

Research Article

Exploring the Limitations of Peripheral Blood Transcriptional Biomarkers in Predicting Influenza Vaccine Responsiveness

Luca Marchetti,¹ Emilio Siena,² Mario Lauria,^{1,3} Denise Maffione,² Nicola Pacchiani,² Corrado Priami,^{1,3,4} and Duccio Medini²

¹The Microsoft Research-University of Trento Centre for Computational and Systems Biology (COSBI), 38068 Rovereto, Italy

²GSK Vaccines, 53100 Siena, Italy

³Department of Mathematics, University of Trento, Povo, 38123 Trento, Italy

⁴Department of Computer Science, Stanford University, Stanford, CA, USA

Correspondence should be addressed to Luca Marchetti; marchetti@cosbi.eu

Received 26 April 2017; Revised 2 August 2017; Accepted 10 August 2017; Published 28 September 2017

Academic Editor: Min Li

Copyright © 2017 Luca Marchetti et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Systems biology has been recently applied to vaccinology to better understand immunological responses to the influenza vaccine. Particular attention has been paid to the identification of early signatures capable of predicting vaccine immunogenicity. Building from previous studies, we employed a recently established algorithm for signature-based clustering of expression profiles, SCUDO, to provide new insights into why blood-derived transcriptome biomarkers often fail to predict the seroresponse to the influenza virus vaccination. Specifically, preexisting immunity against one or more vaccine antigens, which was found to negatively affect the seroresponse, was identified as a confounding factor able to decouple early transcriptome from later antibody responses, resulting in the degradation of a biomarker predictive power. Finally, the broadly accepted definition of seroresponse to influenza virus vaccine, represented by the maximum response across the vaccine-targeted strains, was compared to a composite measure integrating the responses against all strains. This analysis revealed that composite measures provide a more accurate assessment of the seroresponse to multicomponent influenza vaccines.

1. Introduction

Vaccines represent one of the most effective interventions to control infectious diseases. Despite the many successes [1, 2], however, we are still missing an effective vaccine against current global pandemics such as HIV, malaria, and tuberculosis.

Most vaccines have been developed empirically, against pathogens characterized by limited antigenic variation and that can be neutralized by antibodies alone. Protection against those organisms displaying a fast rate of antigenic variation, such as the HIV virus, or complex host-pathogen interaction biology, such as *Plasmodium falciparum*, requires vaccines that activate multiple arms of the immune response and that do not rely only on the production of neutralizing serum antibodies [3]. Such vaccines will only be accomplished through a careful design involving new-generation

antigens, designed to induce optimal and broadly protective immune responses [4], and adjuvants, which are able to guide the type of adaptive response to produce the most effective forms of immunity for a specific pathogen [5].

Paramount to this novel approach will be the exploitation of new high-throughput technologies combined with the use of advanced algorithms to extract meaningful information from the large sets of generated data. Conventional immunological methods, such as ELISA and flow cytometry, can only assess a limited number of components of the immune system at a given time and, as such, they are not suited for addressing the complexity of the human immune system. Novel technologies, in contrast, allow easily quantifying the abundance of cells, RNA, proteins, and other metabolites across different tissues with high throughput. The availability of high dimensional data, coupled with computational modeling, holds the potential to provide new insights into

TABLE 1: Transcriptional data sets used in this study.

Set	Accession	Reference	Samples (subjects)	Platform	Description
Set #1	GSE29614	Nakaya et al., Nat Immunol 2011	27 (9)	Affymetrix U133 Plus 2.0	Time course of young adults vaccinated with influenza TIV during 2007/08 flu season
Set #2	GSE29617	Nakaya et al., Nat Immunol 2011	80 (28)	Affymetrix U133 Plus 2.0	Time course of young adults vaccinated with influenza TIV during 2008/09 flu season
Set #3	GSE74817	Nakaya et al., Immunity 2015	621 (212)	Affymetrix U133 Plus PM	Time course of adults vaccinated with influenza TIV during the years 2009–2012
Set #4	GSE48018	Bucasas et al., J Infect Dis 2011	431 (116)	Illumina HumanHT-12 V3.0	Time course of adults vaccinated with influenza TIV

the mechanisms underlying the immunological response to vaccination [6].

In addition to being foundational to a rational vaccine design, an increased understanding of the mechanisms involved in the response to vaccination will also guide the optimization of immunogenicity and reactogenicity profiles of existing vaccines. Also, deciphering the early events following vaccination will provide molecular signatures predictive of vaccine efficacy or safety which, in turn, will enable a prospective identification of subjects with a suboptimal response profile and speed up future clinical trials [7].

Recently, several studies [8–11] have analyzed the transcriptome profile of peripheral blood, or blood-derived cells, to study immunity to the trivalent influenza vaccine (TIV) in humans. These studies decoded the global pattern of transcriptional response to TIV vaccination and independently reported that the extent of upregulation of a set of interferon-inducible genes, one to three days after vaccination, and of genes involved in plasmablasts differentiation and activity, seven days after vaccination, correlates with the magnitude of serum functional antibody titers measured after one month. Among those, Nakaya and colleagues [10, 11] extended this further, by applying a machine learning algorithm to identify sets of genes that could predict subjects’ seroresponse across independent TIV vaccination trials.

Despite their extensive use, most machine learning approaches only provide information in the form of disease- or treatment-specific gene signatures, along with their associated predictive power. This limited set of information hardly provides any supporting evidence in the investigation of the causes of an incorrect prediction. This is especially true for vaccine response prediction, due to the complexity and multicomponent nature of the immune system and the genetic heterogeneity across patients.

With this study we sought to expand our understanding of the early events following vaccination by looking for transcriptional signatures, derived from peripheral blood mononuclear cells (PBMCs), whose combined response is associated with the magnitude of functional antibody responses measured four weeks after vaccination. Using two sets of published profiles covering two vaccination campaigns (2007/2008 and 2008/2009, resp., for set #1 and set #2, see Table 1), we identified a biomarker consisting of 207 transcripts whose early (3–7 days after vaccination) response was capable of prospectively discriminating low- from high-responder subjects. Following best practices in biomarker

identifications, we used one cohort (2008/2009 campaign) for training and a different one (2007/2008 campaign) for validation. The good performance of the identified biomarker was further confirmed using a third set of profiles from a separate study including vaccination campaigns spanning a total of three years (2009/2010/2011, set #3). Despite the consistently high prediction accuracy achieved by the biomarker across cohorts, we noticed that not all the subjects were correctly classified. We investigated the possible reasons for the observed misclassifications and identified preexisting immunity against one or more of the vaccine antigens to be a confounding factor. While this finding confirms earlier reports on preexisting immunity negatively affecting the seroresponse to vaccination, we additionally identified a difference in the response between subjects with and without preexisting immunity. Specifically, we found preexisting immunity to affect the 24 h response and to have little to no effect on whole blood transcriptomes profiled 3 or 14 days after vaccination. Therefore, preexisting immunity can decouple early transcriptome from later antibody responses, resulting in the degradation of a biomarker predictive power. This finding was confirmed using a fourth set of profiles (set #4).

We finally related our predictions to a composite HAI response measure, which integrates the responses to all strains represented in the vaccine, and showed how this can be a more informative measure for influenza vaccine responsiveness compared to the widely accepted maximum-across strains one.

2. Results and Discussion

We analyzed a total of four transcriptome datasets, whose essential features are reported in Table 1.

2.1. A Biomarker of 207 Transcripts, Primarily Involved in Cell Proliferation and Cytokine Signaling, Predicts Seroresponse to Seasonal Influenza Vaccination. We employed an optimized version of SCUDO [12–15], an algorithm for signature-based clustering of expression profiles that relies on a completely new way of addressing the classification problem. The algorithm is based on the concept of subject-specific signatures, rather than disease-specific signatures, to divide samples. Briefly, the method first seeks to summarize the characteristics of each sample by means of a subject-specific, rank-based signature and then it performs a systematic, all-to-all signature comparison to segregate samples into different

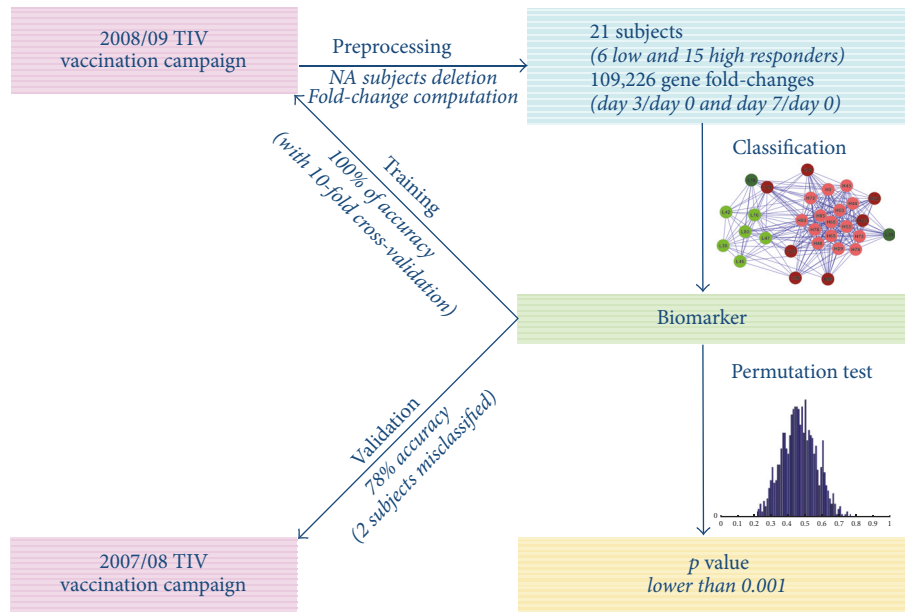


FIGURE 1: Graphical representation of the analysis pipeline implemented to compute biomarkers (see Materials and Methods). The analysis started with the dataset related to the 2008/09 TIV vaccination campaign (set #2). After a preliminary preprocessing of the dataset for removing subjects with intermediate answer to the vaccine (“NA subjects,” maximum $\text{HAI}_{\text{Day 28/Day 0}}$ across strains equal to 4) and for computing gene fold changes (day 3/day 0 and day 7/day 0), we applied the classification algorithm SCUDO to obtain the biomarker. The classification process on the training dataset exhibited 100% accuracy with a 10-fold cross-validation scheme. Moreover, its statistical significance was assessed by a permutation test providing a p value < 0.001 . Finally, the identified biomarker was validated using an independent dataset (2007/08 TIV vaccination campaign, set #1). The classification on the validation dataset exhibited 78% accuracy.

groups. The result of the comparison can be represented in the form of a graph, which allows for a qualitative, as opposed to solely quantitative, interpretation of the classification performance. Therefore, the algorithm combines the benefits of rank-based classification, dimensionality reduction, and the power of network analysis. The computation of subject-specific signatures also allows the characterization of specific differences between subjects of the same class by highlighting intraclass variability that may cause incorrect predictions. Subject-specific signatures can then be merged together to obtain a unique disease-specific biomarker.

After stratifying vaccines into high responders (maximum hemagglutinin inhibition (HAI) titer fold increase > 4) and low responders (maximum HAI fold increase < 4) according to [10], we applied SCUDO for predicting subjects’ membership to the correct category, based on transcriptome response profiles derived from PBMCs (see Materials and Methods and Figure 1). SCUDO produced a biomarker of 207 transcripts whose regulation 3 or 7 days after immunization was able to predict seroconversion with 100% and 78% accuracy within the training (2008-2009 trial, set #2) and an independent validation (2007-2008 trial, set #1) sets, respectively (see Supplementary File S1 in Supplementary Material available online at <https://doi.org/10.1155/2017/3017632>). The length of the biomarker produced by the algorithm was optimized to maximize both the classification accuracy and the amount of information about the mechanisms responsible for seroconversion. In a previous work [12] we showed that the length of the signatures is not critical, because often

the same level of classification accuracy can be obtained with a range of signature sizes. For this reason, we also sought to identify the minimal set of genes that could predict vaccination outcome without a significant loss in the prediction accuracy compared to the identified biomarker. The algorithm identified a set of 12 transcripts, which showed prediction accuracy comparable to that of the longer biomarker (see Supplementary File S2). Interestingly, all 12 selected transcripts were also included in the 207 transcripts set. Similarly to the 207 transcripts biomarker, this minimal biomarker relied on the fold change of some transcripts at day 3 and of some others at day 7 after immunization. We verified that transcript regulation at day 3 or day 7 alone is not enough for an accurate subject classification (not shown).

One hundred and sixty-seven (167) of the 207 transcripts could be mapped into functionally annotated genes (see Supplementary File S1). Functional analysis revealed that the biomarker is enriched for genes involved in cell proliferation (IFIT3, PML, PTPN6, PTPRU, TBGR1, WNK2, and ZEB1) and in cytokine signaling (DDX58, IFIT3, PIK3R1, PIK3R5, PML, PTPN6, and SKP1), supporting previous evidence of cell- and interferon-mediated immune responses to be linked to humoral responses to influenza vaccination. Of note, also the CAMK4 kinase, which was experimentally shown to be involved in the regulation of antibody response [10], was included in the biomarker.

2.2. Preexisting Immunity Interferes with Response Outcome Predictability. Subject classification was based on a similarity

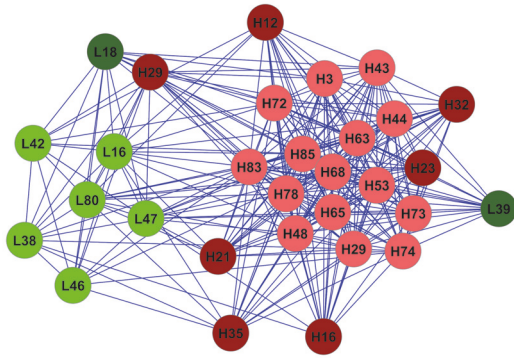


FIGURE 2: *Prediction of seroresponse to TIV vaccination.* The graph represents sample classification by means of the 207 transcripts biomarker. Nodes represent subjects and the length of connecting edges is proportional to the distance between subject signatures. Node colors indicate subjects' class (green for low responders and red for high responders); darker colors represent samples included in the validation set (2007-2008 TIV campaign, set #1), while lighter colors represent samples in the training set (2008-2009 TIV campaign, set #2). Node labels indicate the sample class (H for high responders and L for low responders) and sample identifier.

map that could be represented in the form of a graph in which nodes correspond to subjects and edge length encodes the level of similarity between subject-specific signatures (short edge = high similarity, long edge = low similarity, and no edge = negligible similarity). This provided a graphical representation of the underlying sample structure (Figure 2) showing how the biomarker successfully stratified the tested subjects into high responders and low responders. One of the two misclassified subjects, subject H29, showed some similarity with members from both clusters, reflecting a suboptimal signature performance. The other misclassified subject, subject L39, was closely associated with the high responders group despite its modest HAI response. This same subject was consistently misclassified also by the 12 transcripts biomarker.

As a further validation of the identified biomarker, we tested it on an additional set of transcriptional profiles derived from three independent vaccination campaigns. Specifically, we used the testing cohort described in Nakaya et al. [11], which is related to the TIV vaccination campaigns of 2009, 2010, and 2011 (set #3). The testing cohort includes 21 subjects out of a total of 212, selected because only for them the classification in high/low responders was available. Despite the fact that none of the subjects from these vaccination campaigns were used for training and that the data was collected as part of a different study, we obtained a classification accuracy of 81% (4 out of 21 subjects were misclassified; see Figure 3), which is in close agreement with what was reported in the original study [11].

Motivated by the consistently good performance of the biomarker across cohorts, we conducted an investigation on the possible reasons explaining why some subjects escape correct classification. Focusing on subject L39 as a representative case, we identified a correlation with an unusually high

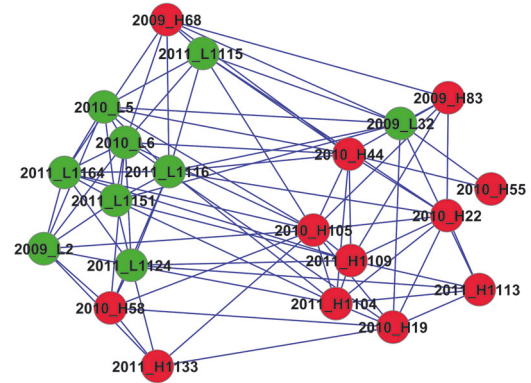


FIGURE 3: *Additional validation of the 207 transcripts biomarker.* The graph represents samples classification by means of the 207 transcripts biomarker of subjects included in the testing group of [11], which is related to the TIV vaccination campaigns of 2009, 2010, and 2011 (set #3). Nodes represent subjects and the length of connecting edges is proportional to the distance between subjects' signatures. Node colors indicate subjects' class (green for low responders and red for high responders). Node labels indicate the year of the vaccination campaign, sample class (H for high responders and L for low responders), and sample identifier. Four out of 21 subjects were misclassified (subjects 2009_H68, 2010_H58, 2011_H113, and 2009_L32).

level of preexisting antibodies against the strain H3N2 (HAI titer = 2560), which possibly interfered with the subject's seroresponse. This particular evidence was in line with the several independent studies that reported that the existence of preexisting immunity (defined as the presence of preimmune antigen-specific antibodies) is associated with a reduced vaccine responsiveness [16–18]. Subject L39's proximity to the high responders cluster (Figure 2) indicates that these experienced a similar transcriptional response to the vaccine stimulus, suggesting that the seroresponse inhibition played by preexisting immunity depends on mechanisms that are not captured by the day 3 or day 7 PBMCs transcriptomes.

In order to substantiate this conclusion, which was derived from a single subject, we run a further investigation aimed at characterizing the influence of preimmune antigen-specific antibodies on the transcriptional response to influenza vaccination. Whole Blood (WB) transcriptomes, assessed in a pool of 119 healthy individuals vaccinated with a 2008-2009 trivalent influenza vaccine [8], were used for this analysis (set #4). From the initial pool, subjects were selected and stratified based on their baseline antigen-specific antibody level. Fifteen (15) subjects were found to have negligible preexisting immunity to the vaccine (maximum HAI titer across the three antigens ≤ 16) while 20 showed signs of preexposure to one or more of the vaccine antigens (maximum HAI titer across the three antigens ≥ 512). Transcriptome responses (fold changes from baseline measured 1, 3, and 14 days after vaccination) were then compared among the two groups. Due to inconsistencies (use of WB rather than PBMCs and lack of day 7 transcriptome profiling) between this and the previously introduced studies, it was not possible

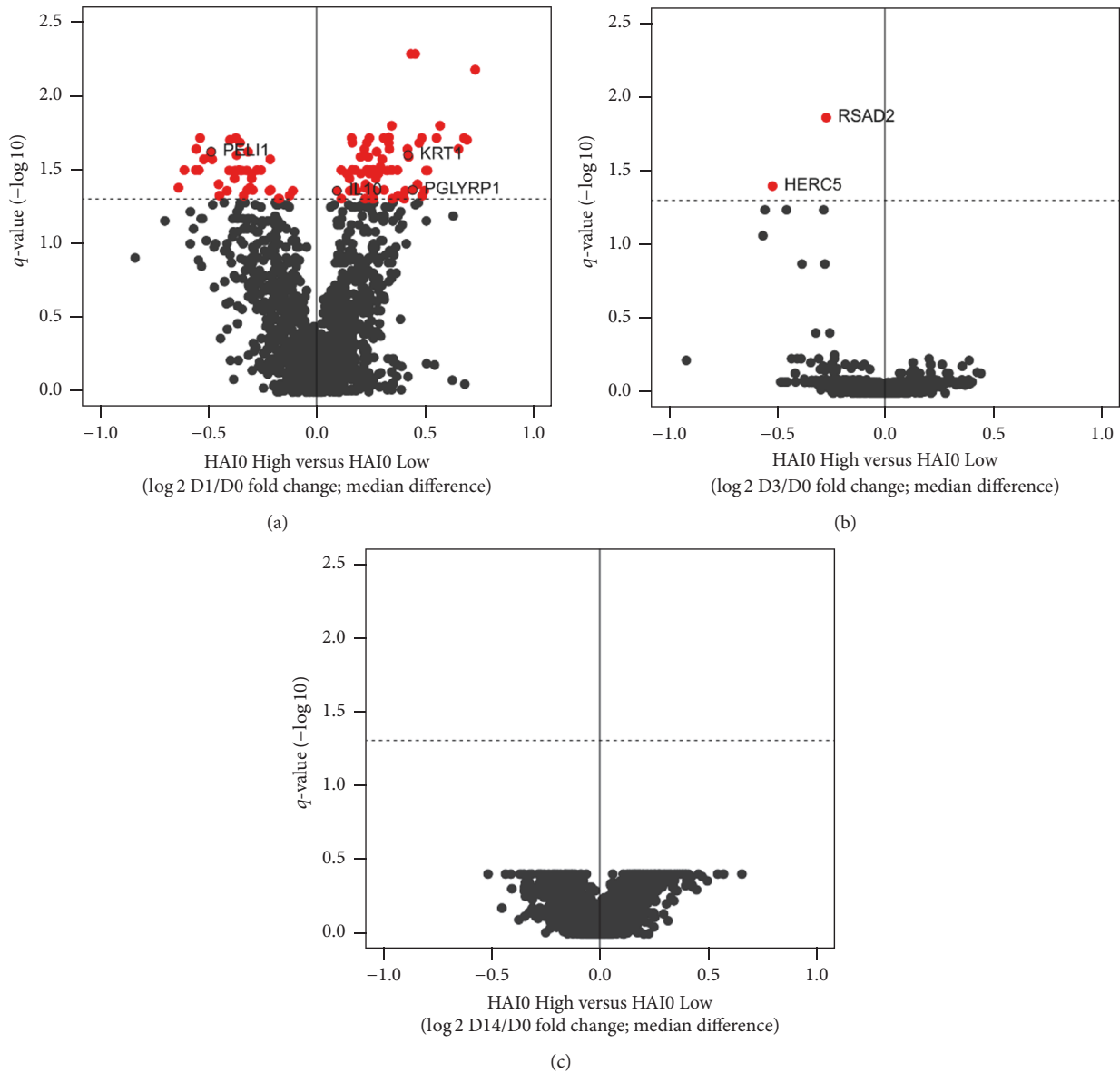


FIGURE 4: Differentially responsive genes in subjects with low and high preexisting immunity to the influenza vaccine, set #4. Volcano plots showing the differences in gene responsiveness (fold change from baseline) between subjects with high (High) and low (Low) baseline antibody titers ($HAI_{Day 0}$). Y-axes represent the q -values computed by adjusting Wilcoxon test p values using the Benjamini-Hochberg procedure. X-axes represent the difference in the median fold change from baseline response between subjects with high and low baseline immunity. Results are reported for day 1 (a), day 3 (b), and day 14 (c) time points.

to apply SCUDO to this dataset for confirmatory purposes. We note that our use of WB gene expression profiles is not ideal to confirm a hypothesis originally formulated while observing PBMC profiles. However previous studies have shown that there is a substantial overlap between the transcriptional responses for these two cell lines for a number of conditions similar to ours (see, e.g., Bondar et al. [19], for the case of inflammatory response). Thus the general evolution of transcriptional profiles in WB following vaccination is highly suggestive of those in PBMC and provides strong support if not conclusive evidence for our

hypothesis of negligible influence of preexisting immunity on transcriptional response at days 3–7 after vaccination.

Differential gene expression analysis identified 118 transcripts whose response 1 day following vaccination differed among subjects with high and low preimmune status (Figure 4(a)). Among these, 77 were found to be more responsive in subjects with high baseline immunity, while 41 were less responsive (see Supplementary File S3). Functional analysis of these genes did not result in any enriched gene ontology (data not shown). Nonetheless, multiple genes participating in the inflammatory response (IL10, KRT1, PELI1,

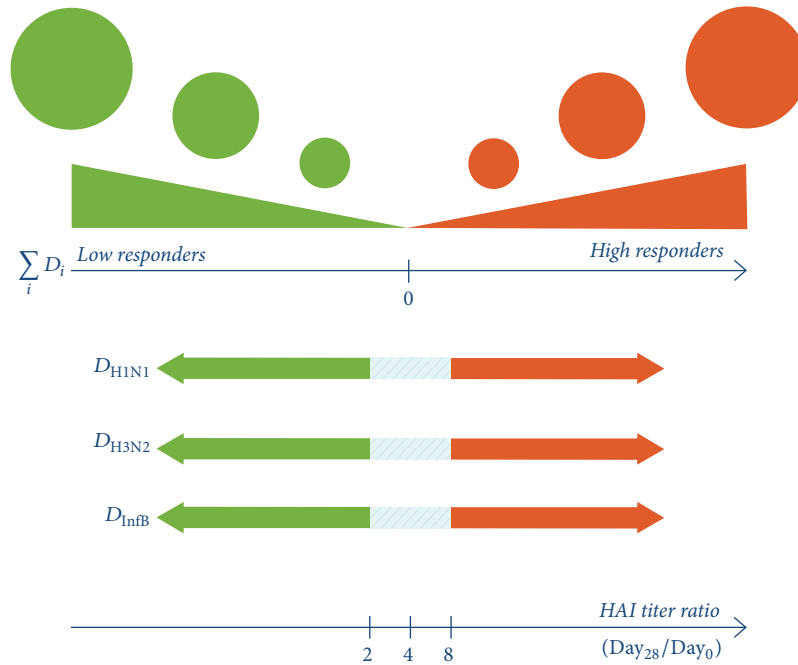


FIGURE 5: *Rationale for the computation of the titer response score.* The score integrates the responses obtained with the three strains represented in the TIV (H1N1, H3N2, and influenza B) by summing the logarithmic deviations of the $HAI_{Day\ 28/Day\ 0}$ ratios from the threshold value of the class to which the subject belongs (2 for low responders, 8 for high responders). The circles of different radius on the top of the figure represent how the TRS was encoded as node size in Figure 6.

and PLYRPI) were present. Interestingly, IL10, KRT1, and PLYRPI, all negative regulators of inflammatory response, were more responsive in subjects with preexisting immunity, while PELI1, a TRIF-dependent Toll-like receptor agonist, was less responsive in this group (Figure 4(a)). This specific response pattern suggests that preexisting immunity may act as negative feedback on innate immune activation when triggered by a recurring antigen. Differently, later time points (≥ 3 days after vaccination) did not show such a strong effect (Figures 4(b) and 4(c)). Only two interferon-inducible genes, HERC5 and RSAD2, were found to be more responsive in individuals with no preexisting immunity 3 days after vaccination.

Overall, these findings are in agreement with our initial hypothesis that while preexisting immunity can inhibit seroresponse to influenza vaccination, it does not affect peripheral blood-derived transcriptome profiles assessed 3 or more days after vaccination. This directly translates in the inability of transcriptional biomarkers, based on information collected from peripheral bloodstream 3 or more days after vaccination, to correctly account for the preimmune status of a vaccinee.

It must be noted, however, that most of the subjects (28 out of 30) employed in this specific analysis were sero-responders (maximum $HAI_{Day\ 28/Day\ 0}$ fold increase across strains > 4), making the proper assessment of transcriptional differences between responder and nonresponder subjects in case of presence or absence of preexisting immunity not possible.

2.3. Maximum HAI Response across Strains Is an Oversimplified Measure for Assessing Influenza Vaccine Responsiveness. The outcome of a classification strongly depends on the definition of the response categories. For this reason we devised a titer response score (TRS; see Materials and Methods, Figure 5 and Supplementary File S4) that integrated the responses obtained from all the three strains represented in the TIV and investigated how this is related to the maximum response across strains. Figure 6 represents TRS responses, encoded as node radius, mapped on the classification graph of Figure 2. According to the TRS, subjects showing more extreme responses will be assigned higher scores and, therefore, will appear as bigger nodes in the graph. Interestingly, subjects with high TRS were placed in peripheral regions of the graph, with smaller nodes populating more central regions. This is indicative of the fact that subjects with high TRS were better discriminated (greater distance between high and low responder groups) by the biomarker. Coherently with these observations, both misclassified subjects (H29 and L39) had a small TRS, indicating that their membership in their predicted class, the wrong one, was not strong. Subject H29, which was classified as high responder based on the response relative to the H3N2 strain ($HAI_{Day\ 28/Day\ 0} = 32$), did not respond as strongly against the other two strains (see Supplementary File S4). Similarly, while most of the low responders did not respond to any of the strains ($HAI_{Day\ 28} = HAI_{Day\ 0}$), subject L39 showed a 2-fold increase in the HAI titer against the H1N1 strain (see Supplementary File S4).

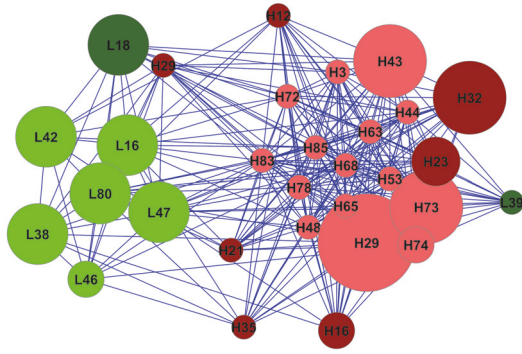


FIGURE 6: *Prediction of seroresponse to TIV vaccination.* The graph represents sample classification by means of the 207 transcripts biomarker as in Figure 2, but here the node size encodes the titer response score (bigger nodes correspond to higher indices, as displayed in Figure 5). Nodes represent subjects and the length of connecting edges is proportional to the distance between subject signatures. Node colors indicate subjects class (green for low responders and red for high responders); darker colors represent samples included in the validation set (2007-2008 TIV campaign, set #1), while lighter colors represent samples in the training set (2008-2009 TIV campaign, set #2). Node labels indicate the sample class (H for high responders and L for low responders) and the sample identifier.

Overall, this evidence indicates that while using the maximum HAI fold change across strains may be accurate enough for the discrimination between high responders and low responders, composite measures integrating information from all strains represented in the vaccine provide a more informative measure.

3. Conclusions

Overall, we have identified a set of 207 transcripts, derived from PBMCs, whose early (3–7 days) regulation after vaccination was predictive of the magnitude of later (1 month) serum antibody response. Overall, this biomarker performed similarly to the one described in the original work published by Nakaya et al. [10], by providing a 100% within-dataset classification accuracy and a cross-datasets accuracy of 81% (see Figure 3). Moreover, the applied classification method provided a graphical representation of the classification result, which enabled for a qualitative, as opposed to solely quantitative, interpretation of the results obtained. This provided the opportunity to investigate on the possible reasons behind the observed misclassifications and allowed identifying preexisting immunity, against one or more of the vaccine antigens represented in the vaccine, to be a possible confounding factor. Specifically, preexisting immunity, which was found to negatively affect the seroresponse, was shown to have little to no effect on whole blood transcriptome profiles assessed 3, or more, days after vaccination. This implies a decoupling of early transcriptome from later antibody responses, highlighting a major limitation in transcriptome-based biomarkers.

Differently from 3 and later days responses, 24-hour whole blood transcriptomes revealed substantial differences in gene responsiveness between subjects with preexisting

immunity and subjects without it. Part of these differences included a decreased responsiveness of the transcriptional inflammatory response in subjects with preexisting immunity, providing a first clue of a possible mechanistic link between preimmune HAI titers and reduced serological response to influenza vaccination.

The present study represents an example of how innovative computational tools can improve our data mining capabilities and helps to reveal latent factors that can impact the response to vaccination. We envision that extending this analysis pipeline to other studies will help in identifying additional confounding factors and produce more accurate predictions.

4. Materials and Methods

Definition and testing of the biomarkers were based on data published by Nakaya et al. [10, 11]. Gene expression derived from peripheral blood mononuclear cells and HAI response data were downloaded from the Gene Expression Omnibus repository (GEO, <https://www.ncbi.nlm.nih.gov/geo>) using the two accession identifiers GSE29614 and GSE29617 for the 2007-2008 and 2008-2009 vaccination trials, respectively. Gene expression data were imported using the ArrayExpress Bioconductor package and processed using the RMA normalization procedure. Gene-level expression data were derived by computing the geometric mean of multiple probes mapping to the same gene, where applicable. The dataset related to the additional validation presented in Figure 3 was downloaded from the GEO repository using the accession identifier GSE74817. Analysis of differential gene expression between subjects with high and low baseline HAI titers was performed on the dataset published in [8] and downloaded from GEO using the accession identifier GSE48018. In that study, 119 healthy individuals vaccinated with a 2008-2009 trivalent influenza vaccine were profiled for both whole blood transcriptome and HAI responses. Fifteen (15) subjects had a preimmune HAI titer ≤ 16 (maximum across strains) and were considered to be naive to the vaccine. In order to generate a contrast group comparable in size, 20 subjects with a HAI titer ≥ 512 (maximum across strains) were selected as high-preimmune individuals. After removing genes that did not show substantial variation across the entire dataset (\log_2 interquartile range ≤ 0.5), fold changes from baseline in gene expression were computed for both groups 1, 3, and 14 days after vaccination. Gene responsiveness was then compared between the two groups through the Mann-Whitney-Wilcoxon test and resulting p values were corrected for multiple testing using the Benjamini-Hochberg procedure. Transcripts with a q -value ≤ 0.05 were assumed to be affected by preexisting immunity.

According to [10], seroresponder and nonseroresponder subjects were identified based on whether they achieved a maximum $\text{HAI}_{\text{Day 28}/\text{Day 0}}$ fold increase across strains > 4 or < 4 , respectively. Afterwards, day 3 and day 7 fold changes from baseline transcript abundances were employed for the prediction of subjects' class membership using an enhanced version of the previously described classification

algorithm SCUDO [12, 13, 15]. The present version of SCUDO has been extended to increase its discriminatory power and equipped with a parameter optimizer that allows the automatic selection of the algorithm parameters according to user-defined criteria as we did to compute the transcriptional biomarker presented in [20, 21].

SCUDO is a rank-based classification method and this makes the computation normalization-free. Therefore, we did not apply any data preprocessing with batch effect removal algorithms (e.g., quantile normalization) before running SCUDO [12]. This feature of the classification algorithm played a crucial role in the additional validation presented in Figure 3, in which samples from separate datasets could be tested without taking batch effects into account.

The classification accuracy of the analysis was evaluated using a 10-fold cross-validation scheme over the training dataset (2008-2009 trial) and by validating over another independent dataset (2007-2008 trial). Statistical significance of the identified biomarkers was assessed by comparing the classification accuracy of the method with the accuracy of an empirical null distribution obtained by 10000 random permutations of the transcript labels (permutation test p value < 0.001). An overview of the analysis pipeline is provided in Figure 1. The robustness of the identified 207 transcripts biomarker was further assessed by applying it to independent TIV vaccination campaigns (2009, 2010, and 2011 campaigns included in GSE74817).

The titer response score (TRS; Figures 5 and 6 and Supplementary File S4) was generated by integrating the responses obtained with all the three strains represented in the TIV vaccine. Differently from a previously proposed integrated antibody response measure [8], the TRS of each subject was computed as the sum of the logarithmic deviations of the three considered $HAI_{Day\ 28}/Day\ 0$ ratios from the threshold value of the class to which the subject belongs (2 for low responders, 8 for high responders). An additional penalty was added for subjects with high (>160) baseline HAI titers. Additional information can be found in the Supplementary File S4.

Conflicts of Interest

Emilio Siena, Nicola Pacchiani, and Duccio Medini are employees of GSK Vaccines s.r.l. The other authors do not have any conflicts of interest.

Authors' Contributions

Luca Marchetti and Emilio Siena contributed equally to this work.

References

- [1] J. G. Breman and I. Arita, "The certification of smallpox eradication and implications for guinea worm, poliomyelitis, and other diseases: Confirming and maintaining a negative," *Vaccine*, vol. 29, no. 4, pp. D41–D48, 2011.
- [2] J. Smith, R. Leke, A. Adams, and R. H. Tangermann, "Certification of polio eradication: Process and lessons learned," *Bulletin of the World Health Organization*, vol. 82, no. 1, pp. 24–30, 2004.
- [3] R. Rappuoli, "Bridging the knowledge gaps in vaccine design," *Nature Biotechnology*, vol. 25, no. 12, pp. 1361–1366, 2007.
- [4] P. R. Dormitzer, J. B. Ulmer, and R. Rappuoli, "Structure-based antigen design: a strategy for next generation vaccines," *Trends in Biotechnology*, vol. 26, no. 12, pp. 659–667, 2008.
- [5] D. T. O'Hagan and C. B. Fox, "New generation adjuvants - From empiricism to rational design," *Vaccine*, vol. 33, no. 2, pp. B14–B20, 2015.
- [6] B. Pulendran, S. Li, and H. I. Nakaya, "Systems vaccinology," *Immunity*, vol. 33, no. 4, pp. 516–529, 2010.
- [7] R. Rappuoli and A. Aderem, "A 2020 vision for vaccines against HIV, tuberculosis and malaria," *Nature*, vol. 473, no. 7348, pp. 463–469, 2011.
- [8] K. L. Bucasas, L. M. Franco, C. A. Shaw et al., "Early patterns of gene expression correlate with the humoral immune response to influenza vaccination in humans," *Journal of Infectious Diseases*, vol. 203, no. 7, pp. 921–929, 2011.
- [9] G. Obermoser, S. Presnell, K. Domico et al., "Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines," *Immunity*, vol. 38, no. 4, pp. 831–844, 2013.
- [10] H. I. Nakaya, J. Wrammert, E. K. Lee et al., "Systems biology of vaccination for seasonal influenza in humans," *Nature Immunology*, vol. 12, no. 8, pp. 786–795, 2011.
- [11] H. I. Nakaya, T. Hagan, S. S. Duraisingham et al., "Systems Analysis of Immunity to Influenza Vaccination across Multiple Years and in Diverse Populations Reveals Shared Molecular Signatures," *Immunity*, vol. 43, no. 6, pp. 1186–1198, 2015.
- [12] M. Lauria, "Rank-based transcriptional signatures: A novel approach to diagnostic biomarker definition and analysis," *Systems Biomedicine*, vol. 1, pp. 35–46, 2013.
- [13] M. Lauria, "Rank-based miRNA signatures for early cancer detection," *BioMed Research International*, vol. 2014, Article ID 192646, 2014.
- [14] A. L. Tarca, M. Lauria, M. Unger et al., "Strengths and limitations of microarray-based phenotype prediction: lessons learned from the IMPROVER Diagnostic Signature Challenge," *Bioinformatics*, vol. 29, no. 22, pp. 2892–2899, 2013.
- [15] M. Lauria, P. Moyses, and C. Priami, "SCUDO: A tool for signature-based clustering of expression profiles," *Nucleic Acids Research*, vol. 43, no. 1, pp. W188–W192, 2015.
- [16] W. E. P. Beyer, A. M. Palache, M. J. W. Sprenger et al., "Effects of repeated annual influenza vaccination on vaccine sero-response in young and elderly adults," *Vaccine*, vol. 14, no. 14, pp. 1331–1339, 1996.
- [17] S. Sasaki, X.-S. He, T. H. Holmes et al., "Influence of prior influenza vaccination on antibody and B-cell responses," *PLoS ONE*, vol. 3, no. 8, Article ID e2975, 2008.
- [18] S. Ng, D. K. M. Ip, V. J. Fang et al., "The Effect of Age and Recent Influenza Vaccination History on the Immunogenicity and Efficacy of 2009-10 Seasonal Trivalent Inactivated Influenza Vaccination in Children," *PLoS ONE*, vol. 8, no. 3, Article ID e59077, 2013.
- [19] G. Bondar, M. Cadeiras, N. Wisniewski et al., "Comparison of whole blood and peripheral blood mononuclear cell gene expression for evaluation of the perioperative inflammatory response in patients with advanced heart failure," *PLoS ONE*, vol. 9, no. 12, Article ID e115097, 2014.
- [20] L. Caberlotto, L. Marchetti, M. Lauria, M. Scotti, and S. Parolo, "Integration of transcriptomic and genomic data suggests candidate mechanisms for APOE4-mediated pathogenic action in

Alzheimer's disease," *Scientific Reports*, vol. 6, Article ID 32583, 2016.

- [21] A. Matone, E. Derlindati, L. Marchetti et al., "Identification of an early transcriptomic signature of insulin resistance and related diseases in lymphomonocytes of healthy subjects," *PLOS ONE*, vol. 12, no. 8, p. e0182559, 2017.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

