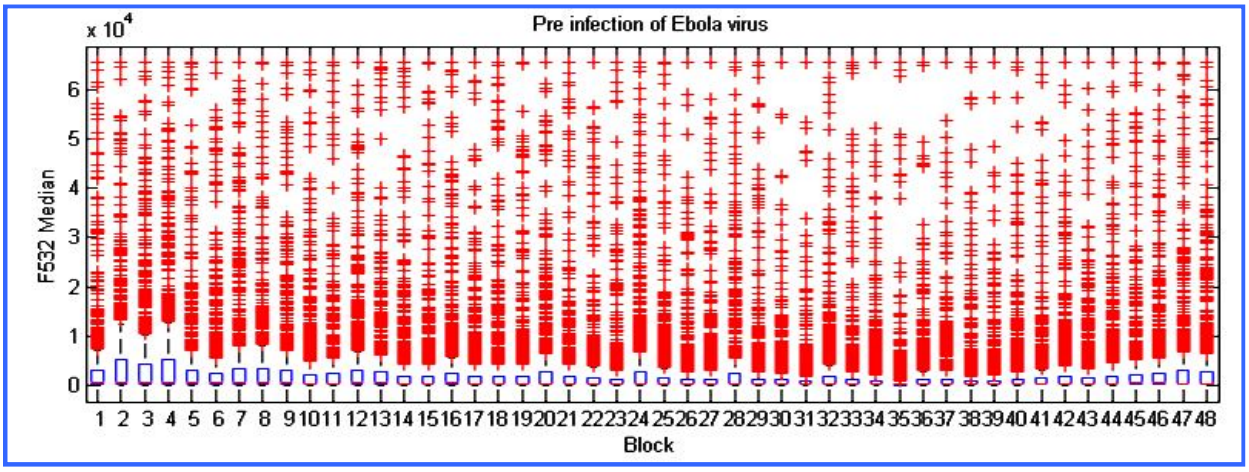


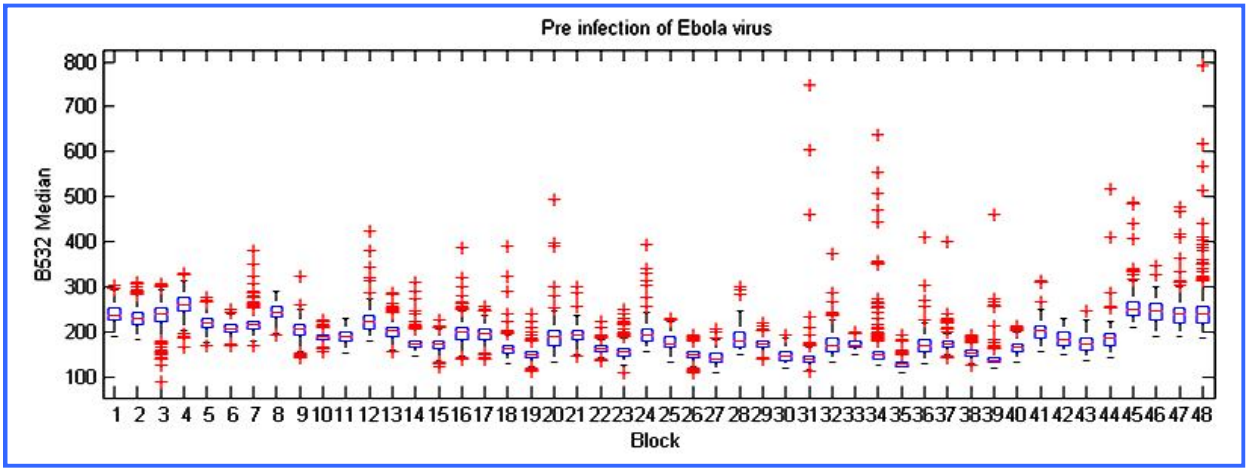
The colormap of Ebola Hemorrhagic fever genes is shown in Figure 4.3. In Figure 4.2 the background colour intensity is more than that of gene expression level. Hence by changing it to colormap option gives more clear view of gene expression.

4.1.1 SATISTICS OF MICROARRAY DATA

Figure 4.4

SPATIAL EFFECTS IN THE FORGROUND AND BACKGROUND INTENSITIES OF PRE INFECTION OF EBOLA VIRUS

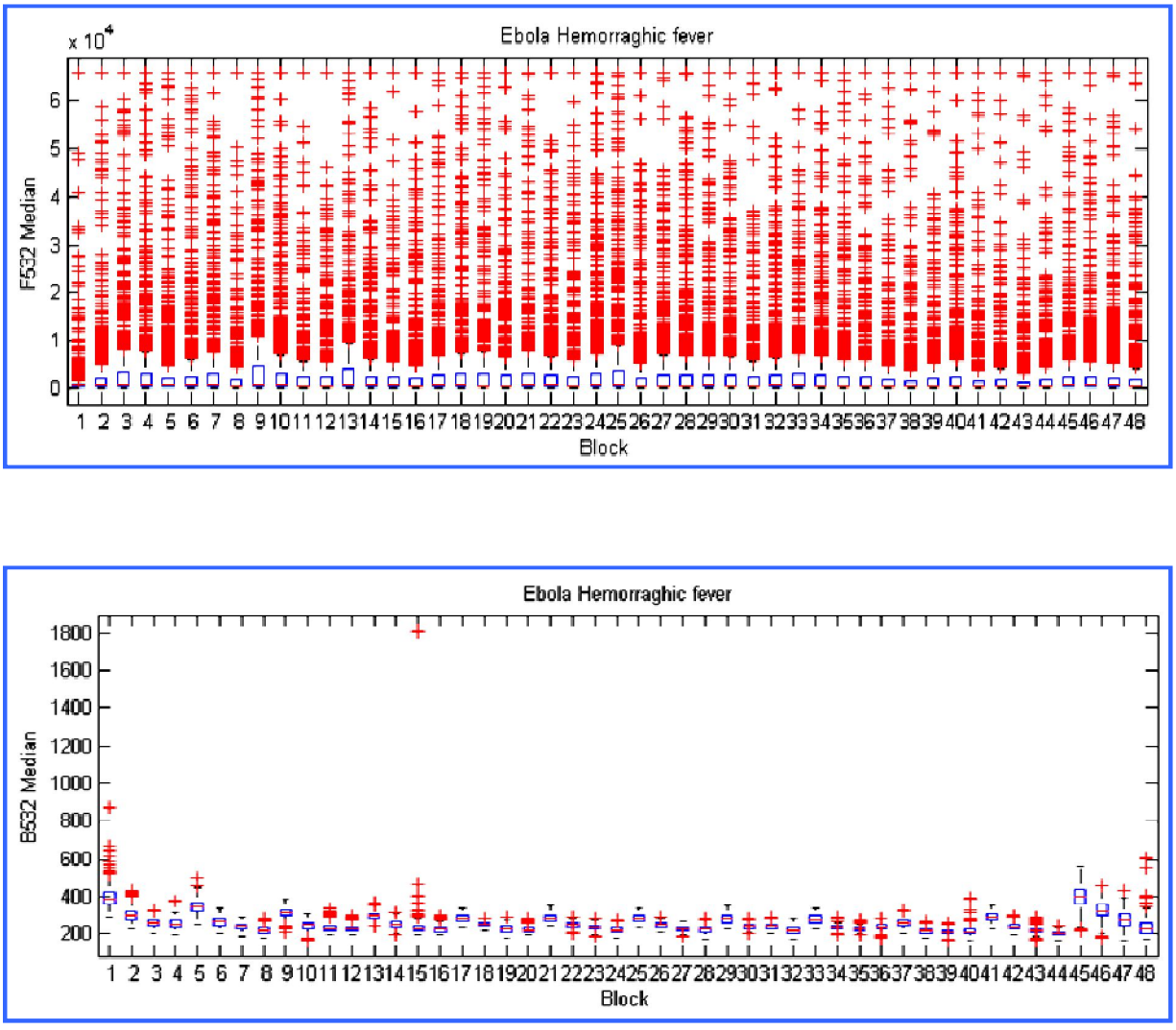




In plot shown in Figure 4.4, the spatial effects of gene expression show the measure of the actual expression levels of Pre-infection of Ebola Hemorrhagic Fever. In B532 Median plot the points are scattered. It is easy to notice the express level of genes.

Figure 4.5

SPATIAL EFFECTS IN THE FORGROUND AND BACKGROUND INTENSITIES OF EBOLA HEMORRHAGIC FEVER

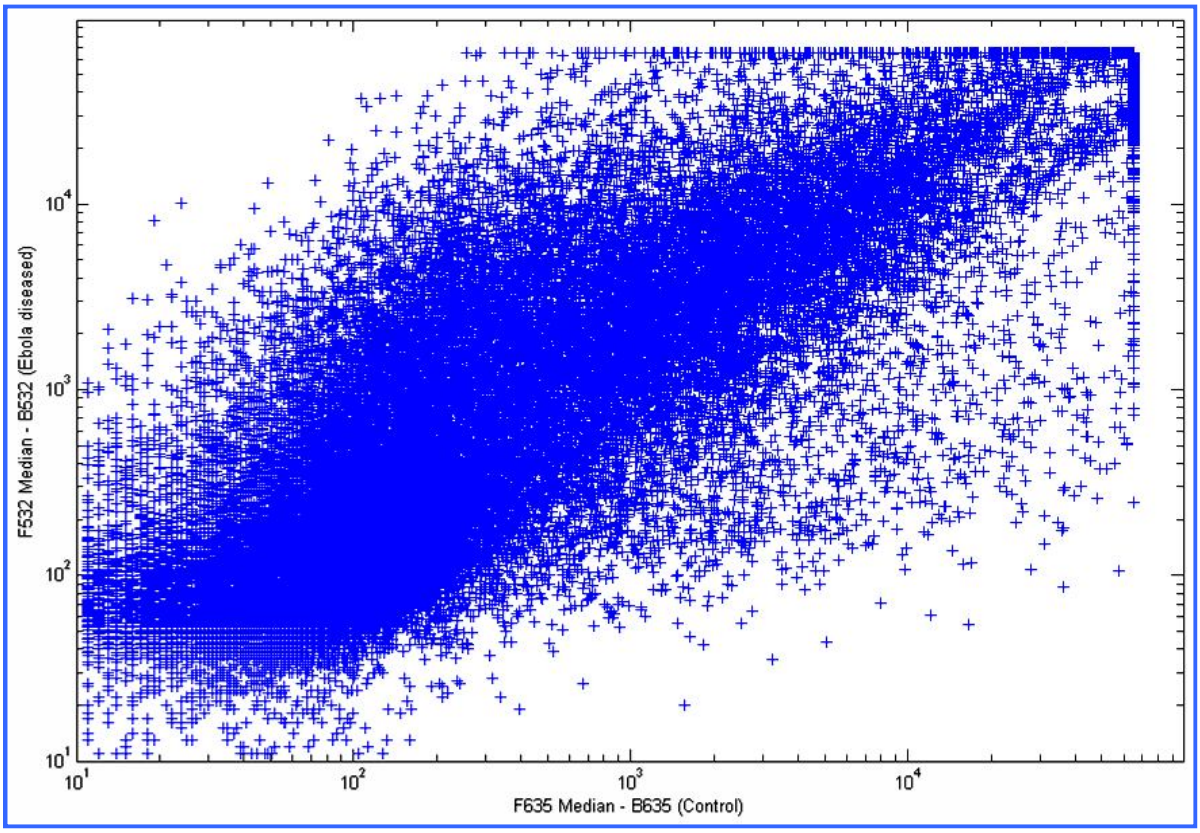


In the plot shown in Figure 4.5, the spatial effects of the actual expression levels of genes involved in Ebola Hemorrhagic Fever. The diseased genes are highly expressed. In F532 Median and B532 Median plots the points are less scattered hence Ebola Hemorrhagic Fever genes are less noticeable.

4.1.2 SCATTER PLOT ANALYSIS OF MICROARRAY DATA

Figure 4.6

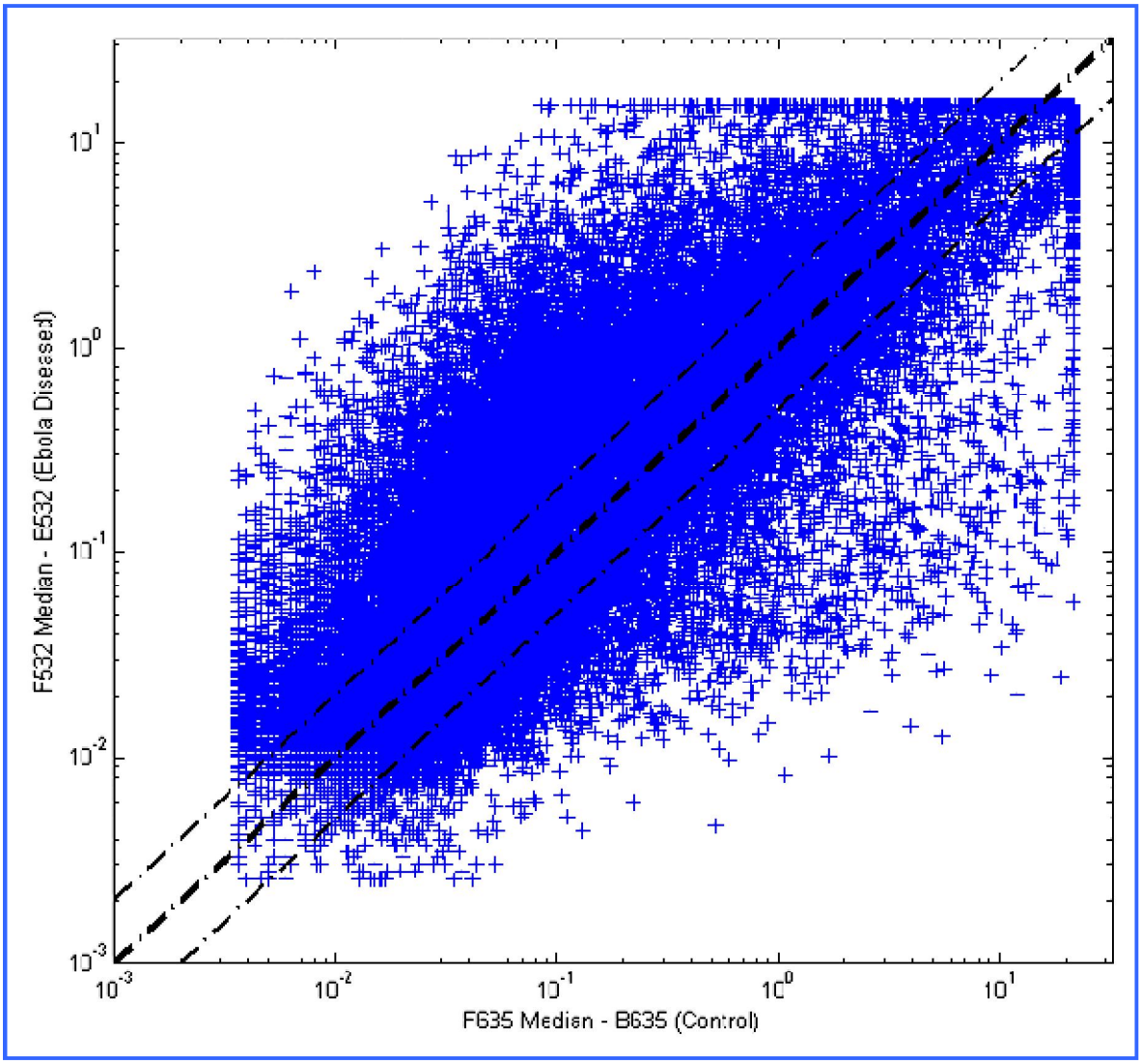
SCATTER PLOT OF EBOLA HEMORRHAGIC FEVER VS CONTROL



The plot shown in Figure 4.6 represents the scatter plot drawn between diseased gene and control gene, using loglog plot function of MATLAB 7.0.4. The plot was constructed before the normalization of the selected dataset. Points that are above the diagonal correspond to genes that have higher expression levels in the Ebola Hemorrhagic Fever condition.

Figure 4.7

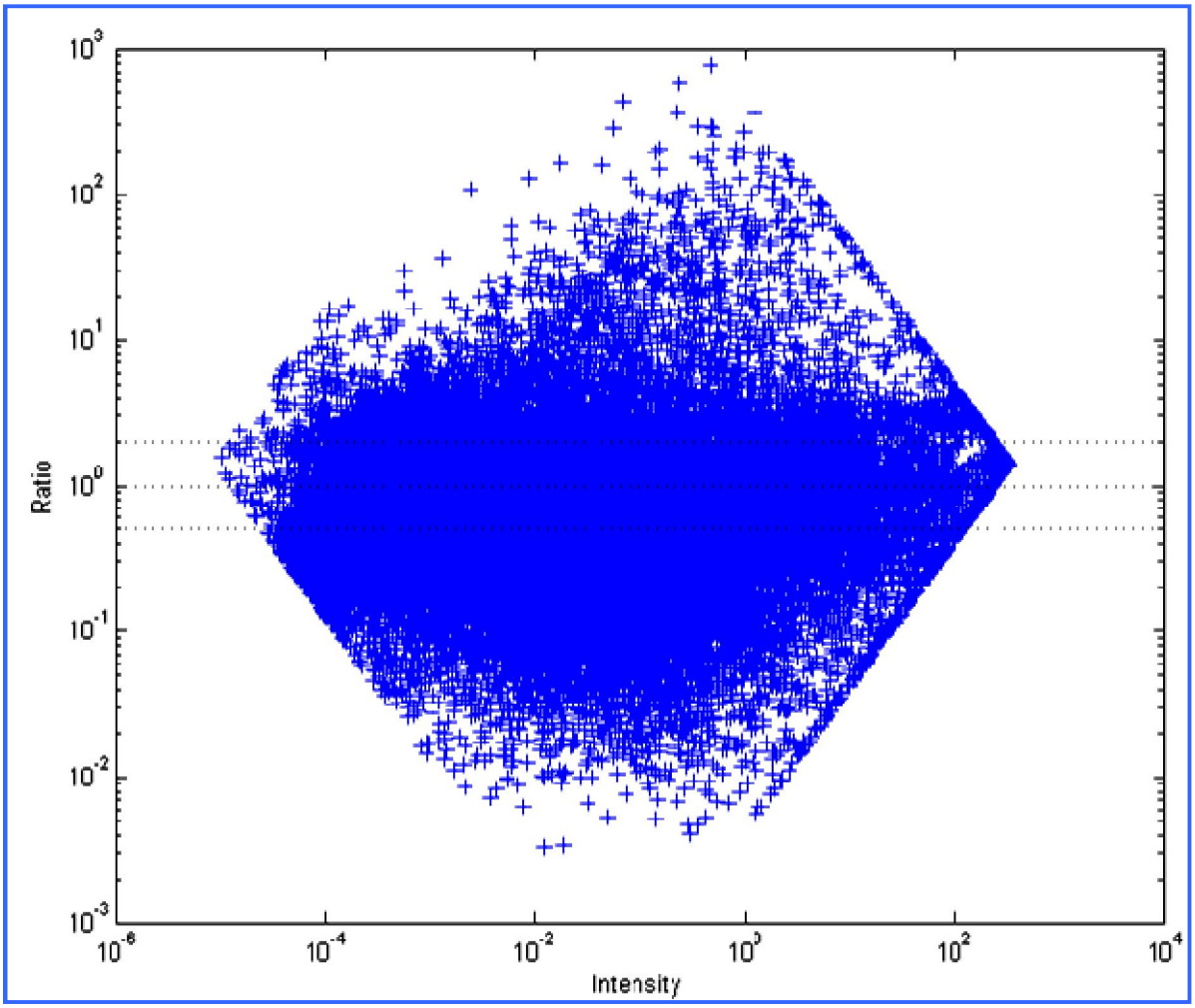
SCATTER PLOT OF EBOLA HEMORRHAGIC FEVER VS CONTROL AFTER NORMALIZATION



The plot shown in Figure 4.7 represents the scatter plot of the normalized data of Ebola diseased genes and Control genes. Normalization of data was done using manorm function in MATLAB 7.0.4. The manorm function was used to perform global mean normalization. The normalized data are more evenly distributed about the y = x line.

Figure 4.8

SCATTER PLOT FOR INTENSITY VS RATIO



The plot shown in Figure 4.8 represents the scatter plot drawn using the function mairplot to generate Intensity vs. Ratio plot for the normalized data using MATLAB 7.0.4. The mairplot function was used to perform Lowess normalization on the data. The microarray data were normalized and gene expression level of diseased and control genes were examined.

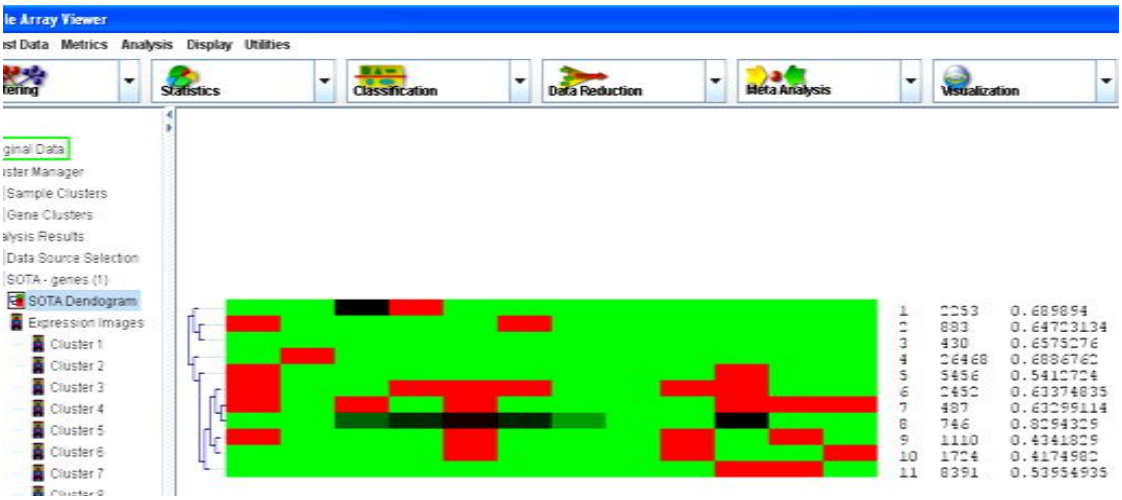
4.2 CLUSTERING GENE EXPRESSION PROFILES

Gene expression microarrays provide a popular technique to monitor the relative expression of thousands of genes under a variety of experimental conditions. Cluster analysis partitions gene data into meaningful clusters which capture the natural structure of the data to find genes that have similar functionality (Pati and Das, 2011). There are three common types of clustering methods such as hierarchical clustering, k-means clustering and self-organizing maps.

The gene expression data dataset GSE24943, used for cluster analysis, was downloaded from GEO database. Then it was imported into MeV and analyzed using various tools in it.

Figure 4.9

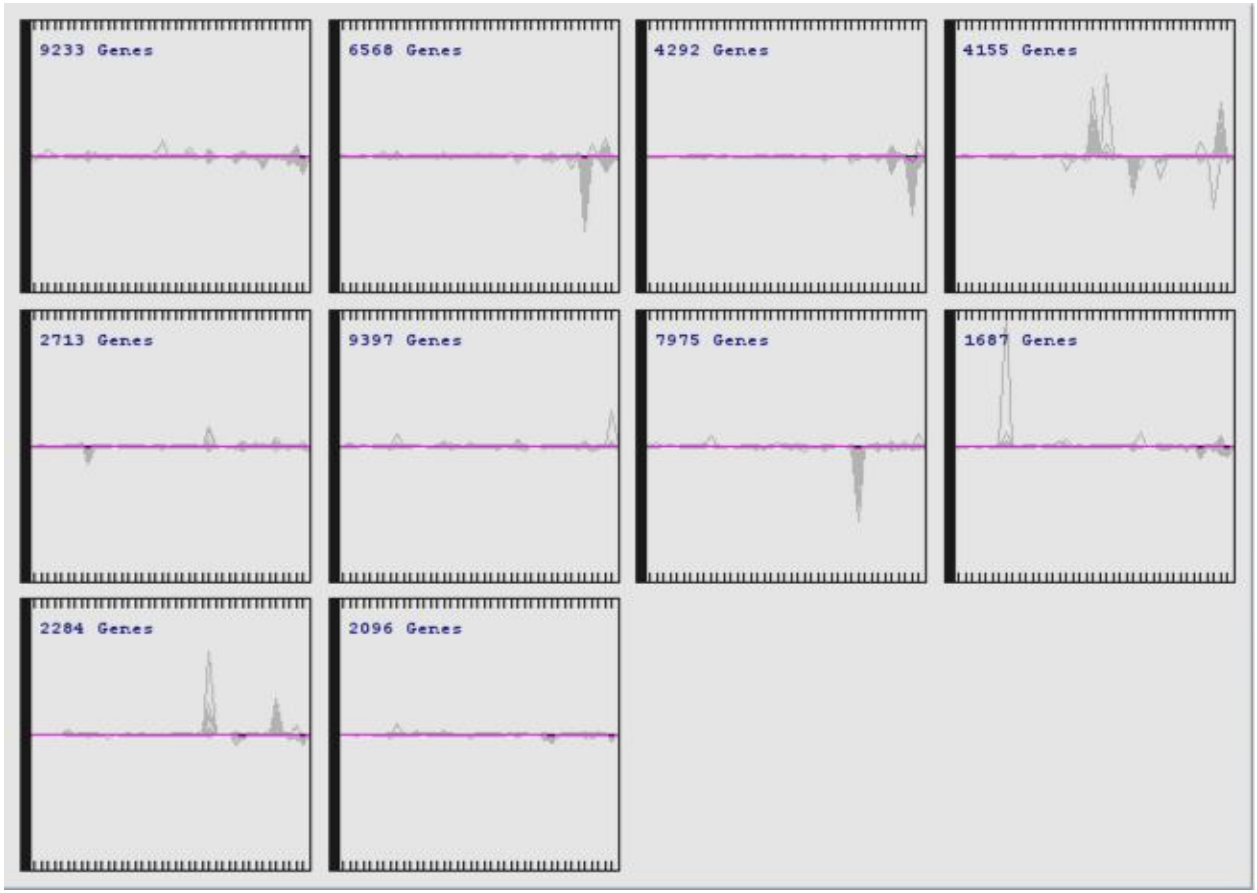
SOTA DENDOGRAM



SOTA dendrogram shown in Figure 4.9 displays the tree generated with the expression image of each resulting cluster's centroid gene. The cluster id number, the cluster population (number of genes in the cluster), and the cluster diversity (mean gene to centroid distance) are shown at the right end of the Figure 4.9 respectively.

Figure 4.10

GENE CLUSTERS GRAPH USING KMC



The genes with similar expression patterns were clustered using K-Means clustering. Among 50400 genes 10 clusters were formed with different set of genes and they are represented in Figure 4.10.

Figure 4.11

METAGENE GRAPGH USING NMF

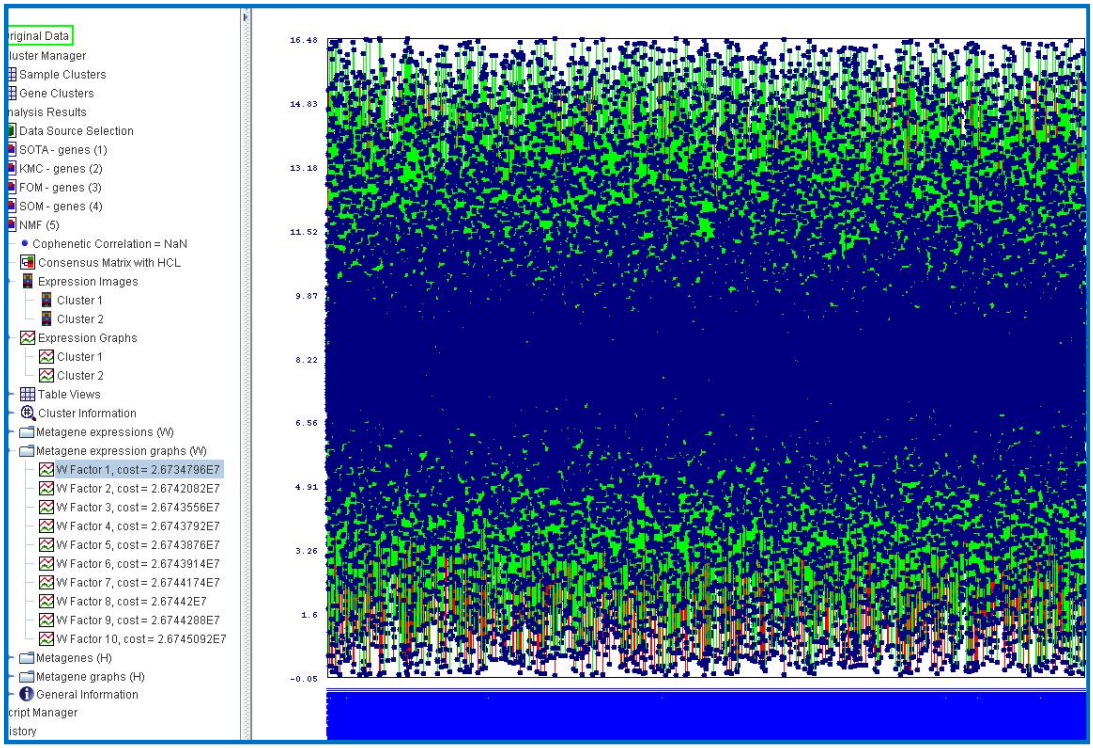
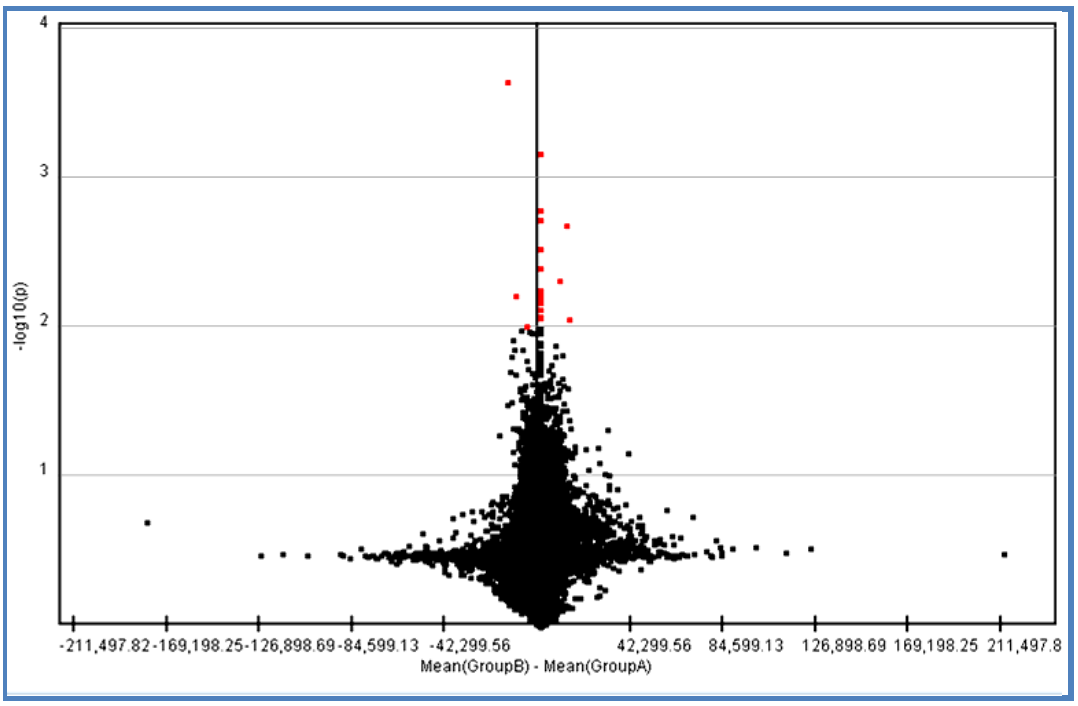


Figure 4.12

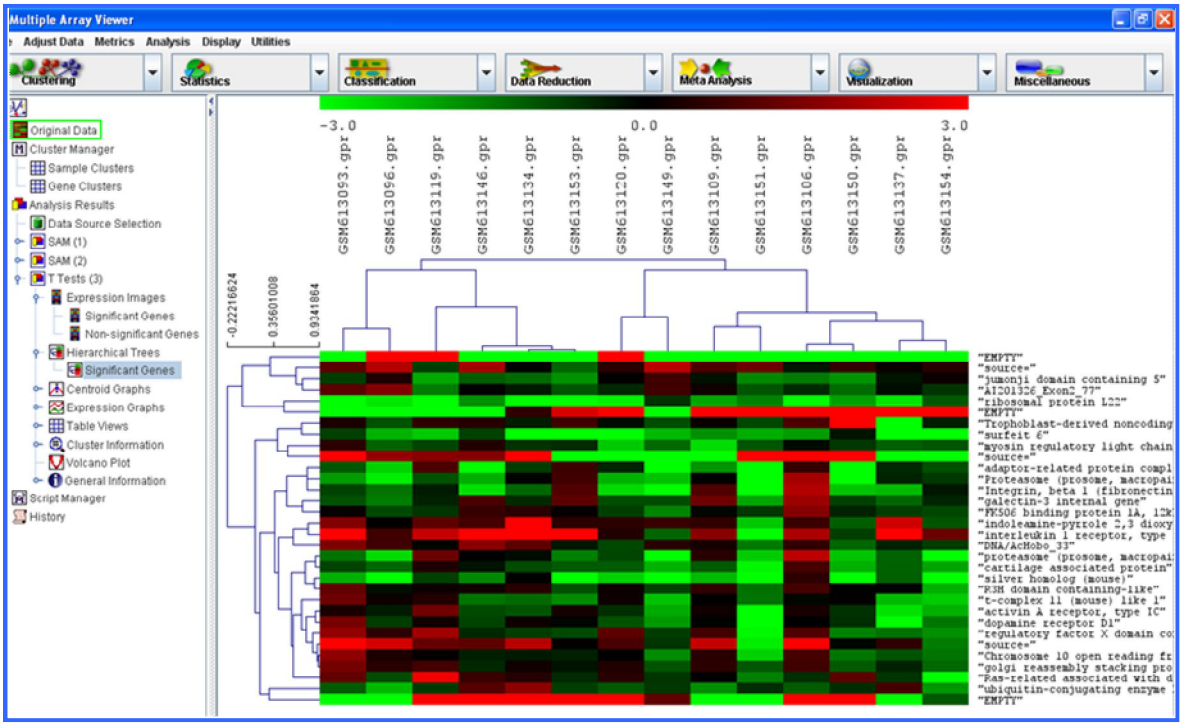
VOLCANO PLOT USING t-TEST


4
3
2
1
-log10(p)
-211,497.82
-169,198.25
-126,898.69
-84,599.13
-42,299.56
42,299.56
84,599.13
126,898.69
169,198.25
211,497.8
Mean(GroupB) - Mean(GroupA)

When a large number of genes are significant in correlating microarray gene expression data with patient prognosis, clustering of significant genes may be effective not only for further dimension reduction but also for identifying co-regulated genes that belong to the same molecular pathway related to disease biology and aggressiveness (Matsui *et al*., 2008).

Figure 4.13

HIERARCHICAL CLUSTERING OF SIGNIFICANT GENES USING t-TEST



In Figure 4.13, the significant genes with similar expression pattern were hierarchically clustered using t-TEST (among 44 sample datasets) between Ebola diseased and Control containing 50,400 genes, based on Pearson correlation.

4.3 SIGNIFICANCE ANALYSIS OF MICROARRAY

For each area of interest, the microarray dataset were analyzed to detect transcriptional alterations between the EHF and control using SAM. SAM generates an interactive plot of the observed vs. expressed (based on the permuted data) delta-values was shown in Figure 4.14. Out of 50400 genes in the dataset chosen, 9 genes were found

to be significantly changed in EHF condition. In the majority of the significant genes 3 were upregulated in EHF, whereas 6 genes were down regulated.

Delta is a vertical distance (in graph units) from the solid line of slope 1(i.e., where observed=expected). The two dotted lines represent the region within + or – delta units from the "observed=expected" line. The genes whose plot values are represented by black dots are considered non-significant, those coloured red are positive significant and the green ones are negative significant.

Figure 4.14

GRAPH SHOWING UPREGULATED AND DOWNREGULATED GENES IN EBOLA HEMORRHAGIC FEVER USING SAM

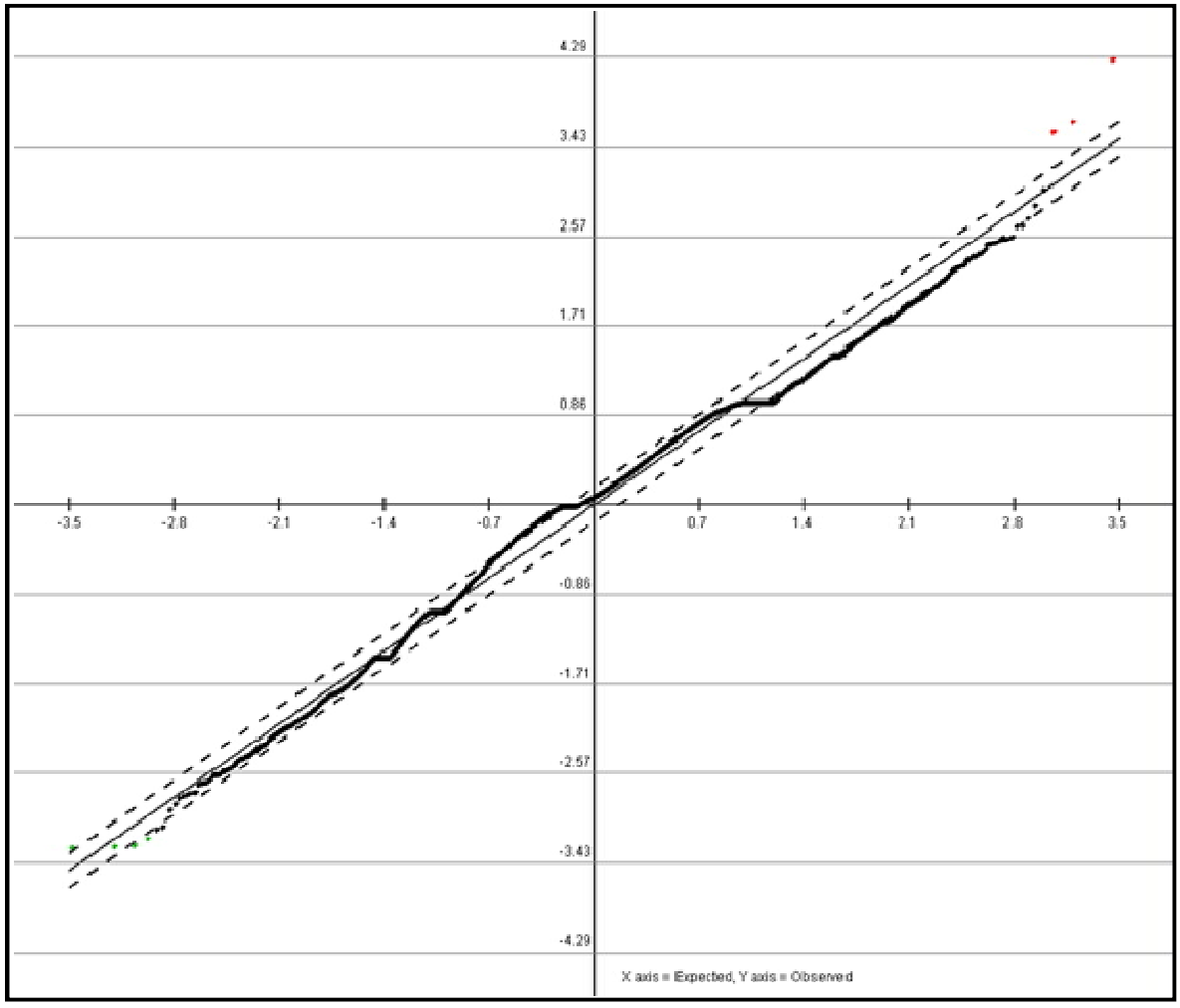


Table 4.1

DELTA TABLE FOR SIGNIFICANT GENES

A maximum of 9 significant genes were observed in the delta value ranges from 0.12-0.160 respectively (Table 4.1)

Delta value	Median False Value	Number of significant Genes
0.124-0.160	3.0	9

The present analysis of microarray dataset identified dysregulations in transcriptional levels of several genes previously not implicated in Ebola Hemorrhagic fever, which have a role in Ebola Hemorrhagic fever. The genes were analyzed for their biological function and role in EHF.

Figure 4.15

LIST OF SIGNIFICANT GENES IN EBOLA HEMORRHAGIC FEVER USING t-TEST

Expression li...	Name	ID	Absolute t va...	Degrees of fr...	Raw p value	Adj p value	False Disco...
	"3-hydroxy-3...	"hCT001237"	-2.8487375	21.0	0.009616305	0.009616305	13.098967
	"Sortilin-rela...	"hHR003817"	2.9654462	21.0	0.007381494	0.007381494	16.910332
	"source="	"hHR012169"	2.8797417	21.0	0.00896642	0.00896642	13.291399
	"ADP-ribosy...	"hHR015733"	3.6133506	21.0	0.00163116...	0.00163116...	11.744353
	"source="	"hHR027637"	3.184148	21.0	0.004464662	0.004464662	16.072783
	"zinc finger ...	"hHC001789"	-2.9816456	21.0	0.00711382...	0.00711382...	17.073183
	"FK506 bind...	"hHC002941"	-2.885086	21.0	0.008858727	0.008858727	13.529693
	"interleukin ...	"hHC010513"	2.9315078	21.0	0.007974	0.007974	14.353199
	"Trophoblas...	"hHR002881"	2.9055877	21.0	0.00845698	0.00845698	13.319744
	"3end_LINE...	"hCX048589"	3.0087771	21.0	0.006686302	0.006686302	16.84948
	"EMPTY"	"EMPTY"	-3.2050915	20.0	0.004444262	0.004444262	17.230062
	"EMPTY"	"EMPTY"	-3.4715705	21.0	0.002280124	0.002280124	14.364781
	"Kruppel-lik...	"hHR027171"	3.092512	21.0	0.00551736	0.00551736	17.379684
	"CD47 antig...	"hHC014476"	-2.9508271	21.0	0.007631341	0.007631341	16.025816
	"EMPTY"	"EMPTY"	5.041301	21.0	5.4371674E...	5.4371674E...	2.7403326
	"Hypothetic...	"hHR002657"	-3.0907793	21.0	0.005539413	0.005539413	16.422731
	"ubiquitin C"	"hHC023837"	3.911748	21.0	8.021258E-4	8.021258E-4	10.106785
	"receptor-int...	"hHC009989"	3.2718263	21.0	0.00364147...	0.00364147...	16.684565
	"neutrophil ...	"hHC032273"	2.9514658	21.0	0.007620258	0.007620258	16.698305
	"SINE/Alu_6...	"hCX048677"	4.10323	21.0	5.076487E-4	5.076487E-4	8.528498
	"Actin, beta"	"hCP048894"	4.7009764	21.0	1.2178264E...	1.2178264E...	3.0689228
	"phosphoin...	"hHA038226"	2.8699703	21.0	0.009166566	0.009166566	13.199854
	"myelin expr...	"hHC007951"	-2.9326859	21.0	0.007952687	0.007952687	14.8450165
	"EMPTY"	"EMPTY"	-2.839339	21.0	0.009822034	0.009822034	12.69309
	"zinc finger, ...	"hHC007459"	-2.8455198	21.0	0.00968627...	0.00968627...	12.847057

Table 4.2

LIST OF SIGNIFICANT GENES USING t-TEST AND SAM

S.No	Gene symbol	Gene ID	Gene name	Gene Ontology			Function
				Molecular function	Biological function	Cellular component	
1	HMGCR	hCT001237	3-hydroxy-3 methylglutaryl-Coenzyme A reductase	Hydroxymethylglutaryl-CoA reductase (NADPH) activity	Cholesterol biosynthetic process	Peroxisome	This is a transmembrane glycoprotein is involved in the control of cholesterol biosynthesis.
2	SORCS1	hHR003817	Sortilin-related VPS10 domain containing receptor 1	Protein binding	Neuropeptide signaling pathway	Membrane	This gene is strongly expressed in the central nervous system.
3	ARF1	hHR015733	ADP-ribosylation factor 1	Nucleotide binding	GTP catabolic process	Golgi membrane	Central role in intra-Golgi transport.
4	ZFY	hHC001789	Zinc finger protein, Y-linked	DNA binding	Regulation of transcription, DNA-	Intracellular	It acts as transcription factor.

					dependent		
5	FKBP14	hHC002941	FK506 binding protein 14, 22 k Da	Calcium ion binding	Protein folding	Endoplasmic reticulum	-
6	IL6	hHC010513	Interleukin 6 (interferon, beta 2)	Receptor binding	Neutrophil apoptosis	Interleukin-6 receptor complex	This gene encodes a cytokine that functions in inflammation and the maturation of B cells.
7	KLF13	hHR027171	Kruppel-like factor 13	DNA binding	Regulation of transcription from RNA polymerase II promoter	Intracellular	KLF13 belongs to a family of transcription factors. These transcription factors bind to GC-rich sequences and related GT and CACCC boxes.
8	UBC	hHC023837	Ubiquitin C	Ubiquitin-protein ligase activity	Cell cycle checkpoint	Nucleoplasm	Ubiquitination has been associated with protein degradation, DNA repair, cell cycle regulation, kinase modification, endocytosis, and regulation of other cell signaling pathways.
9	RIPK2	hHC009989	Receptor-interacting serine-threonine	Nucleotide binding	Activation of MAPK activity	Intracellular	It is a potent activator of NF-kappaB and inducer of apoptosis in response to various stimuli.

			kinase 2				
10	NCF1	hHC032273	Neutrophil cytosolic factor1	GTPase activity	Leukocyte mediated cytotoxicity	Cytoplasm	Mutations in this gene have been associated with chronic granulomatous disease
11	SOBP	hCX048677	SINE/Alu_61	Metal ion binding	Locomotory behavior	-	The protein encoded by this gene is a nuclear zinc finger protein that is involved in development of the cochlea.
12	PIK3R1	hHA038226	phosphoinositide-3-kinase, regulatory subunit 1	Receptor binding	Protein phosphorylation	Intracellular	Phosphatidylinositol 3-kinase plays an important role in the metabolic actions of insulin, and a mutation in this gene has been associated with insulin resistance.
13	MYEF2	hHC007951	Myelin expression factor 2	Nucleotide binding	Transcription, DNA-dependent	Nucleus	-

The above table 4.2 represents the Gene ontology of all significant genes.

Table 4.3

LIST OF GENES THAT ARE UPREGULATED AND DOWNREGULATED USING SAM

S.No	Regulation Status of Genes	Name of the Genes
1	Upregulated genes	HMGCR IL6 SORCS1
2	Downregulated genes	ZFY UBC RIPK2 NCF1 PIK3R1 ARF1

The transcriptional changes of the gene expression were identified in the earlier findings in Ebola Hemorrhagic Fever and they include GADD45, DUSP2, IDO, ISG 45K, IL-10, TLR3, GNA13, COX-2, GCH1, GM-CSF, MIP-3α, t-Pa, GADD45A, and IDO (Jensen *et al.,* 2011). The IL-6 upregulated gene and PIK3R1, ARF1 downregulated gene was found to be mainly involved in Ebola Hemorrhagic Fever. Alteration in these genes leads to Ebola Hemorrhagic Fever development.

4.4 ROLE OF ADP-RIBOSYLATION FACTOR 1 IN EBOLA HEMORRHAGIC FEVER

ADP-ribosylation factor 1 (ARF1) is a member of the human ARF gene family. The gene products, including 6 ARF proteins and 11 ARF-like proteins, constitute a family of the RAS superfamily. The ARF proteins are categorized as class I (ARF1, ARF2 and ARF3), class II (ARF4 and ARF5) and class III (ARF6), and members of each class share a common gene organization. The ARF1 protein is localized to the Golgi apparatus and has a central role in intra-Golgi transport (http://www.ncbi. nlm.nih.gov/gene/375).

In *Ebolavirus*, ARF1 executes a critical regulatory role in COPI assembly and the regulation of vesicle transport between the ER and Golgi compartment; its overexpression may disturb the balance between anterograde and retrograde ER-to-Golgi compartment transport and thereby interfere with the efficient formation of *Ebolavirus* virus-like particles (VLP)s. Coexpression of matrix proteinVP40 with the dominant negative ARF1 mutant or the constitutively active variant, which is known to affect membrane trafficking, further reduced the amounts of *Ebolavirus* VP40-induced VLPs. Consistent with these findings, overexpression of wild-type or mutant ARF1 reduced the number of VP40-induced cell protrusions relative to control cells (Yamayoshi *et al.*, 2010). Gene expression studies done were predicted to be good compared with the study of Yamayoshi *et al.* (2010). Hence, ARF 1 gene was found to be downregulated during the Ebola Hemorrhagic Fever condition.

4.5 ROLE OF INTERLEUKIN 6 IN EBOLA HEMORRHAGIC FEVER

Interleukin 6 gene encodes a cytokine that functions in inflammation and the maturation of B cells. In addition, the encoded protein has been shown to be an endogenous pyrogen, capable of inducing fever in people with autoimmune diseases or infections. The protein is primarily produced at sites of acute and chronic inflammation, where it is secreted into the serum and induces a transcriptional inflammatory response through interleukin-6 receptor, alpha (http://www.ncbi.nlm.nih.gov/gene/3569). In the dataset chosen, Ebola Hemorrhagic Fever IL-6 gene was found to be upregulated. Hence it is clear that Interleukin 6 plays important role in control of the severity of disease.

The changes in cytokine response are consistent with findings from other groups that analyzed the transcriptional response to other negative-sense RNA viruses and demonstrated transcriptional upregulation of cytokines, such as IL-6, IL-10, and IL-15 that are also involved in T and B cell activation (Basler *et al*., 2000).

4.6 ROLE OF PHOSPHOINOSITIDE-3-KINASE, REGULATORY SUBUNIT 1 IN EBOLA HEMORRHAGIC FEVER

The Phosphoinositide-3-Kinase, a regulatory subunit 1, was found to be downregulated. It plays an important role in cell signaling pathway that regulates diverse

cellular activities including proliferation, differentiation, apoptosis, migration, metabolism, and vesicular trafficking (Saeed *et al*., 2008).

The PI3K pathway is vital for regulation of diverse cellular activities, including growth, survival, differentiation and motility. The finding that the PI3K pathway is also essential for entry of ZEBOV is therefore highly relevant for design of new therapeutic strategies and provides new potential opportunities where PI3K inhibitors developed for treatment ZEBOV infection (Saeed *et al*., 2008).

Thus, Phosphoinositide-3-Kinase, Regulatory Subunit 1 inhibitors can be produced in order to inhibit infection by ZEBOV at an early step during the replication cycle. It can be used as potent target for drug development.

SUMMARY & CONCLUSION

5.0 SUMMARY AND DISCUSSION

Ebola hemorrhagic fever is a severe, often-fatal disease in humans and nonhuman primates (monkeys, gorillas, and chimpanzees). *Zaire ebolavirus* was first identified in 1976 and was found to be the most virulent, killing approximately up to 90% of infected individuals. It is one of the most deadly diseases in world wide.

Microarray is a high-throughput technique to analyze the expression levels of thousands of genes simultaneously between control and diseased samples. The dataset chosen for *in silico* studies of Ebola Hemorrhagic Fever was GSE24943 –"Therapeutics of Ebola Hemorrhagic Fever: Whole-Genome Transcriptional Analysis of Successful Disease Mitigation". The dataset was downloaded from GEO and imported into MATLAB 7.0.4 to study gene expression level intensity between the diseased gene and Control gene.

Clustering was done using various modules in MeV. The K-Mean Clustering and Hierarchical clustering were done to analyze the samples with similar expression pattern. A dendogram was drawn using SOTA, a module in MeV. A tree view of clustered data was obtained. The data were then imported to t-TEST and SAM in order to predict significant differentially expressed genes. A total of 9 significant genes were obtained, out of 50,400 genes of the selected dataset. 3 out of 9 genes were found to be upregulated and 6 genes were downregulated. One of the main advantages of Global expression profiling is the ability to detect concerted changes in functionally related groups of genes. Finally, Gene Ontology of Significant genes was predicted using INNATEDB.

Three genes were found to be, involved in the cause of Ebola Hemorrhagic Fever are ADP-ribosylation factor 1 (ARF1) involved in the vesicle transport between the ER and Golgi compartment; Interleukin 6 (IL-6) involve in immune response and Phosphoinositide-3-Kinase, a Regulatory Subunit 1(PIK3R1) function in cell Signaling Pathway.

Further study can be carried out by understanding the mechanism of the pathways involving ARF1, IL-6 and PIK3R1 genes in both control and diseased genes. The proteins responsible for the cause of the disease can also be targeted with natural antiviral compounds.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Alter, O., Brown, P. O and Botstein, D (2000), Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl Acad. Sci. USA, (*97):10101–10106.

- Babu, M (2004), Introduction to Microarray Data Analysis, *Computational Genomics*: 225-249.

- Bashar, A and Myaeng, S (2009), "Initializing K-Means using Genetic Algorithms", *World Academy of Science*, Engineering and Technology: 54-60.

- Basler, C.F and Amarasinghe, G.K (2009), Evasion of interferon responses by Ebola and Marburg viruses. *J Interferon Cytokine Res*; 29:511-20.

- Basler, C.F., Wang, X., Muhlberger, E., Volchkov, V and Paragas J (2000), The Ebola virus VP35 protein functions as a type I IFN antagonist. *Proc Natl Acad Sci* U S A 97: 12289–12294.

- Bell (2007), Viral hemorrhagic fevers. In: L Goldman, and D Ausiello (eds.), *CecilMedicine, Saunders Elsevier, Philadelphia*: 1-20.

- Bosio, C. M., Aman, M. J., Grogan, C., Hogan, R., Ruthel and G., Negley, D (2003), Ebola and Marburg viruses replicate in monocyte-derived dendritic cells without inducing the production of cytokines and full maturation, *J. Infect. Dis.*, *188*(11):1630–1638.

- Bray, M and Geisbert, T.W (2005), Ebola virus: the role of macrophages and dendritic cells in the pathogenesis of Ebola hemorrhagic fever. *Int J Biochem Cell Biol*, 37:1560-1566.

- Bray, M and Murphy, F.A (2007), Filovirus research: knowledge expands to meet a growing threat. *J Infect Dis*; 196:438-443.

- Brazma A and Vilo J (2000), Gene expression data analysis, *FEBS Lett*, 480(1):17-24.

- Butte, A 2002, The Use and Analysis Of microarray Data, Drug Discovery, *Nature Publishing Group*, (1):951-960.

- Bwaka, M. A., Bonnet, M. J., Calain, P., Colebunders, R.., Roo, A.D and Guimard, Y (1999), Ebola hemorrhagic fever in Kikwit, Democratic Republic of the Congo: Clinical observations in 103 patients, *J. Infect. Dis.*, *179*, S1–S7.

- CDC (2010), Ebola Hemorrhagic Fever Information Packet, U.S. Department of Health and Human Services: 1-12.

- CFSPH (2009), Viral Hemorrhagic Fevers–Ebola and Marburg: 5-9.

- Chandrasekhar, T., Thangavel, K and Elayaraja, E (2011), Effective Clustering Algorithms for Gene Expression Data, *International Journal of Computer Applications*, 32(4):25-29.

- Chen, G., Banerjee, N., Jaradat, S.A., Tanaka, T.S., Ko, M.S.H. and Zhang, M.Q (2002), Evaluation and Comparison of Clustering Algorithms in Analyzing ES Cell Gene Expression Data, *Statistica Sinica,* 12: 241-262

- Chen, T.S., Tsai, T.H., Zen, Y.T., Lin, C.C., Chen, R.C., Li, S.Y and Chen, H.Y (2005), "A combined k-means and hierarchical clustering method for improving the clustering efficiency of microarray", Proceedings of international symposium on intelligent signal processing and communication systems, *IEEE:256-275.*

- Colebunders, R and Borchert, M (2000), Ebola haemorrhagic fever-a review. *J Infect*, 40: 16-20.

- Dharan, S and Nair, A.S (2009), Biclustering of gene expression data using reactive greedy randomized adaptive search procedure, *BMC Bioinformatics*: 1-10.

- Dube, D., Brecher, M.B., Delos,S.E., Rose,S.C., Park,E.W., Schorn- berg,K.L., Kuhn,J.H and White, J.M (2009), The primed Ebolavirus glycoprotein(19-kilo dalton GP1,2): sequence and residues critical for host cell binding. *J. Virol.* 83, 2883–2891.

- Edgar, R., Domrachev, M. and Lash, A.E (2002), Gene ExpressionOmnibus: NCBI gene expression and hybridization array datarepository, Nucleic Acids Res., 30, 207–210.

- Eisen, M.B., Spellman, P.T., Brown, P.O and Botstein, D (1998) "Cluster analysis and display of genome-wide expression patterns," Proceedings of the National Academy of Sciences of the United States of America, 95 (25):14863–14868.

- Falzarano, D, Geisbert, T.W, Feldmann H (2011), Progress in filovirus vaccine development: evaluating the potential for clinical use. *Expert Rev Vaccines*, 10:63-77.

- Fan, J and Ren, Y (2006), Statistical Analysis of DNA Microarray Data in Cancer Research, *Clin Cancer Res*, 12(15):4469-4473.

- Feldmann, H and Geisbert, T.W (2010), Ebola haemorrhagic fever, *Lancet*:1-12.

- Garden, G.M and Littlejohn, T.G (2001), A comparison of microarray databases, *Brief Bioinform*, 2(2): 143-158.

- Geisbert, T.W., Lee, A, C and Robbins, M (2010) Postexposure protection of non-human primates against a lethal Ebola virus challenge with RNA interference: a proof-of-concept study, *lancet* (375):1896–1905.

- Geisbert, T.W., Young, H. A., Jahrling, P. B., Davis, K. J., Larsen, T., Kagan, E (2003), Pathogenesis of Ebola hemorrhagic fever in primate models: Evidence that hemorrhage is not a direct effect of virus-induced cytolysis of endothelial cells. *Am. J. Pathol.*, *163*(6), 2371–2382.

- Georgii, E., Richter, L., Ruckert, U and Kramer, S (2005), Analyzing microarray data using quantitative association rules, *Bioinformatics,* 21, Suppl 2:123-129.

- Guttula, S.V., Allam, A and Gumpeny, S. R (2012), Analyzing Microarray Data of Alzheimer's Using Cluster Analysis to Identify the Biomarker Genes, *IJAD*, Article ID 649456:1-5.

- Hartman, A. L., Towner, J. S and Nichol, S. T (2004), A C-terminal basic amino acid motif of Zaire ebolavirus VP35 is essential for type I interferon antagonism and displays high identity with the RNA-binding domain of another interferon antagonist, the NS1 protein of influenza A virus, *Virology, 328*(2), 177–184.

- Hensley, L. E., Young, H. A., Jahrling, P. B., and Geisbert, T. W (2002), Proinflammatory response during Ebola virus infection of primate models: Possible involvement of the tumor necrosis factor receptor superfamily, *Immunol. Lett, 80*(3), 169–179.

- Hoch, F.S and McCormick, J (1999), Experimental Filovirus infections. *Curr.Top.Microbiol. Immunol,* 235: 117–143.

- Ito, H., Watanabe, S., Sanchez, A., Whitt, M and Kawaoka,Y (1999), Mutational analysis of the putative fusion domain of Ebolavirus glycoprotein. *J. Virol.* 73, 8907–8912.

- Jensen,V.W., Kurz,S., Feldmann, F., Buehler, L.K., Kindrachuk, J., Filippis, V.D., Correia,J., Fruh, K., Kuhn, J.H., Burton, D.R and Heinz Feldmann, H (2011), Ebola Virion Attachment and Entry into Human Macrophages Profoundly Effects Early Cellular Gene Expression, www.plosntds.org, 5(10):1-16.

- Johnson, E, Jaax, N, White, J and Jahrling, P (1995), Lethal experimental infections of rhesus monkeys by aerosolized Ebola virus, *IJEP* 76 (4): 227–236.

- Kimberlin, C.R., Bornholdt, Z.A., Li, S., Woods ,V.L, Jr and MacRae, I.J (2010), Ebolavirus VP35 uses a bimodal strategy to bind dsRNA for innate immune, *Proc Natl Acad Sci U S A*, 107(1):314-329.

- Leroy, E.M., Rouquet, P., Formenty, P., Souquie`re, S., Kilbourne, A., Froment, J.M., Bermejo, M., Smit, S., Karesh, W., Swanepoel, R., Zaki, S.R and Rollin, P.E (2004), Multiple Ebola Virus Transmission Events and Rapid Decline of Central African Wildlife, *Science* (303).

- Leung, D.W., Prins,K.C., Basler, C.F and Amarasinghe, G.K (2010), *Ebolavirus* VP35 is a multifunctional virulence factor; Virulence, *Landes Bioscience* ,1:6:526-531.

- Levy, D.E and Darnell, J.E (2002), Stats: transcriptional control and biological impact, *Nat Rev Mol Cell Biol* 3: 651–662.

- Manicassamy, B., Wang, J., Rumschlag, E., Tymen, S., Volchkova, V., Volchkov, V and Rong,L (2007), Characterization of Marburgvirus glycoproteinin viral entry. *Virology* 358, 79–88.

- Matlab, (2012), *Bioinformatics toolbox, user's guide,* The Mathworks, inc, Chapter-4:1-33.

- Matsui, S., Yamanaka, T., Barlogie, B., Shaughnessy, J.D and Crowley, J (2008), Clustering of significant genes in prognostic studies with microarrays: application to a clinical study for multiple myeloma, *Stat Med*, 27(7):1106-20.

- Matsuno,K., Nakayama,E., Noyori, O.,Marzi,A.,Ebihara,H.,Irimura, T., Feldmann,H and Takada,A (2010),C-type lectins do not act as functional receptors for filovirus entry into cells. *Biochem.Biophys. Res.Commun*. 403, 144–148.

- Mohamadzadeh, M., Chen, L and Schmaljohn, A.L, How Ebola and Marburg viruses battle the immune system (2007), *Nat Rev Immunol*, 7:556-567.

- Ou, W., Delisle, J., Jacques, J., Shih, J., Price, G., Kuhn, J.H., Wang, V., Verthelyi, D., Kaplan,G and Wilson, C.A (2012), Induction of ebolavirus cross-species immunity using retrovirus-like particles bearing the Ebola virus glycoprotein lacking the mucin-like domain, *Virology Journal* 9:(32).

- Pati, S.K and Das, A.K (2011), Cluster Analysis of Microarray Data Based on Similarity Measurement, *IJBR*, 3(2): 207-213.

- Peters, C.J (2005) Marburg and Ebola virus hemorrhagic fevers. In: G L Mandell, Bennett, J and R Dolin (eds.), Principles and Practice of Infectious Diseases. Churchill, Livingstone Elsevier, *Philadelphia*: 1-10.

- Phoolcharoena, W., Dyec, J.M., Kilbournea, J., Piensooka, K., Prattc, W.D., Arntzena, C.J., Chend, Q., Masona, H.S and Herbst-Kralovetza, M.M (2011), A nonreplicating subunit vaccine protects mice against lethal Ebola virus challenge, *PNAS*, 108 (51): 20695–20700.

- Raychaudhuri, S., Stuphin, P. D., Chang, J.T and Altman, R.B (2001), Basic Microarray analysis: Grouping and feature reduction, *Trends Biotechnol*, 19(5):189-193.

- Reed, D.S and Mohamadzadeh, M (2007), Status and challenges of filovirus vaccines. *Vaccine*; 25:1923-1934.

- Reid, S.P., Leung, L.W., Hartman, A.L., Martinez O and Shaw, M.L (2006) Ebola virus VP24 binds karyopherin alpha1 and blocks STAT1 nuclear accumulation. *J Virol* 80: 5156–5167.

- Retuya, T.J.A., Miranda, M.E.G., Ksiazek, T, G., Khan, A, S and Sanchez, A (1997), Is the ebola-reston virus a threat to occupationally exposed humans, *J Clin Epidemiol* 50: 32S.

- Richardson, J.S., Dekker, J.D., Croyle, M.A and Kobinger, G.P (2010), Recent advances in Ebolavirus vaccine development. *Hum Vaccin*, 6:439-449.

- Rueda, L and Qin, L (2005), A New Method for DNA Microarray Image Segmentation, *Springer-verlag berlin Heidelberg*: 886-893.

- Ruf, W (2004). Emerging roles of tissue factor in viral hemorrhagic fever. *Trends Immunol*, *25*(9), 461–464.

- Saeed, M. F., Kolokoltsov, A.A., Freiberg, A. N., Holbrook, M.R and Davey, R.A (2008), Phosphoinositide-3 Kinase-Akt Pathway Controls Cellular Entry of Ebola Virus, *PLoS Pathogens* , 4 (8) : 1-11.

- Sanchez, A., Geisbert, T.W and Feldmann, H (2006), Filoviridae: Marburg, Ebola Viruses, Fields Virology, *Philadenph*ia, 1409-1448.

- Sanchez, A., Geisbert, T.W and Feldmann, H (2007), "Filoviridae: Marburg and Ebolaviruses," in *FieldsVirology*: 1409–1448.

- Sanchez, A., Khan, A.S., Zaki, S.R., Nabel, G.J and Ksiazek, T.G (2001), Filoviridae: Marburg and Ebola Viruses, *Fields Virology, Philadelphia*: 1279–1304.

- Sanchez, A., Yang, Z. Y., Xu, L., Nabel, G. J., Crews, T., and Peters, C. J (1998), Biochemical analysis of the secreted and virion glycoproteins of Ebola virus. *J. Virol., 72*(8), 6442–6447.

- Saravanakumar Selvaraj, Jeyakumar Natarajan (2011), Microarray Data Analysis and Mining Tools, *Bioinformation* 6(3): 95-99.

- Sherlock, G (2000), Analysis of large-scale gene expression data. *Curr. Opin. Immunol.* 12, 201–205

- Svrakic, N.M., Nesic, O., Dasu, M.R.., Herndon, D and Polo, P. J (2003), Statistical approach to DNA chip analysis. *Recent Prog Horm Res*, (58):75-93.

- Takada, A (2012), Filovirus tropism: cellular molecules for viral entry, *www.frontiers in.org* , 3(34): 1-9.

- Takada, A., Feldmann,H., Stroeher, U., Bray,M., Watanabe,S., Ito,H., McGregor,M and Kawaoka,Y (1997), Identificationofprotective epitopeson Ebola virus glycoprotein at the single aminoacid level by using recombinantvesicularstom- atitis viruses. *J. Virol.* 77, 1069–1074.

- Takada, A., Fujioka, K., Tsuiji, M., Morikawa, A., Higashi, N and Ebihara, H (2004), Human macrophage C-type lectin specific for galactose and *N*-acetylgalactosamine promotes filovirus entry, *J. Virol., 78*(6), 2943–2947.

- Tavazoie S, Hughes, J.D, Campbell, M.J, Cho RJ, Church, G .M (1999), Systematic determination of genetic network architecture, *Nat Genet.*, 22(3):281-295.

- Tibshirani, R, Hastie, T., Narasimhan, B and Chu, G (2002), Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci USA* (99):6567-6572.

- Toronen, P., Kolehmainen, M., Wong, G and Castren, E (1999) Analysis of gene expression data using self-organizing maps. *FEBS Lett.* 451, 142–146

- Transfusion, Volume 49, August 2009 Supplement.

- Tusher, A.G., Tibshirani, R and Chu, G (2001), Significance analysis of microarrays applied to the ionizing radiation response, *PNAS*, 98(9): 5116–5121.

- Tyagi, S., Kumar, S and Singla, M (2010), Clinical aspects of Ebola hemorrhagic fever: a review, *International Journal of Pharma and Bio Sciences.,* ISSN. , 1(3) 0975-6299.

- Ustun, C and Ozgurler, O (2005), Ebola: A Significant Threat as an Infectious Disease, and as a Potential Bioterrorism Agent, *Turk J Med Sci*; 35:1-4.

- Valarmathie, P., Ravichandran, T and Dinakaran, K (2012), Survey on Clustering Algorithms for Microarray Gene Expression Data, *EuroJournals Publishing*, 69(1) :5-20.

- Venet, D (2003), MatArray: a Matlab toolbox for microarray data, Bioinformatics, 19(5): 659–660.

- Wauquier N, Becquart P, Padilla C, Baize S, Leroy EM (2010) Human fatal zaire ebola virus infection is associated with an aberrant innate immunity and with massive lymphocyte apoptosis. *PLoS Negl Trop Dis* 4: e837.

- Wool, R.J and Bates,P (1998), Characterization of Ebolavirus entry by using pseudotyped viruses: identification ofreceptor-deficient celllines. *J. Virol.* 72, 3155–3160.

- Yamayoshi, S., Neumann, G and Yoshihiro, K (2010), Role of the GTPase Rab1b in *Ebolavirus* Particle Formation, *J. Virol*, 84(9): 4816-4820.

- Yang, Z.Y., Duckers, H.J., Sullivan, N. J., Sanchez, A., Nabel, E.G and Nabel, G.J (2000), Identificationof the Ebolavirus glycoprotein as the main viral determinant of vascular cell cytotoxicity and injury. *Nat. Med.* 6, 886–889.

- Yen, J., Garamszegi, S., Geisbert, J., Rubins, K., Geisbert, T., Honko, A., Xia, Y. A., Connor, H. J and Hensley, E. L (2011), Therapeutics of Ebola Hemorrhagic Fever: Whole-Genome Transcriptional Analysis of Successful Disease Mitigation, *JID*:1043-1052

- Zhang, A.P., Bornholdt, Z.A., Liu, T., Abelson, D.M, Lee, D.E., Li, S., Woods, V.L and Saphire, E. O (2012), The Ebola Virus Interferon Antagonist VP24 Directly Binds STAT1 and Has a Novel, Pyramidal Fold, *PLoS Pathogens*, 8(2):1-12.

WEBOGRAPHY

- http:// www.ncbi.nlm.nih.gov /geo/info/profiles.html
- http://allnurses.com/general-nursing-articles/t-tests-one-378029.html
- http://focus.rw/wp/2011/05/ministry-of-health-takes-measures-to-prevent-ebola-outbreak/
- http://microbewiki.kenyon.edu/images/9/93/Transfer.jpg
- http://phys.org/news152290263.html
- http://www.geneontology.org/
- http://www.innatedb.ca/
- http://www.ncbi. nlm.nih.gov/gene/375
- http://www.ncbi.nih.gov/geo/
- http://www.ncbi.nlm.nih.gov/gene/3569
- http://www.ncbi.nlm.nih.gov/geo/
- http://www.ncbi.nlm.nih.gov/geo/info/datasets.html
- http://www.statisticallysignificantconsulting.com/Ttest.htm
- http://www.tm4.org/mev/about
- http://www.tm4.org/mev/features
- http://www.tm4.org/mev_manual/test.html
- http://www.utmb.edu/virusimages/
- http://www.who.int/csr/disease/ebola/en/
- http://www-stat.stanford.edu/~tibs/SAM/