

# New insights about host response to smallpox using microarray data

Gustavo H. Esteves<sup>\*1</sup>, Ana C. Q. Simoes<sup>1,2</sup>, Estevao Souza<sup>3</sup>, Rodrigo A. Dias<sup>1</sup>, Raydonal Ospina<sup>3</sup> and Thiago M. Venancio<sup>2</sup>

<sup>1</sup>BIOINFO, Instituto de Matemática e Estatística, Universidade de São Paulo

<sup>2</sup>Laboratório de Bioinformática, Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo

<sup>3</sup>Departamento de Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo

\*Corresponding author

Email addresses:

GHE: [gesteves@vision.ime.usp.br](mailto:gesteves@vision.ime.usp.br)

ACQS: [anakqui@ime.usp.br](mailto:anakqui@ime.usp.br)

ES: [estevao@sigmabbs.com.br](mailto:estevao@sigmabbs.com.br)

RAD: [rdias@ime.usp.br](mailto:rdias@ime.usp.br)

RO: [rospina@ime.usp.br](mailto:rospina@ime.usp.br)

TMV: [venancio@iq.usp.br](mailto:venancio@iq.usp.br)

# Abstract

## Background

Smallpox is a lethal disease that was endemic in many parts of the world until eradicated by massive immunization. Due to its lethality, there are serious concerns about its use as a bioweapon. Here we analyze publicly available microarray data to further understand survival of smallpox infected macaques, using systems biology approaches. Our goal is to improve the knowledge about the progression of this disease.

## Results

We used KEGG pathways annotations to define groups of genes (or modules), and subsequently compared them to macaque survival times. This technique provided additional insights about the host response to this disease, such as increased expression of the cytokines and ECM receptors in the individuals with higher survival times. These results could indicate that these gene groups could influence an effective response from the host to smallpox.

## Conclusions

Macaques with higher survival times clearly express some specific pathways previously unidentified using regular gene-by-gene approaches. Our work also shows how third party analysis of public datasets can be important to support new hypotheses to relevant biological problems.

## Background

Large scale gene expression analysis with microarray technology is expanding and generating a large amount of high quality, publicly available data. In the present work we analyzed a dataset derived from monkeys infected by smallpox, published by Rubins et al. [1]. Smallpox is a lethal disease that was endemic in many parts of the world until eradicated by a massive immunization program developed by the World Health Organization. Its fatality rate was estimated to be 30%, and the survivors often had disfiguring scars [2].

There are serious concerns about the use of smallpox as a bioweapon [3, 4]. Recently, some health care workers were vaccinated by the UK government for the analysis of antibody responses [5]. Pox viruses display unique abilities to interfere with the host immune system, producing immune modulators [6] and there are at least 16 viral genes involved in combating the host immune response [7]. The original study's goal was to analyze the evolution of the gene expression of the peripheral blood cells of variola-infected monkeys, so as to clarify the biological processes associated with host-pathogen interactions [1].

Among the important results was the absence of a TNF- $\alpha$ /NF- $\kappa$ B-activated transcriptional mechanism during systemic infection, which could suggest that variola interaction with the mammalian host immune system may be ablative to this response[6].

Given the importance of this dataset, we have analyzed the data with systems biology approaches considering the expression profiles of gene groups instead of individual genes. These approaches are more powerful, less vulnerable to common experimental variation and usually more advantageous for obtaining biologically meaningful results.

## Results and Discussion

### Active Modules

The goal of our analysis was to identify differentially expressed modules in monkeys with early death, i.e., died in the first three days, and monkeys with a later death (ED and LD, respectively). Our hypothesis is that these modules play crucial roles in the host response to smallpox infection. The gene groups were defined according to the KEGG database (updated on 04/17/2006) [8], as described in the Methods section.

Initially, this analysis was based on the ED and LD monkey groups using all KEGG Orthology (KO) categories (at level 4), indicating three highly active modules in LD monkeys: Cytokines and Cell Adhesion Molecules (CAMs), which are consistent with the classical model of immune and inflammatory responses, where an activation and proliferation of leukocytes take place. The Alzheimer's Disease module also appeared as differentially activated, most likely due to its hybrid composition (20/22 genes that compose the AD module are present in the microarray platform used by Rubins et al. [1], and 14/20 are also members of other modules).

In addition to the ED and LD analysis, modules were also analyzed during the disease course. We have found 27 modules presenting differential activation in at least one point (Figure 2). Several of those modules are related to immune response, like CD molecules, CAM ligands, antigen processing and presentation and the previously considered activated in a late death Cytokines and Cell adhesion molecules.

A novel finding highlighted by our systems biology approach for all macaques (Figure 2) was an altered expression in the Natural Killer (NK) mediated cytotoxicity module. NK cells play a major role in the host-rejection of virally infected cells. Due to their strong cytolytic activity and the potential for auto-reactivity, they are tightly regulated. The activation of NK cells requires the action of proinflammatory cytokines in combination with cell surface receptors [9]. Under normal conditions, inhibitory NK cell receptors engage MHC class I (MHC-I) molecules. Such inhibitory activity normally predominates over that provided by the activating receptors, which bind to ligands expressed on virus-infected and other stressed cells [10]. Cells with normal levels of MHC-I are generally protected from NK cell-mediated lysis. Virally infected cells often express reduced levels of MHC-I and this inability to engage the inhibitory receptors makes the cells susceptible to NK cell attack (Cerwenka and Lanier, 2001). The down regulation of Natural killer cell mediated cytotoxicity module reported here could indicate a susceptibility to a NK cell mediated lysis (days 1-3, Figure 2).

Two modules with significant alterations between ED and LD groups were further explored using the Relevance Network (RN) approach, as discussed below. Additional File 1 and 2 shows other gene groups significantly altered in the time course infection and not discussed here.

### Relevance Networks (RNs)

To understand the profiles of interaction between the genes of the groups studied, we used a procedure known as relevance networks (RNs), initially proposed by Butte et al. [11]. Interaction here represents significant correlation between two variables (i.e. molecules), which may (but not necessarily will) physically interact with each other.

Cytokines and cell adhesion molecules (CAMs) modules were highly expressed in LD monkeys in the early phase of the infection (Figure 2) and were further explored in more detail by using the RN approach. A metamodule (CAMs and ECM receptors combined) was created to gain a better understanding of leukocyte migration to the infection locus. The connections between ICAM2 and ITGAV and ITGA6 shows a significant change ( $p$  value  $< 10^{-3}$ ) when comparing day 0 versus the earliest infection groups (days 1-3), see Figure 3A.

To establish a successful infection, pox viruses affect the release of immune modulators by the infected cells, so it is crucial to understand the dynamics of the host immune system[6]. The population of the immune cells varies during the infection, as do their exposed membrane proteins. Therefore, the Relevance Networks approach can reveal a more detailed perspective on the dynamics of the molecules involved in the disease. A better assessment of the variation of those molecules during the onset of the infection could give a better understanding of the host response to smallpox, and ultimately of what may be the key factors that set the 30% mortality rate.

To reach the infection focus and kill the pathogen, the immune cells have to adhere and overpass the endothelium. These events are possible due to the interaction between the integrins and the extra cellular matrix receptors and are modulated by cytokines. More specifically, ESAM is an endothelial cell adhesion molecule and ITGAV, ITGB1 and ITGB5 are integrins, which are heterodimeric integral membrane proteins composed of an alpha chain and a beta chain. ITGAV encodes an alpha chain; ITGB1 and ITGB5 encode beta chain integrins, known to participate in the cell-surface mediated signaling. ICAM2 is an intercellular adhesion molecule that promotes an inflammatory response and the activation of AKT (a kinase). AKT phosphorylates components that block the apoptosis pathway [12, 13]. To explore further the graphs produced by RNs for the ECM receptors + CAMs the relations concerning ICAM2 and ESAM are detailed. The complete network is available in Additional file 1 and 2. RNs of day 0 and days 1-3 show statistically significant alterations in the interactions between ICAM2 and the integrins ITGAV and ITGA6 (Figure 3A), which are important extracellular matrix receptors, involved in the adhesion, migration, differentiation, survival and proliferation of cells and cytoskeleton organization. Day 0 versus days 4-6 RNs revealed a significant relation between ICAM2 and other molecules that play crucial role in immune response against pathogens, PLXNC1 [14] and CDH5 [15] (Figure 3B). The relevance networks produced for cytokine group is available in Additional file 3.

ESAM shows a strong relation with ITGAV, ITGB1 and ITGB5 with variable patterns along the time course of infection. The quartet of ESAM, ITGAV, ITGB1 and ITGB5 reveals the relationship between alpha and beta chains in the composition of integrins and the relation of such heterodimeric proteins with ESAM. It is worth mentioning that ITGB3, ITGB6 and ITGB8, elements that interact with ITGAV, are present in the microarray platform design, but were excluded from analysis in the filtering steps (see Methods section).

Active modules and the RNs results pointed out drastic changes in the expression levels and interactions of important genes involved in essential cellular processes of the host response, such as cellular adhesion, apoptosis and immune response. These

results constitute interesting new biological hypotheses about smallpox virus infection that could be further explored by expert virology groups.

### **Overlapping with Rubins et al. and Jahrling et al..**

The data analyzed in our manuscript was previously analyzed by Rubins et al [1] and Jahrling et al [16]. The former focused on the microarray data and the later on the clinical findings. Rubins et al[1] reported the expression profile in IFN response, cell cycle/proliferation response, TNF- $\alpha$  / NF- $\kappa$ B regulated genes, dose response and lymphocyte and Ig responses. Meanwhile Jahrling et al [16] discuss the hemorrhagic diathesis, cytokines activation (IL6, IL8, INFG, CCL2 and CCL4) and lymphocyte depletion.

Rubins et al[1] performed a detailed analysis of IFN response genes. The KEGG module of cytokines includes IL6, IL8, IFN  $-\alpha$ ,  $-\beta$  and  $-\gamma$ , CCL2 (MCP-1), CCL4 (MIP1 $\beta$ ), TNF- $\alpha$  and NF  $-\kappa$ B. Additional files 4 and 5 show the influence of each molecule in the modules activation. Once KEGG separates the molecules into functional classes, it is not possible to define cascading effects, like Rubins did [1].

The decrease in the relative abundance of B cells and MHC II reported by Rubins et al[1] are corroborated by the profile of module Antigen processing and presentation (Figure 2). The Cell cycle, Complement and coagulation cascades, Cytokines and Hematopoietic cell lineage modules identified here were also reported by Rubins et al [1] as differentially expressed using gene by gene approaches.

Altered D-dimer levels were interestingly investigated by Jahrling et al. [16], which reflect the serious influence of smallpox virus inhibitors in the complement and coagulation cascades. Our active modules are also in agreement with such findings. Our results suggest an unbalanced cytokine levels, which is strongly associated with sepsis and organ failure death in infected animals.

In addition, we have also observed alterations in the hematopoietic cell lineage module which could be a consequence of the influx of immature leukocytes into the bloodstream, particularly neutrophils (aka "left shift", using medical jargon). Rubins et al [1] observed a considerable depletion of T cells, and a large number of apoptotic cells in the splenic periarterial lymphatic sheath, possibly due to the viral replication in antigen presenting cells in the lymphatic tissues, which could lead to apoptosis in the lymphocytes. The lysis of vulnerable cells induced by NK cells (described earlier) and the down regulation of the NK cell mediated cytotoxicity module supports such hypothesis.

## **Conclusions**

The usage of active modules pointed out the major role of the cytokines and ECM receptors in an effective host response. We have also pointed out that the population of the ECM receptors and cytokines significantly changes during the infection, possible due to a host response to eliminate the virus. In order to have a closer perspective of the dynamics of the molecules involved in this response we used the

Relevance Networks. Despite the need of an experimental validation of these new biological hypotheses, one of our goals was to exemplify and stress the importance of public microarray datasets analysis by using different and novel bioinformatics models and thus expand knowledge and information about important biological problems.

## Methods

### Data description

In this work we used cDNA microarray data containing 37,632 elements that represents approximately 11,000 unique genes and approximately 12,000 genes without identification [1, 17]. The slides were produced with human cDNA clones and the test samples were extracted from primate models (cynomolgus macaques). The samples of interest were always labeled with Cy5 and the reference samples were labeled with Cy3. The original dataset was obtained from the SMD [18, 19] and is associated with the *KRUBINS* key (experimenter). In this work, only data associated to studies 1 (Harper or India 7124 variola strains through a combination of aerosol exposure [5x10<sup>8</sup> plaque-forming units (pfu)] and high-titer i.v. inoculation [109 pfu]) and 2 (India 7124 strain via i.v. route alone [109 pfu]) [1] were used. Our goal is to analyze the host response to a successful infection; therefore the third study (varied viral dose) data was not used here.

The raw data was submitted to exploratory analyses. The main problem found was a bias of log ratio values from left to right in the majority of the chips, see Figure 1A. To address this problem, we used a normalization procedure proposed by Futschik and Crompton[20] known as OLIN. The OLIN method has some clear advantages: (i) it relies on strong statistical formalism; (ii) has the capacity to remove spatial and intensity (systematic) errors in microarray data, which is the main problem we found in the original dataset; (iii) Futschik and Crompton[20] precisely addressed the importance of optimization of model parameters on microarray data normalization (instead of simply accepting the method default values); (iv) the method performance was conveniently compared to other proposed methods[20]. Basically, the effect of log ratio values in relation to mean expression values (M and A, respectively) was estimated using LOWESS regression [21] with an optimization step for the selection of the parameter  $\alpha_A$  that control the smoothing of the regression. Subsequently, a new LOWESS regression was performed with log ratio values based on the x and y directions on the slide, and again, the optimization step of the parameters  $\alpha_x$  and  $\alpha_y$  was performed. This model successfully corrected the spatial-dependent artifact observed in the dataset (Figure 1B). Also, in a filtering step, we used only spots with signal intensity greater than background intensity in any sample and all spots with signal intensity greater than 2.5 times background intensity in 80% of the samples.

### Active modules

In order to find significantly altered gene groups we used the “active modules” approach, originally proposed by Segal et al [22]. The rationale of this method is to evaluate differential activation or repression of gene modules, which represent the biological processes.

The first step was to organize the data in a matrix  $E_{ij}$ ,  $i=1, \dots, n_s, j=1, \dots, n_l$ , where  $e_{ij}$  represents the gene  $i$  in the sample  $j$ . In the present study, 199 groups were selected according to the level 4 of KEGG[8] (updated on 04/17/2006), so each group has  $n_g$  genes,  $g=1, \dots, 199$ .

After the data normalization described earlier, expression values were standardized by subtracting the average expression calculated over all the arrays. This process adjusted the mean values of each gene to zero, which were then used to build a discrete version of  $E_{ij}$ , named  $E_{ij}^*$ , by applying the indicator function  $e_{ij}^* = I(e_{ij} \geq 1) - I(e_{ij} \leq -1)$  for all  $i$  and  $j$  (an indicator function assumes value 0 or 1 for false or true, respectively). Thus,  $E_{ij}^*$  is a matrix where each element can assume the discrete values 1 (induction), 0 (equal expression) or -1 (repression). Using the matrix  $E_{ij}^*$  the following operations are performed:

$$q_{1j} = \sum_{i=1}^{n_s} I(e_{ij}^* = 1) \quad q_{2j} = \sum_{i=1}^{n_s} I(e_{ij}^* = -1).$$

For each of the  $g$  groups of interest, each array  $j$  was then splitted in two disjoint sets of genes: one with exactly the  $n_g$  elements of the group, and other constituted by the remaining  $n_s - n_g$  genes. One should also consider two random variables  $Y_{1j}^{(g)}$  and  $Y_{2j}^{(g)}$  representing the number of induced genes in the first and second group, respectively.

Under the null hypothesis of no biological activity for  $g$  in sample  $j$ , Segal et al [22] propose that  $Y_{1j}^{(g)} \sim \text{hypergeometric}(n_g; n_s - n_g; q_{1j})$  and  $Y_{2j}^{(g)} \sim \text{hypergeometric}(n_g; n_s - n_g; q_{2j})$ . Intuitively, the absence of biological activity makes the probability of an induced (or repressed) gene to belong to the first group equal to  $1/2$  and  $p$ -values associated with the observed values  $y_{1j}^{(g)}$  and  $y_{2j}^{(g)}$  can be easily calculated. This procedure defined a new matrix  $P_{gj}$  where the element  $p_{gj}$  represents the fraction of induced genes of group  $g$  in the array  $j$ , or zero for the non-significant modules (we assumed a significance level  $\alpha=0.05$ ). In the cases where the significance is for the repressed genes, we just made the fraction  $p_{gj}$  negative. The arrays were then categorized in biological types  $t=1, \dots, t$  and for each pair of gene set  $g$  and biological type  $t$ , we count the number of arrays of type  $t$  that showed gene set  $g$  induced (repressed). We then used the hypergeometric distribution to classify each gene group as significantly induced or repressed across the biological types ( $p$  value  $< 0.05$ ).

The biological types are defined as following:  $n_t = 1$  means that the macaque was not infected,  $n_t = 2$  means death 1 to 3 days after infection,  $n_t = 3$  means death 4 to 6 days after infection and  $n_t = 4$  means death after the 6<sup>th</sup> day.  $n_t = 2$  was also labeled Early Death (ED) and  $n_t = 3$  and  $n_t = 4$  were labeled Late Death (LD). The significant contribution of molecules for activations in a module can be found in additional file 3.

## Relevance Networks (RNs)

For a better understanding of the Relevance Networks (RN) method, a definition of the term *interaction* is required. Here we are interested in a more general definition of protein interaction, which goes beyond physical interaction and includes also functional or phylogenetic relationships. Interaction here is inferred by significant

correlation between two variables (i.e. molecules), which may (but not necessarily do) physically interact with each other. Aiming to identify such correlations, the Relevance Networks (RN) method, initially proposed by Butte et al.[11], was employed. The method computes the squared linear Pearson correlation coefficient,  $r_{ij}^2$ , between all gene pairs for each biological type (as described in the active modules section of methods), defining a fully connected graph. Using a re-sampling method, the algorithm estimates a cut-off value,  $c$ , to split the graph into small sub-graphs where  $r_{ij}^2 > c$ . These sub-graphs are the relevance networks (RNs).

In order to identify significant differences between two RNs, we adapted this method to contrast two distinct conditions. Therefore, the linear Pearson correlation coefficients between every pair of genes in each condition,  $r_{ij}^1$  and  $r_{ij}^2$  are computed and then a Fisher's Z-transformation [23] is used to translate the probability distribution of the random variable  $r_{ij}^1 - r_{ij}^2$  into an approximated standard normally distributed random variable, which permits the identification of pairs of genes that displayed significant (positive or negative) differences between  $r_{ij}^1$  and  $r_{ij}^2$  ( $p < 0.001$ ). This type of analysis was accomplished for the 199 gene groups used in the active modules procedure, linking the findings with biological knowledge.

## Authors' contributions

All authors conceived the study. GHE implemented and adapted the methods. ACQS implemented the retrieval of the groups from KEGG. ACQS and TMV analyzed the results and drafted the latest version of the manuscript. ES, RAD, RO participated in the statistical modeling. All authors read and accepted the final version of this manuscript.

## Acknowledgements

Special thanks to Lucas Perin for the English revision and the three reviewers for critical and insightful suggestions. The authors would like also to thank Dr. Eduardo Jordão Neves for discussions about the statistical models and for reading the manuscript; Dr. Luiz Fernando Reis, Dra. Erna Kroon, Dr. Ricardo Vêncio, Tobias Mann, Leonardo Varuzza, Elier Cristo and Dr. Roberto Hirata Jr. for helpful discussions. GHE is supported by CAPES grant; ES is supported by CNPq grant; ACS, TMV, RAD and RO are supported by FAPESP grants. GHE, ACS and TMV are PhD students in Bioinformatics at Universidade de São Paulo; ES and RO are PhD students in Statistics at Instituto de Matemática e Estatística, Universidade de São Paulo; RAD is PhD student in Computer Science at Instituto de Matemática e Estatística, Universidade de São Paulo.

## References

1. Rubins KH, Hensley LE, Jahrling PB, Whitney AR, Geisbert TW, Huggins JW, Owen A, Leduc JW, Brown PO, Relman DA: **The host response to smallpox: analysis of the gene expression program in peripheral blood cells in a nonhuman primate model.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(42):15190-15195.
2. Fenner F, Anderson D, Arita I, Jezek Z, Ladnyi I: **Smallpox and Its Eradication.** *World Health Organization* 1988.
3. Henderson DA, Inglesby TV, Bartlett JG, Ascher MS, Eitzen E, Jahrling PB, Hauer J, Layton M, McDade J, Osterholm MT *et al*: **Smallpox as a biological weapon: medical and public health management. Working Group on Civilian Biodefense.** *Jama* 1999, **281**(22):2127-2137.
4. Bozzette SA, Boer R, Bhatnagar V, Brower JL, Keeler EB, Morton SC, Stoto MA: **A model for a smallpox-vaccination policy.** *The New England journal of medicine* 2003, **348**(5):416-425.
5. Putz MM, Midgley CM, Law M, Smith GL: **Quantification of antibody responses against multiple antigens of the two infectious forms of Vaccinia virus provides a benchmark for smallpox vaccination.** *Nature medicine* 2006, **12**(11):1310-1315.
6. Turner PC, Moyer RW: **Poxvirus immune modulators: functional insights from animal models.** *Virus research* 2002, **88**(1-2):35-53.
7. Seet BT, Johnston JB, Brunetti CR, Barrett JW, Everett H, Cameron C, Sypula J, Nazarian SH, Lucas A, McFadden G: **Poxviruses and immune evasion.** *Annual review of immunology* 2003, **21**:377-423.
8. Kanehisa M, Goto S, Kawashima S, Nakaya A: **The KEGG databases at GenomeNet.** *Nucleic acids research* 2002, **30**(1):42-46.
9. Andoniou CE, Andrews DM, Degli-Esposti MA: **Natural killer cells in viral infection: more than just killers.** *Immunological reviews* 2006, **214**:239-250.
10. Cerwenka A, Lanier LL: **Natural killer cells, viruses and cancer.** *Nature reviews* 2001, **1**(1):41-49.
11. Butte AJ, Tamayo P, Slonim D, Golub TR, Kohane IS: **Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks.** *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**(22):12182-12186.
12. Perez OD, Kinoshita S, Hitoshi Y, Payan DG, Kitamura T, Nolan GP, Lorens JB: **Activation of the PKB/AKT pathway by ICAM-2.** *Immunity* 2002, **16**(1):51-65.
13. Brunet A, Bonni A, Zigmond MJ, Lin MZ, Juo P, Hu LS, Anderson MJ, Arden KC, Blenis J, Greenberg ME: **Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor.** *Cell* 1999, **96**(6):857-868.
14. Comeau MR, Johnson R, DuBose RF, Petersen M, Gearing P, VandenBos T, Park L, Farrah T, Buller RM, Cohen JI *et al*: **A poxvirus-encoded semaphorin induces cytokine production from monocytes and binds to a novel cellular semaphorin receptor, VESPR.** *Immunity* 1998, **8**(4):473-482.
15. Tzima E, Irani-Tehrani M, Kiosses WB, Dejana E, Schultz DA, Engelhardt B, Cao G, DeLisser H, Schwartz MA: **A mechanosensory complex that**

- mediates the endothelial cell response to fluid shear stress.** *Nature* 2005, **437**(7057):426-431.
16. Jahrling PB, Hensley LE, Martinez MJ, Leduc JW, Rubins KH, Relman DA, Huggins JW: **Exploring the potential of variola virus infection of cynomolgus macaques as a model for human smallpox.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(42):15196-15200.
  17. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X *et al*: **Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling**, vol. 403; 2000.
  18. Sherlock G, Hernandez-Boussard T, Kasarskis A, Binkley G, Matese JC, Dwight SS, Kaloper M, Weng S, Jin H, Ball CA *et al*: **The Stanford Microarray Database.** *Nucleic acids research* 2001, **29**(1):152-155.
  19. **SMD Advanced Search** [<http://smd.stanford.edu/cgi-bin/search/QuerySetup.pl>]
  20. Futschik M, Crompton T: **Model selection and efficiency testing for normalization of cDNA microarray data.** *Genome biology* 2004, **5**(8):R60.
  21. Cleveland W: **Robust locally weighted regression and smoothing scatterplots.** *J Am Stat Ass* 1979, **74**:8.
  22. Segal E, Friedman N, Koller D, Regev A: **A module map showing conditional activity of expression modules in cancer.** *Nature genetics* 2004, **36**(10):1090-1098.
  23. Zar: **Biostatistical analysis:** Prentice Hall; 1999.

## Figures

### Figure 1 - OLIN normalization

OLIN normalization method proposed by Futschik & Crompton [20]. **A-** The spatial bias effect found. This effect was noted in most of the chips; **B-** The same slide after OLIN normalization.

### Figure 2 - Active module analysis

Heat map representation of gene groups (modules) with significant alteration ( $p < 0.05$ ) are shown. The red and green colors represent positive (activation) and negative (repression) regulation, respectively.

### Figure 3 - Relevance networks analysis

Relevance networks ( $p < 10^{-3}$ ). Changes in correlation between genes at day 0 (left panel) and days 1-3 (right panel) (**A**) and day 0 (left panel) versus days 4-6 (right panel) (**B**). Solid and dashed lines represent positive and negative correlation values, respectively.

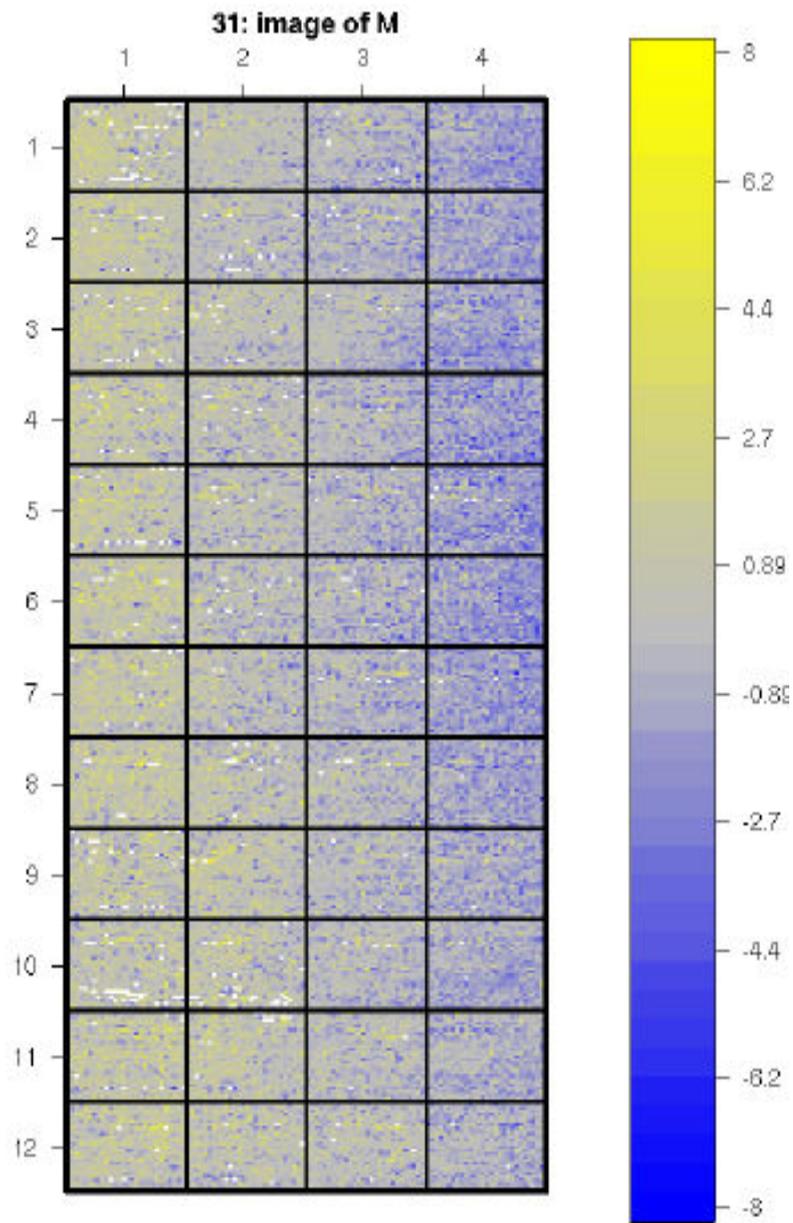
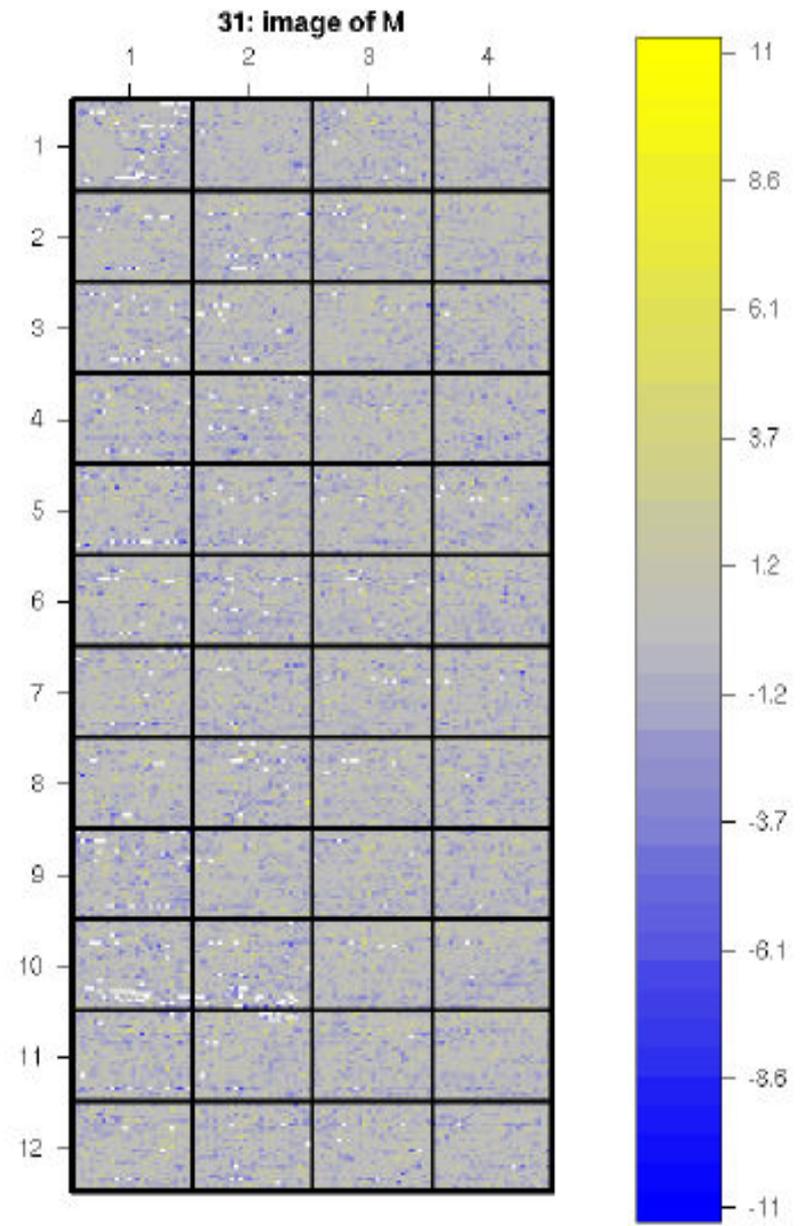
**A****B**

Figure 1

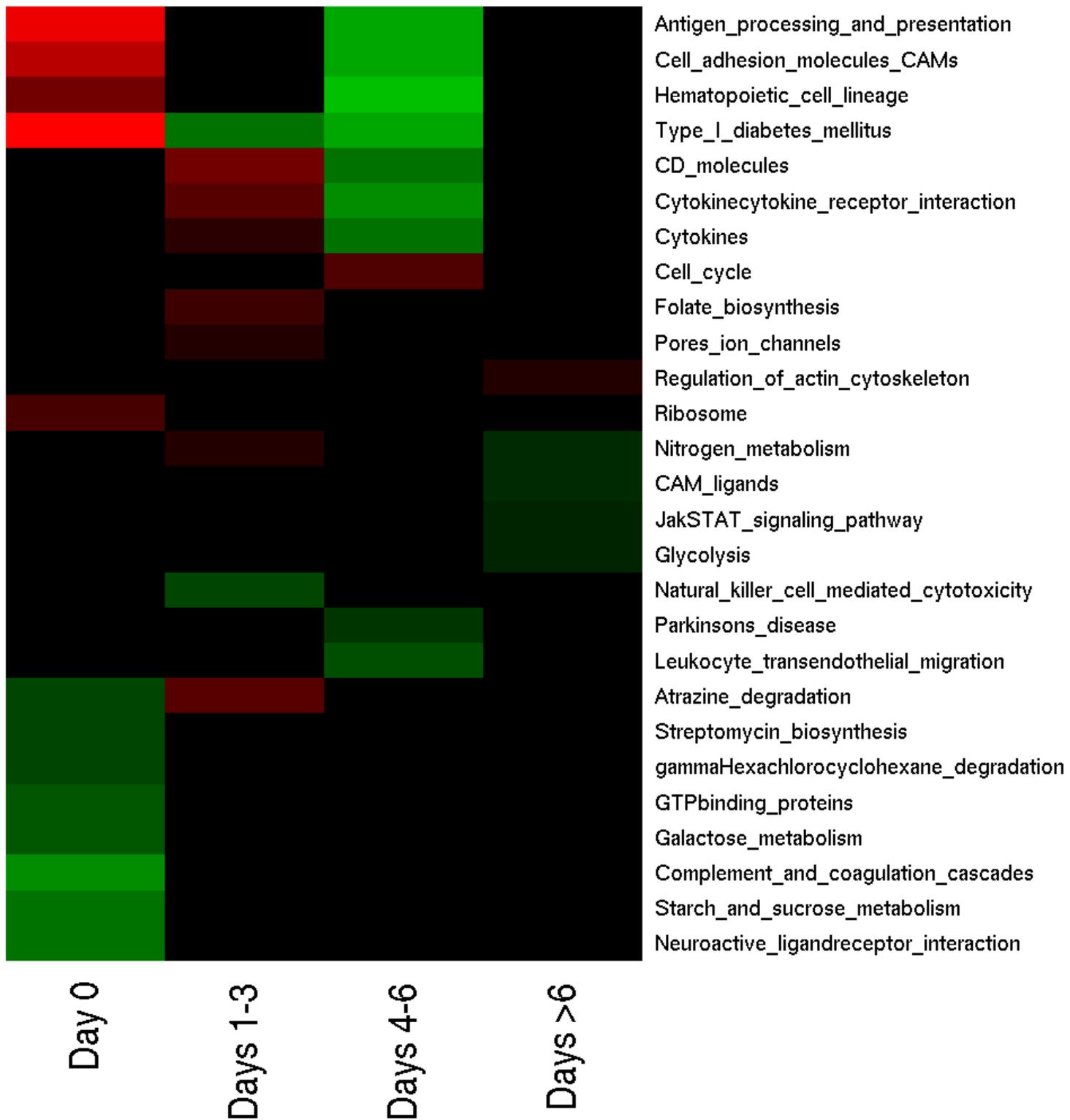
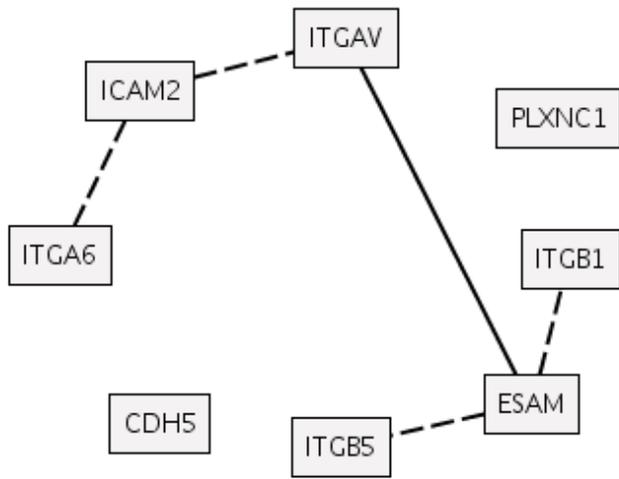


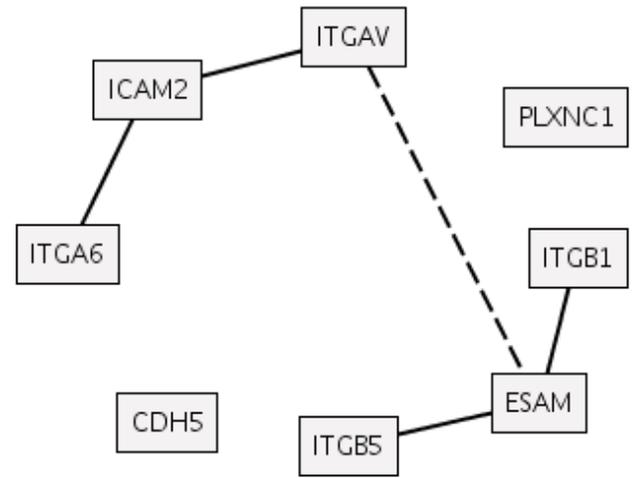
Figure 2

**A**

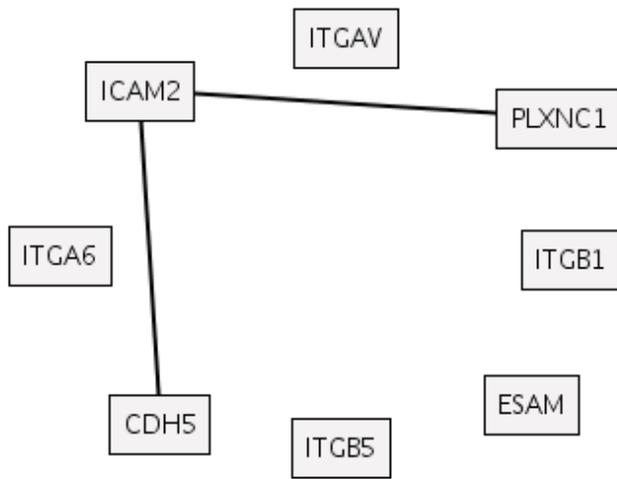
Day 0



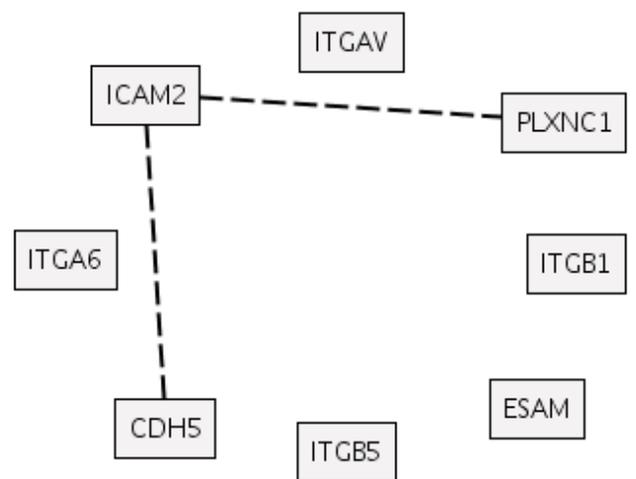
Days 1-3

**B**

Day 0



Days 4-6



**Additional files provided with this submission:**

Additional file 1: addfile1.pdf, 334K

<http://www.biomedcentral.com/imedia/1866969560142644/supp1.pdf>

Additional file 2: addfile2.pdf, 316K

<http://www.biomedcentral.com/imedia/5422341891426449/supp2.pdf>

Additional file 3: addfile3.pdf, 275K

<http://www.biomedcentral.com/imedia/7525111611426450/supp3.pdf>

Additional file 4: addfile4.html, 104K

<http://www.biomedcentral.com/imedia/7783137471426452/supp4.html>

Additional file 5: addfile5.html, 11K

<http://www.biomedcentral.com/imedia/8438312381426452/supp5.html>