Enhancing diagnosis of tuberculosis in children with novel Mycobacterium tuberculosis antigens

Conference Paper

Author(s):
Sutter, Thomas; Meier, Noemi; Jacobsen, Marc; Ottenhoff, Tom H.M.; Vogt, Julia E.; Ritz, Nicole

Publication date:
2020-07-17

Permanent link:
https://doi.org/10.3929/ethz-b-000456061

Enhancing diagnosis of tuberculosis in children with novel Mycobacterium tuberculosis antigens

Thomas M. Sutter [1] **Noemi Meier** [2 3] **Marc Jacobsen** [4] **Tom H.M. Ottenhoff** [5] **Julia E. Vogt** [1] **Nicole Ritz** [2 3 6 7]

Abstract

Diagnosing tuberculosis in children is a challenging task. Microbiological confirmation of tuberculosis disease is often lacking, and standard immunodiagnostics including the tuberculin skin test and interferon-γ release assay for tuberculosis infection have limited sensitivity. Recent research suggests that the inclusion of novel Mycobacterium tuberculosis antigens have the potential to improve standard immunodiagnostic tests for tuberculosis. In this work, we aim to identify new optimal antigen-cytokine combinations using novel Mycobacterium tuberculosis antigens and cytokine read-outs in order to improve immunodiagnostic assays for tuberculosis. As diagnostic tools for tuberculosis in children are still insufficient, new immunodiagnostic tests have the potential to improve diagnosis in this patient group.

1. Introduction

Tuberculosis (TB) remains one of the leading causes of death globally. Current estimates show that one in ten TB cases occur in children below 15 years of age with an annual estimated number of 1 million cases of childhood TB disease in 2017 (WHO, 2018a). Despite being a preventable and curable infectious disease, 233'000 children died of TB in 2017, of which 80% occurred in children below 5 years of age. The recent World Health Organization roadmap towards ending TB in children and adolescents mentions up to 69% underdiagnosis and highlights the development of accurate, non-sputum-based diagnostics tests for TB disease and infection as a key action towards ending TB in children and adolescents (WHO, 2018b). In young children TB disease is often of paucibacillary nature and therefore remains undiagnosed (Perez-Velez & Marais, 2012). In addition collection of samples for microbiological proof in this patient group is challenging and reaches 50% at best (Nemeth et al., 2014). As a consequence, non-sputum-based diagnostic tests based on immunological evidence of TB have been developed. These tests rely on the measurement of a recall cell mediated immune response triggered by in vivo or in vitro mycobacterial antigens. Until two decades ago the tuberculin skin test has been the standard test, measuring a local skin induration after injection of purified protein derivative, a Mycobacterium tuberculosis protein mixture. However, because of its low specificity especially in Bacille Calmette-Guérin vaccinated individuals, interferon-gamma release assays have been developed, and have become the standard immunodiagnostic test of TB infection in adults (Diel et al., 2010). Unfortunately interferon-gamma release assay have two major limitations: lower performance in children with a sensitivity ranging from 62% to 83% and inability to discriminate between TB disease and TB infection (Sollai et al., 2014; Mandalakas et al., 2011). Recent research suggests that incorporation of novel Mycobacterium tuberculosis antigens expressed during different stages of TB (reviewed in Meier et al. (2018)) and the measurement of additional cytokines (Walzl et al., 2011) can improve performance of currently used interferon-gamma release assay. Evaluation of novel diagnostic tests incorporating different Mycobacterium tuberculosis antigens and cytokines are therefore feasible tests suitable for pediatrics and urgently needed (WHO, 2013). The aim of our study is to include novel Mycobacterium tuberculosis antigens and measure additional cytokines for the immune diagnosis of childhood TB. We use supervised and unsupervised machine learning algorithms to compare groups and identify the best antigen-cytokine pairs.

[1]Department of Computer Science, ETH Zurich, Zurich, Switzerland [2]Mycobacterial Research Laboratory, University of Basel Children's Hospital, Basel, Switzerland [3]Faculty of Medicine, University of Basel, Basel, Switzerland [4]Children's University Hospital Düsseldorf, Düsseldorf, Germany [5]Department of Infectious Diseases, Leiden University Medical Center, Leiden, Netherlands [6]Pediatric Infectious Diseases and Vaccinology Unit, University of Basel Children's Hospital, Basel, Switzerland [7]Department of Pediatrics, Royal Children's Hospital Melbourne, Parkville, Australia. Correspondence to: Thomas M. Sutter <thomas.sutter@inf.ethz.ch>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 108, 2020.

2. Data

The Childhood Tuberculosis in Switzerland Study (CITRUS) is a prospective multicenter observational study (registered at ClinicalTrials.gov NCT03044509). Eligible are children undergoing evaluation for TB exposure, infection or disease below the age of 18 years. Children that have been treated previously or that have started treatment more than 5 days before study inclusion are excluded. Upon enrolment baseline characteristics, clinical scores and TB test results done by the treating physician are recorded. The study participants were classified into the following three groups: confirmed/unconfirmed TB, TB infection, and unlikely TB according to previously published case definitions (Graham et al., 2015) (for further details see appendix A). Granulocyte-macrophage colony-stimulating factor (GM-CSF), interferon (IFN)-γ, IFN- γ-inducible protein (IP)-10, interleukin (IL)-1 receptor-antagonist (RA), IL-2, IL-6, IL-10, IL-13, IL-17, soluble cluster of differentiation 40 ligand (sCD40L) and tumor necrosis factor (TNF)-α were measured using a Luminex technology according to manufacturer's instructions. As a pre-processing step, Cytokine concentrations were normalized (Dodge, 2006) both within antigen-cytokine pairs (using a minimum-maximum (min-max) or a mean-standard deviation (mean-std) normalization) and within a patient's distribution of values (using a mean-std normalization) as indicated.

3. Methods

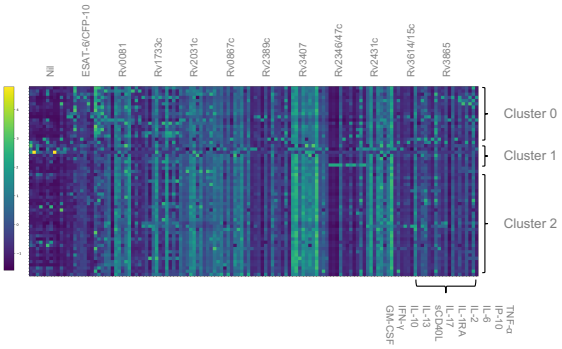
In order to detect novel antigen-cytokine pairs that are able to distinguish between the different patient groups, we trained a well-established classifier. Discrimination of confirmed/unconfirmed TB and TB infection versus TB exposed, based on data containing information on all antigen-cytokine pairs, was achieved using a logistic regression classifier with L2-regularisation (Hoerl & Kennard, 1970) (see appendix A). To get a reliable estimate of the discriminative classifier performance, a five-fold cross-validation was applied to a set of training data to select the model's hyperparameters. The performance of the discriminative classifier was evaluated using area under the receiver operating characteristics (AUROC) (Hanley & Mcneil, 1982). The contribution of each antigen-cytokine pair to our predictive model was evaluated by analyzing the weight in the decision function.

A further goal of this study is to better understand three pre-defined patient subgroups (confirmed/unconfirmed TB disease, TB infection, unlikely TB). To achieve this aim, we analysed the data with an unsupervised, data driven approach. As we pre-defined 3 patient groups, we used K-means clustering algorithm (MacQueen, 1967) with a given number of clusters (n=3) reflecting the anticipated number of patient groups of interest. Patients with incomplete measurements in any of the conditions (e.g. missing values) were excluded from this analysis.

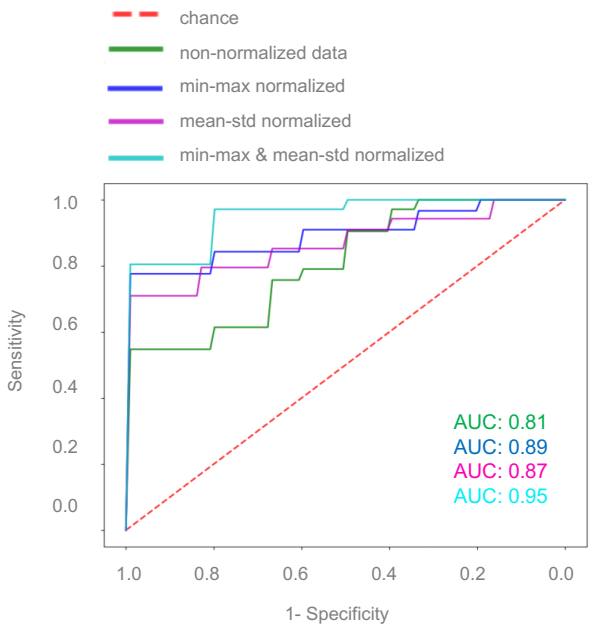
As an additional analysis of the 3 patient sub-groups of interest, we compared the differences in median cytokine concentrations between confirmed/unconfirmed TB, TB infection and unlikely TB. Antigen-cytokine pairs with the greatest differences were selected and K-means clustering approach was performed as above on these selected antigen-cytokine pairs.

4. Experiments & Results

A discriminative classifier distinguishes healthy from sick children and normalization of data results in improvement of the classifier's performance. A total of 59 patients had complete measurements for all antigen-cytokine pairs and were included in this analysis: confirmed TB (n=8), unconfirmed TB (n=2), TB infection (n=17) and unlikely TB (n=32). Different methods of normalization (e.g. non-normalized data, antigen-cytokine pairs either normalized using min-max or mean-std normalization and normalization of antigen-cytokine pairs with min-max and between patient normalization with mean-std were applied to our dataset and resulted in differences between antigen-cytokine pairs and cytokine concentrations (supplementary Figure 3 A-D). These differences influenced the outcome of the discriminative classifier (confirmed/unconfirmed TB and TB infection versus TB exposed). The AUROC was lower without normalization (AUROC = 0.81 ± 015), compared to a normalization of antigen-cytokine pairs (AUROC min-max = 0.89 ± 0.12 and AUROC mean-std = 0.87 ± 0.13) or combining an antigen-cytokine pair normalization with individual patient normalization (AUROC min-max/mean-std = 0.95 ± 0.03) (Figure 1b). The most important antigen-cytokine pairs that contributed to the performance of the discriminative classifier were consistent for the normalization methods used. Rv2346/47c- and Rv3614/15c- induced concentrations of IP-10 were the two antigen-cytokine pairs with the highest weight in the predictive model for all discriminative classifiers with normalized data (Figure 2b, supplementary Figure 4 A-C). The weight of ESAT-6 and CFP-10 - induced concentrations of TNF-γ for the predictive model was consistently high for all normalized and non-normalized data. ESAT-6/CFP-10- induced concentrations of IFN-γ were among the 10 antigen-cytokine pairs that contributed the most to the classifier for all non-normalized and normalized data except when mean-std normalization alone was applied. Rv2031c- induced concentrations of GM-CSF contributed to the performance of the classifier when any normalization method was applied with increasing weight for combined min-max and mean-std normalization. Combining data from the 10 antigen-cytokine pairs with the highest weight in the predictive model using both



(a) Cytokine concentrations for individual patients.



(b) Performance of the binary classifier.

Figure 1. Normalization of data contributes to performance of discriminative classifier. Figure 1a shows the cytokine concentrations for individual patients. Results are sorted by patient group and clusters (2, 1 or 0), and antigen-cytokine pairs. Min-max normalization was applied to cytokine-antigen concentrations, mean-std normalization was applied to between-individual measurements (color change from dark blue to light green represents an increase in relative cytokine concentration). Figure 1b shows the AUROC curve of the performance of the binary classifier (confirmed/unconfirmed TB and TB infection versus TB exposed) in 59 patients using different normalization methods: min-max and mean-std; normalization of antigen-cytokine pairs; min-max/mean-std combining an antigen-cytokine pair normalization with individual patient normalization.

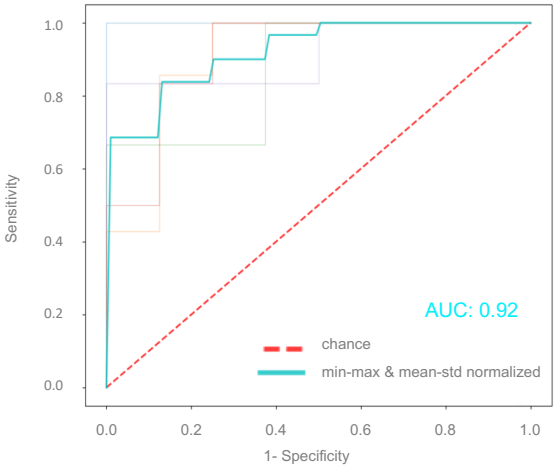
min-max and mean-std normalization resulted in AUROC min-max/mean-std = 0.92 ± 0.04 (Figure 2a).

Unsupervised K-means clustering reveals three groups of children that cannot be explained by disease status. We found three clusters which did not overlap with our patient groups (i.e. confirmed & unconfirmed TB, TB infection, unlikely TB) in the unsupervised analysis approach. All three clusters included patients from all study groups. Figure 1a displays normalized cytokine concentrations of antigen-cytokine pairs of all individual patients sorted by cluster (2, 1 or 0). Cluster 0 consisted of 4 confirmed TB, 1 unconfirmed TB, 6 TB infection and 5 unlikely TB patients (median age = 8.4, 68.7% male). Cluster 1 consisted of 2 confirmed TB, 0 unconfirmed TB, 2 TB infection and 1 unlikely TB patients (median age = 13.6, 20.0% male). Cluster 2 consisted of 2 confirmed TB, 1 unconfirmed TB, 9 TB infection and 26 unlikely TB patients (median age = 7.8, 55.3% male). Clusters could neither be explained by disease classification, nor age, nor gender, nor ethnicity (data not shown). This is a surprising finding which we will investigate in further work.

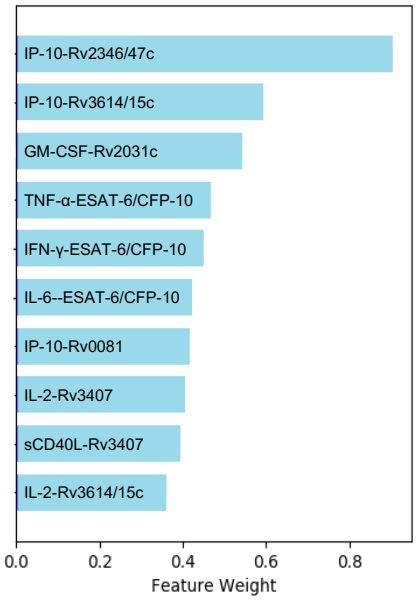
Supervised K-means clustering based on median cytokine differences between three study groups reveals one group that clustered mainly healthy children but no confirmed TB cases. Greatest differences in median cytokine concentrations between confirmed/unconfirmed TB, TB infection and unlikely TB were observed for: ESAT-6/CFP-10-induced concentrations of GM-CSF, IFN-γ and IL-2; Rv0081-induced concentrations of TNF-α; Rv2389c-induced concentrations of GM-CSF and IP-10; and Rv3614/15c-induced concentrations of IFN-γ, IL-2, IP-10 and TNF-α (data not shown). A total of 71 patients had complete measurements for these 10 conditions with the greatest differences and were thus further included in the comparative analysis: confirmed TB (n=10), unconfirmed TB (n=4), TB infection (n=25) and unlikely TB (n=32). K-means clustering with these antigen-cytokine pairs resulted in three clusters grouping the majority of unlikely TB patients and none of the confirmed TB patients in cluster 0 (25 out of 32). Only one unlikely TB patient and none of the unconfirmed TB patients were grouped to cluster 2 (6 confirmed TB, 5 TB infection). Cluster 1 consisted of all four study groups with the majority being TB infected (11 out of 24) (supplementary Figure 5 A-B).

5. Discussion & Conclusion

Diagnosis of childhood TB is one of the key challenges for the global epidemic. As current diagnostic tests are insufficient for detection of TB in children, there is an urgent need for novel tests. Our study is unique as it combines the use of the largest number of novel Mycobacterium tuberculosis antigens and cytokine combinations in a childhood TB di-



(a) Performance of binary classifier using the 10 most important features.



(b) 10 antigen-cytokine pairs contributing the most to performance of trained discriminative classifier.

Figure 2. Effect of normalization of antigen-cytokine pairs and normalization for individual patients. Figure 2a shows the performance of binary classifier using the 10 most important features and applying an antigen-cytokine pair normalization (min-max) and a normalization for individual patients (mean-std). Figure 2b shows the 10 antigen-cytokine pairs contributing the most to performance of trained discriminative classifier with min-max normalization of antigen-cytokine pairs and mean-std individual patient normalization.

agnostic study, exploring the results by applying different machine learning algorithms.

We found that IP-10-responses induced by Rv2346/47c and Rv3614/15c were the two most important features to discriminate diseased from healthy individuals. We showed that further cytokines including GM-CSF, IL-2, IL-6, INF-γ and TNF-α play an important role during immune responses in TB in children. We also demonstrate the importance of data normalization to reduce bias towards highly expressed cytokines and inter-individual heterogeneity in Mycobacterium tuberculosis- specific immune responses. The standard antigens used in the current available test including ESAT-6 and CFP-10 remain important. Our results, however, clearly show that in addition to IFN-γ also IL-6 and TNF-α responses to ESAT-6 and CFP-10 contributed towards distinction of study groups and were among the 10 most important features for the discriminative classifier. Two studies in children also confirm the addition of TNF-α to improve distinction between TB patients and healthy individuals (Tebruegge et al., 2015; 2019).

In our study we demonstrate the impact of normalization on data with improved performance of a discriminative classifier. Performance was best and most robust when both cytokine-antigen concentrations and between-patient values were normalized. IP-10 concentrations induced by Rv2346/47c and Rv3614/15c were found as major contributors to the performance of the discriminative classifier throughout all normalization methods, likely resulting from high concentrations of this cytokine. However, for cytokines that are not expressed at high concentrations, we show that normalization is highly important. For example, IL-2 and IFN-γ concentrations induced by ESAT-6/CFP-10 and Rv3614/15c were only shown to be among the most important features after normalization.

One potential limitation of our study is the sample size which was limited for the two subgroups of TB infection and disease. For optimal training of the classifier and differentiation between TB infection and disease a larger sample size is required. Further studies including a larger number of children are therefore needed to confirm and expand our results. In addition, this study is conducted in a low incidence setting and major factors influencing immune responses such as malnutrition, HIV-infection and other immunocompromising conditions are rare and can therefore not be evaluated.

In conclusion, this is the first study using supervised and unsupervised machine learning algorithms to analyze results from novel Mycobacterium tuberculosis antigens and cytokines for the immunodiagnosis of TB in children. The use of machine learning algorithms is a key tool to evaluate

complex immunological datasets. We identified antigen-cytokine pairs that perform better than the current standard antigen cytokine-pair used in interferon-gamma release assays. These results demonstrate that novel antigen-cytokine pairs have the potential to improve immunodiagnostic tests for tuberculosis in children.

References

Diel, R., Loddenkemper, R., and Nienhaus, A. Evidence-based comparison of commercial interferon-gamma release assays for detecting active TB: a metaanalysis. *Chest*, 137(4):952–968, 2010. ISSN 1931-3543 (Electronic) 0012-3692 (Linking). doi: 10.1378/chest.09-2350. URL http://www.ncbi.nlm.nih.gov/pubmed/20022968.

Dodge, Y. The Oxford Dictionary of Statistical Terms. *Oxford University*, 2006.

Graham, S. M., Cuevas, L. E., Jean-Philippe, P., Browning, R., Casenghi, M., Detjen, A. K., Gnanashanmugam, D., Hesseling, A. C., Kampmann, B., Mandalakas, A., and others. Clinical case definitions for classification of intrathoracic tuberculosis in children: an update. *Clinical Infectious Diseases*, 61(suppl_3):S179–S187, 2015.

Hanley, J. A. and Mcneil, B. J. The Meaning and Use of the Area Under a Receiver Operating Characterstics (ROC) Curve. *Radiology*, 143(1):29–36, 1982.

Hoerl, A. E. and Kennard, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55–67, 1970.

MacQueen, J. Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, (1):281–297, 1967.

Mandalakas, A. M., Detjen, A. K., Hesseling, A. C., Benedetti, A., and Menzies, D. Interferon-gamma release assays and childhood tuberculosis: systematic review and meta-analysis. *The International journal of tuberculosis and lung disease*, 15(8):1018–1032, 2011.

Meier, N. R., Jacobsen, M., Ottenhoff, T. H. M., and Ritz, N. A Systematic Review on Novel Mycobacterium tuberculosis Antigens and Their Discriminatory Potential for the Diagnosis of Latent and Active Tuberculosis. *Front Immunol*, 9:2476, 2018. ISSN 1664-3224 (Electronic) 1664-3224 (Linking). doi: 10.3389/fimmu.2018.02476. URL http://www.ncbi.nlm.nih.gov/pubmed/30473692.

Nemeth, G. O., Nemeth, J., Altpeter, E., and Ritz, N. Epidemiology of childhood tuberculosis in Switzerland between 1996 and 2011. *European journal of pediatrics*, 173(4):457–462, 2014.

Perez-Velez, C. M. and Marais, B. J. Tuberculosis in children. *New England Journal of Medicine*, 367(4):348–361, 2012.

Sollai, S., Galli, L., de Martino, M., and Chiappini, E. Systematic review and meta-analysis on the utility of Interferon-gamma release assays for the diagnosis of Mycobacterium tuberculosis infection in children: a 2013 update. *BMC Infect Dis*, 14 Suppl 1:S6, 2014. ISSN 1471-2334 (Electronic) 1471-2334 (Linking). doi: 10.1186/1471-2334-14-S1-S6. URL http://www.ncbi.nlm.nih.gov/pubmed/24564486.

Tebruegge, M., Dutta, B., Donath, S., Ritz, N., Forbes, B., Camacho-Badilla, K., Clifford, V., Zufferey, C., Robins-Browne, R., Hanekom, W., and others. Mycobacteria-specific cytokine responses detect tuberculosis infection and distinguish latent from active tuberculosis. *American journal of respiratory and critical care medicine*, 192(4):485–499, 2015.

Tebruegge, M., Ritz, N., Donath, S., Dutta, B., Forbes, B., Clifford, V., Zufferey, C., De Rose, R., Robins-Browne, R. M., Graham, S., and others. Mycobacteria-specific mono-and polyfunctional CD4+ T cell profiles in children with latent and active tuberculosis: a prospective proof-of-concept study. *Frontiers in immunology*, 10:431, 2019.

Walzl, G., Ronacher, K., Hanekom, W., Scriba, T. J., and Zumla, A. Immunological biomarkers of tuberculosis. *Nat Rev Immunol*, 11(5):343–354, 2011. ISSN 1474-1741 (Electronic) 1474-1733 (Linking). doi: 10.1038/nri2960. URL http://www.ncbi.nlm.nih.gov/pubmed/21475309.

WHO. Roadmap for childhood tuberculosis towards zero deaths. 2013. URL https://www.who.int/tb/publications/tb-childhoodroadmap/en/.

WHO. *Global tuberculosis report 2018*. World Health Organization, 2018a.

WHO. Roadmap towards ending TB in children and adolescents. 2018b.

A. Methods

A.1. Study design, setting and population

The CITRUS study is registered at ClinicalTrials.gov (NCT03044509) and was approved by the ethics committees of all different study centers and written consent was obtained from all participants and/or their legal representatives (lead ethics Ethikkommission Nordwest- und Zentralschweiz 2016-01094). Baseline characteristics recorded upon enrollment include information on country of birth, age, ethnicity, Bacille Calmette-Guérin immunization status, nutritional status, medical history. Participants are managed according to current Swiss guidelines by the local treating team. The study participants were classified into the following groups confirmed TB (microbiologically confirmed TB disease using culture or nucleic acid amplification techniques), unconfirmed TB (no microbiological proof of TB but symptoms and/or a chest radiograph suggestive of TB and/or routine immunologic evidence of TB by TST ($\geq 5mm$) or positive IGRA), TB infection (criteria for confirmed and unconfirmed TB not met but routine immunologic evidence of TB by TST ($> 5mm$) or positive IGRA), unlikely TB (exposed healthy children that did not meet the above criteria and had no immunologic evidence of TB by routine TST ($\leq 5mm$) or negative IGRA) according to previously published case definitions by (Graham et al., 2015).

A.2. Cytokine measurement

Cytokine concentrations were calculated using a standard curve (5-parameter logistic regression). Concentrations of unstimulated samples served as background and were subtracted from antigen- and mitogen-induced cytokine concentrations. Measurements above or below the limit of quantification (calibration range: 3.2-10'000 pg/ml) were set to 0.1 pg/ml for the lower and to 10'000 pg/ml for the upper limit of quantification.

A.3. Normalization of Data

A min-max normalization, called min-max feature scaling, was applied to the dataset by mapping the entire range of cytokine concentrations to the range [0,1] for every pair. A mean-std normalization, was applied to the dataset by transforming the distribution of cytokine concentrations of every patient to have a mean of zero and a standard deviation of 1.

A.4. Discriminative Classification

L2-regularisation: By adding the L2 norm of the coefficients to the loss function, we reduce the model complexity as the regularization term enforces a shrinking of the coefficients. Large weights in the model are penalized, which helps in reducing overfitting. Five-fold cross-validation: the dataset was split into five equally sized parts, for the model to be trained on four parts and tested on one. This was repeated for all combinations of subsets for a single set of hyperparameters to report a more stable and reliable estimate of the model's predictive power. All data was analyzed using Python (version 3.6) including the python libraries Pandas (version 0.24), Scikit-learn (version 0.21) and Numpy (version 1.16). The plots were generated using the Python library Matplotlib (version 3.0)

B. Additional Figures

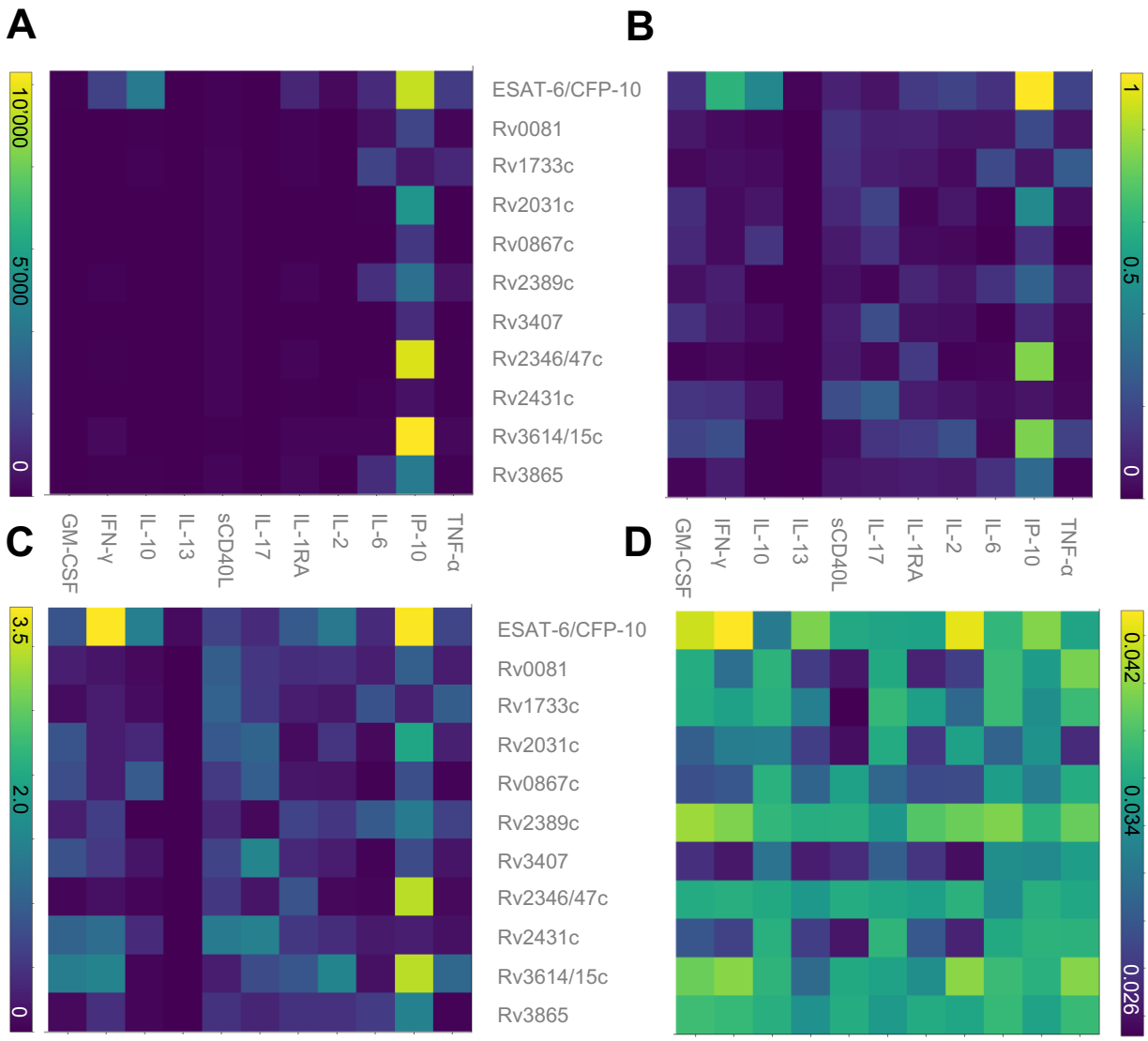


Figure 3. relative median cytokine concentrations, (color change from dark blue to light green indicates an increase in relative cytokine concentration). Non-normalized data (A), min-max normalized data (B), mean-std normalized data (C), min-max normalized and mean-std normalized (between individuals) data (D).

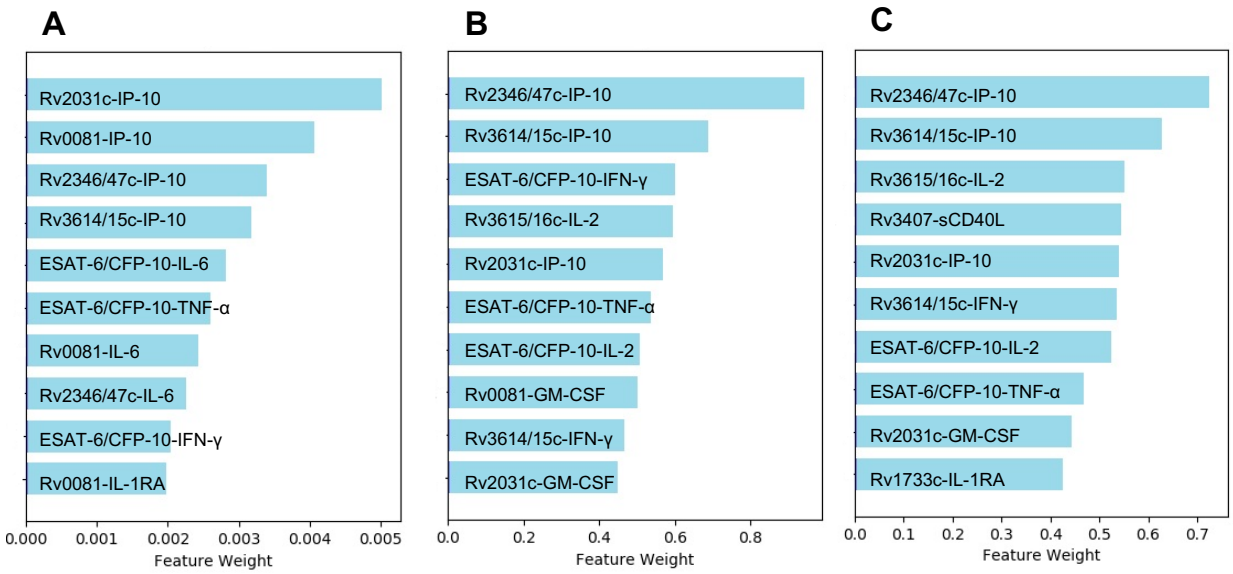


Figure 4. Normalization of data contributes to the performance of a discriminative classifier. Combination of 10 antigen-cytokine pairs contributing the most to the performance of a trained discriminative classifier according to different normalization methods applied: (A) non-normalized data (B) min-max normalized data and (C) mean-std normalized data.

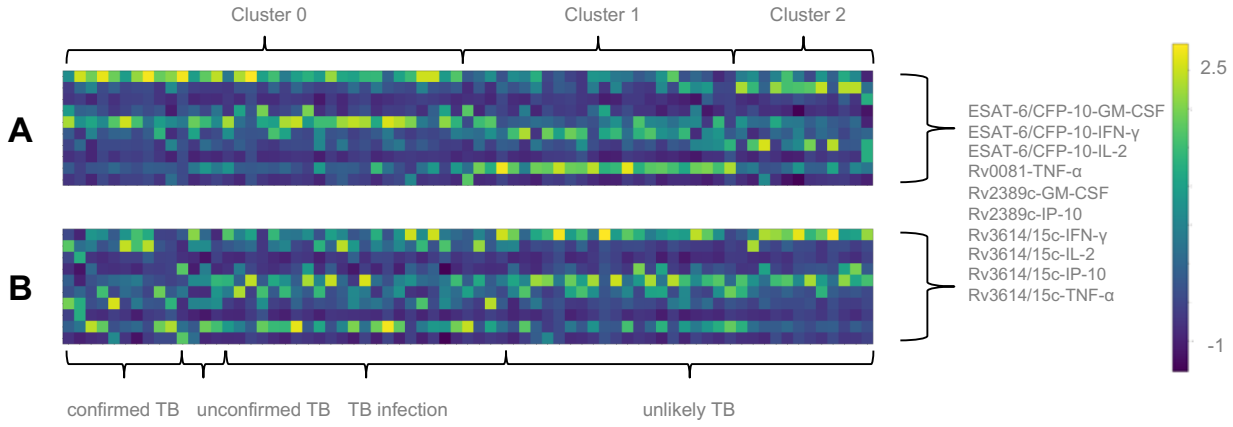


Figure 5. Normalized cytokine concentrations for individual patients (n=71) and selected antigen-cytokine pairs sorted by clusters (A) and study group (B) (color change from dark blue to light green indicates an increase in relative cytokine concentration.