Does the immune system have an influence on malaria parasite gene expression?

ARTICLE

READS
12

4 AUTHORS, INCLUDING:

Marcel van Gerven
Radboud University Nijmegen
82 PUBLICATIONS 774 CITATIONS
SEE PROFILE

Tom Heskes
Radboud University Nijmegen
217 PUBLICATIONS 2,602 CITATIONS
SEE PROFILE

Does the immune system have an influence on malaria parasite gene expression?

Adriana Birlutiu, Tjeerd M.H. Dijkstra, Marcel van Gerven, Tom Heskes

Institute for Computing and Information Sciences (iCIS), Radboud University Nijmegen

Abstract

The malaria parasite has a direct influence on the immune system of its host, as witnessed by periodic fevers. It is unclear whether the host immune system has a direct influence on parasite gene expression. The parasite is known to secrete proteins that are thought to manipulate the host immune system. An open question is whether the malaria parasite regulates these proteins in response to immune system pressure. A recent study of Daily et al. [2] shows that in vivo parasite gene expression from 35 patients can be classified into three clusters. Besides expression data, Daily et al. collected sixteen patient characteristics, some of which are significantly different between the clusters (rank order test).

We take an integrated probabilistic approach by predicting the malaria parasite cluster based on patient characteristics and their interactions. As the number of predictors (136) is larger than the number of patients, we use L1-regularized multinomial logistic regression. We find that the interaction of immune signaling molecules TNFα and TGFα explain cluster membership best, consistent with the notion that parasite gene expression reacts to immune system pressure.

1. Clustering based on gene expression data

Data consist of 5159 gene expression profiles of parasites derived directly from venous blood samples of 35 patients. Clustering the expression data using non-negative matrix factorization ([1]) results in three clusters.

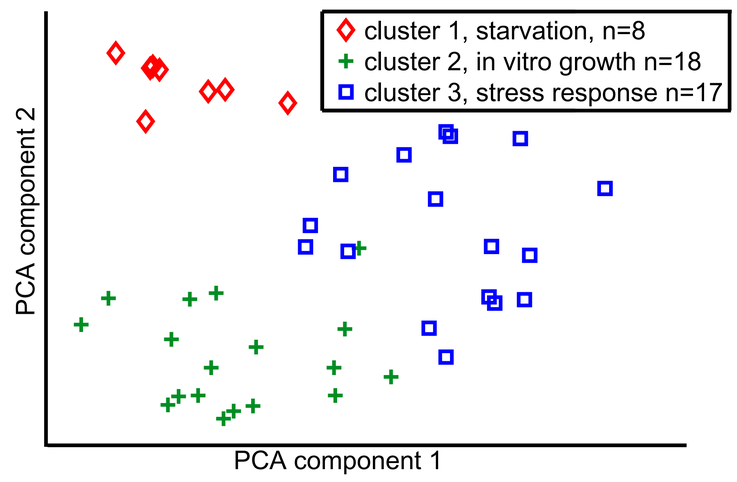

Figure 1: *Parasite gene expression projected from 5159 to 2 dimensions using PCA. Data labels correspond to the clustering as obtained by Daily et al. [2]*

2. Classification based on patient characteristics

Patient characteristics (clinical correlates) consist of general information (age, weight, etc.) and immune markers (interleukin, P-selectin, lymphotactin, etc). First we recreated the analysis of [2] and performed Mann-Whitney U-tests with cluster 2 as reference. We find that cluster 1 and 2 cannot be separated and that TNFα and TGFα are most significantly different between clusters 2 and 3.

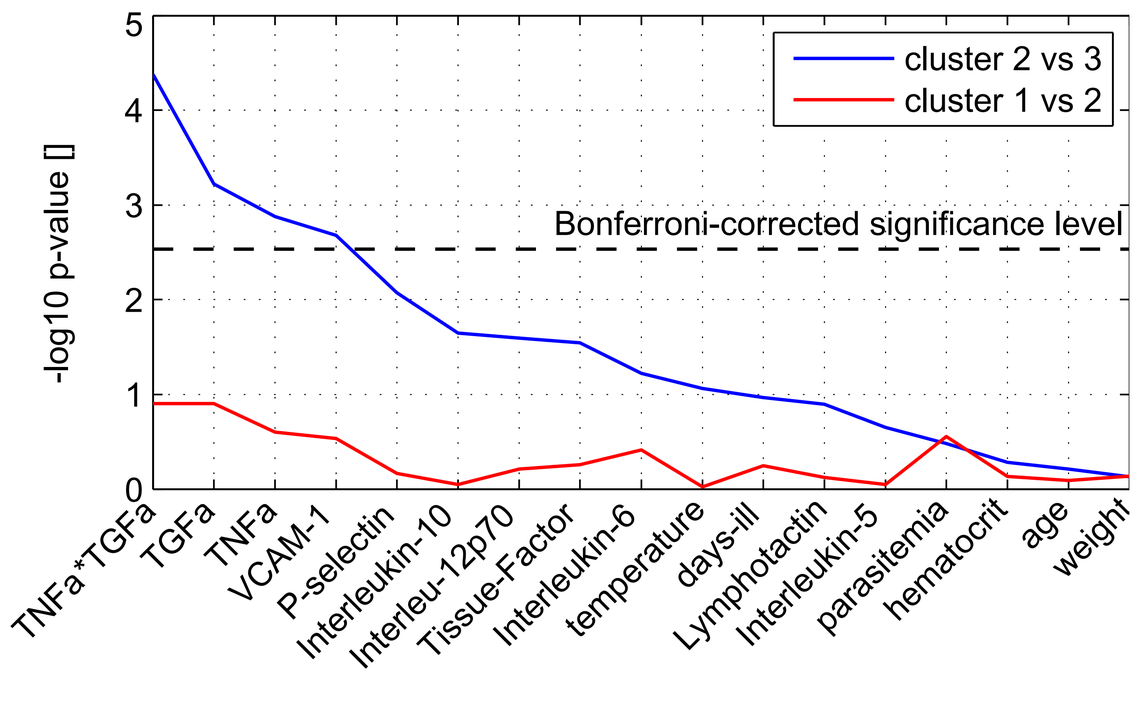

Figure 2: *Significance of Mann-Whitney U-test for all patient characteristics.*

The goal is to build a model that predicts the cluster membership of gene expression data from the patient characteristics. We consider the multinomial logistic regression model

$$P(k|\boldsymbol{b},\boldsymbol{Y}) = \frac{\exp\left(\boldsymbol{b}_k^T\boldsymbol{Y}\right)}{\sum_{k'}\exp\left(\boldsymbol{b}_{k'}^T\boldsymbol{Y}\right)},$$

with cluster k and patient characteristics $\boldsymbol{Y}$. As patient characteristics have different units, we $\log$ transformed and standardized them. Parameters $\boldsymbol{b}$ are found by maximizing the L1-penalized log-likelihood (j runs over patients and m over patient characteristics):

$$\boldsymbol{b} = \underset{\boldsymbol{b}}{\operatorname{argmax}}\left(\sum_j\sum_k \log P(k|\boldsymbol{b},\boldsymbol{Y}^j) - \lambda\sum_{k,m}|b_{k,m}|\right).$$

3. Results

We considered all patient characteristics and the pairwise interactions between them, for a total of 136 predictors. We obtained predictive accuracy using leave-one-out cross-validation. We compared the label predicted by the multinomial logistic model with the one obtained by clustering the gene expression data. The prediction accuracy is 69% with a significance level of $p = 0.09$ obtained by comparing with the classifier which assigns all patients to the largest cluster.

predicted	cluster 1	cluster 2	cluster 3
actual			
cluster 1	0	6	0
cluster 2	0	10	2
cluster 3	0	3	14

As the table shows it is not possible to separate cluster 1 from 2 and 3 but it is possible to separate clusters 2 and 3. The non-separability can also be seen in Fig 2 (red curve). This is surprising as NMF analysis of parasite gene expression suggests that cluster 1 is separate from 2 and 3 (Fig 1). A different analysis of the Daily data by Lemieux et al [3] suggests that clusters 1 and 3 are close and cluster 2 is different. An explanation might be that cluster 1 is a starvation response and that metabolites (glucose) were not assayed.

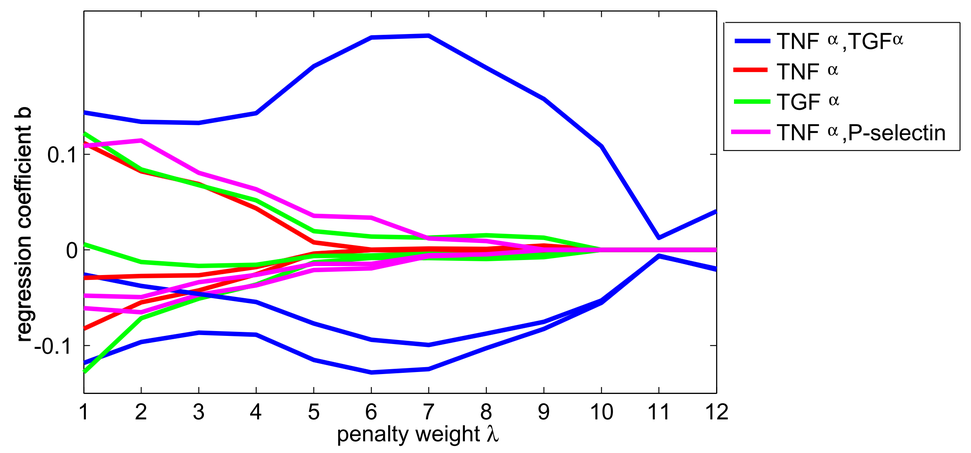

Figure 3: *Regression coefficients $\boldsymbol{b}$ corresponding to four (interactions of) patient characteristics as a function of regularizer weight λ.*

The three most relevant patient characteristics for cluster prediction are:

- TNFα by TGFα interaction term
- TNFα (Tumor Necrosis Factor α),
- TGFα (Transforming Growth Factor α).

We included the TNFα–TGFα interaction also in the re-analysis of Fig 2 where it came out as the most significant factor too. The Pearson correlation coefficient between them is 0.36. TNFα is a universal alarm molecule of the immune system, TGFα less so. Their interaction has been little studied in malaria research.

Patient characteristics can distinguish clusters 2 and 3, suggesting that immune system pressure influences the malaria parasite. Cluster 1 cannot be distinguished using only patient characteristics.

References

[1] J-P. Brunet et al. Metagenes and molecular pattern discovery using matrix factorization. Proc Nat Acad Sci, v 101, p 4164–4169, 2004.

[2] J.P. Daily et al. Distinct physiological states of *Plasmodium falciparum* in malaria-infected patients. Nature, v 450, p 1091–1097, December 2007.

[3] J.E. Lemieux et al. Statistical estimation of cell-cycle progression and lineage commitment in *Plasmodium falciparum* reveals a homogeneous pattern of transcription in ex vivo culture. Proc Nat Acad Sci, v 106, p 7559–7564, 2009.