Prediction of Specific Antibody- and Cell-mediated Responses using Baseline Immune Status Parameters of Individuals Received Measles-mumps-rubella Vaccine

Anna Pavlovna Toptygina , Dmitry Grebennikov , Gennady Bocharov *

Posted Date: 3 January 2023

doi: 10.20944/preprints202301.0027.v1

Keywords: Measles-mumps-rubella; post-vaccination immune response; correlation analysis; multivariable linear regression; predictors; antibody-mediated response; cell-mediated response.

Article

Prediction of Specific Antibody- and Cell-Mediated Responses Using Baseline Immune Status Parameters of Individuals Received Measles-Mumps-Rubella Vaccine

Anna P. Toptygina [1,*], Dmitry Grebennikov [2,3,4,*] and Gennady Bocharov[2,3,5,*]

[1] Gabrichevsky Research Institute for Epidemiology and Microbiology, Moscow, Russian Federation; toptyginaanna@rambler.ru (A.T.)
[2] Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, (INM RAS), Moscow, Russian Federation; g.bocharov@inm.ras.ru (G.B.)
[3] Moscow Center for Fundamental and Applied Mathematics at INM RAS, Moscow, Russian Federation;
[4] World-Class Research Center Digital Biodesign and Personalized Healthcare, Sechenov First Moscow State Medical University, Moscow, Russian Federation; dmitry.ew@gmail.com (D.G.)
[5] Institute of Computer Science and Mathematical Modelling, Sechenov First Moscow State Medical University, Russian Federation
* Correspondence: toptyginaanna@rambler.ru (A.T.); dmitry.ew@gmail.com (D.G.); g.bocharov@inm.ras.ru (G.B.)

Abstract: The successful vaccination implies the induction of effective specific immune responses. We intend to find biomarkers among various immune cell subpopulations, cytokines and antibodies which could be used to predict the levels of specific antibody- and cell-mediated responses after measles-mumps-rubella vaccination. We measured 59 baseline immune status parameters (frequencies of 42 immune cell subsets, levels of 13 cytokines, immunoglobulins) before vaccination and 13 response variables (specific IgA and IgG, antigen-induced IFN-γ production, CD107a expression on $CD8^+$ T lymphocytes, and cellular proliferation levels by CFSE dilution) 6 weeks after vaccination for 19 individuals. Statistically significant Spearman correlations between some baseline parameters and response variables were found for each response variable ($p<0.05$). Due to the low number of observations relative to the number of baseline parameters and missing data for some observations, we utilized three feature selection strategies to select potential predictors of the post-vaccination responses among baseline variables: (a) screening of the variables based on the correlation analysis, (b) supervised screening based on the information of changes of baseline variables at day 7, (c) implicit feature selection is performed using regularization-based sparse regression. We identified optimal multivariate liner regression models for predicting the effectiveness of vaccination against measles-mumps-rubella using the baseline immune status parameters. It turned out that the sufficient number of predictor variables ranges from one to five depending on the response variable of interest.

Keywords: Measles-mumps-rubella; post-vaccination immune response; correlation analysis; multivariable linear regression; predictors; antibody-mediated response; cell-mediated response

1. Introduction

In the beginning of XXI century, the World Health Organization (WHO) announced the measles eradication program. Despite tremendous effort to practically implement the plan, the targeted measles elimination was not achieved by 2015. The current objective is set to eliminate the measles by 2025. However, in Europe characterized by high vaccine coverage in years 2017-2018 several measles resurgence took place with several thousand cases reported [2–4]. Since pre-vaccination era, it is known that natural measles infection induces lifelong protective immunity. It is considered that measles vaccine also induces high protective immunity. However, the effectiveness and durability of

 doi:10.20944/preprints202301.0027.v1

the protection varies between vaccinated individuals. Besides the humoral immunity, the virus-specific cellular immunity influences the protection against measles. It has been documented that host genetics contribute to inter-individual variability of protection, i.e., there are HLA allelic associations with either a high antibody-mediated immunity or a strong cellular immunity [5]. The SNPs in genes controlling the cytokine production, e.g.,the interferon gamma (IFN-γ), were revealed which affect the mode and strength of the measles induced immunity [6]. Primary measles vaccine failure is considered is reported for about 2-12% of children immunized at around one year of age [8]. In Russia, routine measles vaccination was started in 1967, and since 1986, the mandatory two-time measles vaccine has been introduced. In many countries of the European and American regions, but not in all countries of the world, double measles vaccination has also been introduced.

A multivariate analysis of the innate immune responses in humans following vaccination with yellow fever vaccine (YF17D) was used to examine the response predictors in [14]. The high-throughput data on gene expression profiling, multiplex analysis of cytokines and chemokines and flow cytometry in conjunction with computational models allowed to identify gene signatures that correlate with and predict the magnitude of virus-specific $CD8^+$ T cell and neutralizing antibody responses to vaccine [14]. In [15] the immune status parameters were studied to predict the post-vaccination antibody responses to influenza vaccine. It was shown that the baseline PBMC subpopulation frequencies before vaccination are sufficient to construct accurate predictive models. To this end, both the correlation analysis and cross-validation based predictive modelling were used. For training and validating predictive models the subjects were randomly subdivided into training and testing sets (75% and 25%, respectively). The predictor parameters were iteratively selected based on the parameter's strength of correlation with the end-point in the training set [15]. The antibody responses to meningococcus vaccine in healthy adults were studied using an integrative network modelling approach in [16]. Differently expressed genes together with the pathways (signaling and transcription) whose expression was correlated to antibody responses were identified. The mathematical modelling and analysis for linking data to models are based on the use of a broad range of mathematical statistics-, information-theoretic, and machine learning algorithms [17–19]. In general, the development of a predictive model for personalized immune response is the major challenge in vaccinology [20]. A linear regression model linking the measures of vaccine-induced immune responses with various factors such as individual's biomarkers was proposed thus marking a first step towards a personalized predictive vaccinology.

Previously, we showed that the pre-vaccination host immune status, i.e., the peripheral blood lymphocyte subpopulation structure, the levels of major immunoglobulin classes and the serum cytokine profile, impact the height and durability of post-vaccination immunity [1]. In the current study, we examined the relationship between the subsets of peripheral blood lymphocytes and cytokines and the induced virus-specific cellular and humoral immune responses. The purpose of this study is to identify novel biomarker predictors of protective immunity following vaccination with measles-mumps-rubella (MMR) vaccine.

The analysis of high-dimensional data in which the number of variables largely exceeds the small sample size is a challenge, cooming to may problems in life sciences [26]. Various strategies for parameter selection can be use to identify the mots informative ones. In this study, we implemented three complementary approaches, i.e. the correlation analysis, the supervised screening, and implicit feature selection using regularized (or penalized) regression. Several multivariate classification algorithms are explored to identify combinations of biomarkers that demonstrate sufficient predictive performance.

In Section 2, the clinical data and the mathematical tools used for their analysis are described. In Section 3, we present the results of building the predictive models for vaccine-induced immune responses. The study is complemented by discussion in Section 4.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 January 2023 doi:10.20944/preprints202301.0027.v1

2. Materials and Methods

2.1. Study design

The study involved 19 children (9 boys and 10 girls) aged 1 to 2 years (average age 1 year 3 months). These children had not previously been vaccinated against measles, rubella and mumps, did not have these diseases. All children were vaccinated with the Priorix vaccine (GlaxoSmithKline). Before vaccination, a week and 6 weeks after it, in total 6 mL of blood was taken from the cubital vein in test tubes Vacutaner with heparin, and gel to obtain serum. The blood samples were drawn immediately before the vaccine administration and at days 7 and 42 after vaccination. At days 0 and 7 we measured 59 immune status parameters, including the frequencies of 42 immune cell subsets, levels of 13 cytokines and antibody levels. At day 42, i.e., 6 weeks postvaccination, we measured the specific immune responses against each strain: specific IgA and IgG, Ag-induced IFN-γ production, CD107a expression on $CD8^+$ T cells, and cellular proliferation levels by CFSE dilution (totally, 13 response variables). The overall details of study design are summarized in Figure 1A.

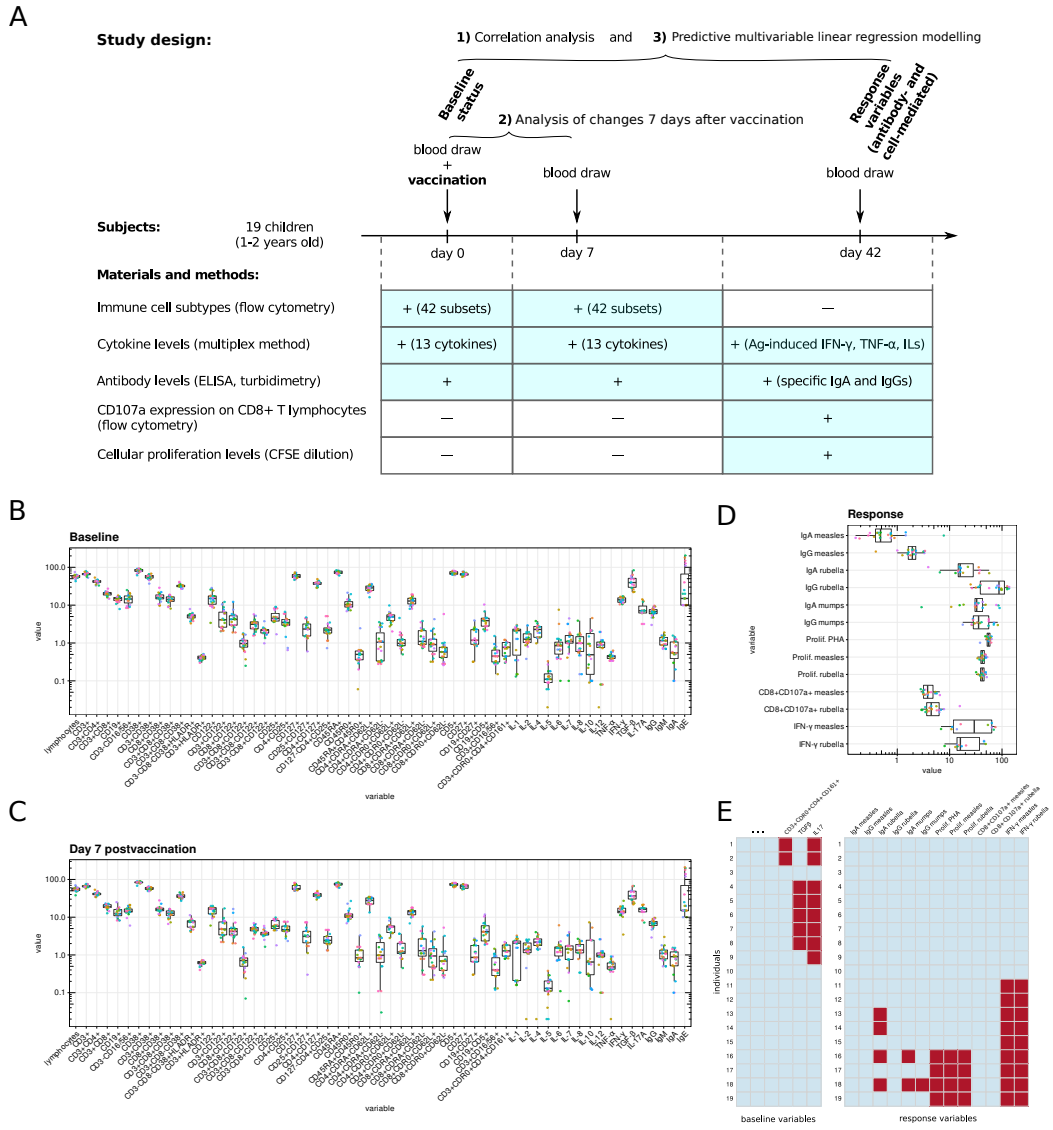

Figure 1. Collected data on baseline immune status and post-vaccination immune response. (**A**) Study design and materials and methods. (**B-D**) Boxplots of measured immune status parameters at day 0 (**B**), day 7 (**C**), and boxplot of post-vaccination response variables at day 42 (**D**). (**E**) Indication of missing data for baseline and response variables (missing observations are colored in red).

 doi:10.20944/preprints202301.0027.v1

2.2. Isolation of lymphocytes from blood

Peripheral blood mononuclear cells (PBMCs) were isolated from heparinized peripheral blood samples by Ficoll-Histopaque density gradient centrifugation. Isolated PBMC were divided into 3 parts (see below).

2.3. Antibodies

Anti-CD3-FITC, anti-CD3-PerCP, anti-CD16/56-PE, anti-CD4-PerCP, anti-CD8-FITC, anti-CD8-PerCP, anti-CD38-PE, anti-DR-APC, anti-CD122-PE, anti-CD45RA-FITC, anti-CD45R0-PE, anti-CD62L-APC, anti-CD161-APC, anti-CD25-FITC, anti-CD127-PE, anti-CD27-FITC, anti-CD5-PerCP, anti-CD19-APC, CD107a-RE-Su5 (BD).

2.4. Immunofluorescent staining

The first part of PBMCs was incubated with anti-surface antigen antibodies at 4°C for 20 min, washed in phosphate buffered saline (PBS) and fixed with 1% formaldehyde in PBS. The specific cellular immune response to measles and rubella antigens was evaluated in 3 ways: Ag-specific cells detection, Ag-specific cell proliferation, and Ag-specific cytokines production. As antigens, a tween-ether extract of measles or rubella virus cultures was used.

2.5. Ag-specific cells detection

The second part of PBMCs was resuspended in 1 mL of RPMI-1640 medium supplemented with 10% calf fetal serum, 2 mM L-glutamine and gentamicin. Cell viability (according to trypan blue staining) was not lower than 95%. In three sterile conical tubes (volume 1.5 mL) were added: 50μL of monensin solution (to a final concentration of 10μM), a suspension of PBMCs (10^6 cells per tube), monoclonal antibodies to CD107a-PE-Cy5 (final dilution 1:100) and 10μL of measles or rubella virus antigens; 10μL of medium was added to the control tube instead of antigens. The contents of each tube were mixed once with a pipette, and then tubes were centrifuged to precipitate cells (200 g, 1 min) and incubated for 15 h at 37°C in an atmosphere of 5% CO_2 and 100% humidity. After incubation, the tubes were again centrifuged (500 g, 1 min), The supernatant was carefully taken, and the cells were resuspended in PBS, then the cells were stained with anti-CD8-FITC antibodies, washed in PBS and fixed with 1% formaldehyde in PBS.

2.6. Cell proliferation

A third part of PBMCs was resuspended in 1 mL of RPMI-1640 medium with gentamicin at a concentration of 10^7 cells per mL. Cells were placed in a sterile round-bottom tube (2 mL volume) and 5- and 6-carboxyfluorescein diacetate-succinylmidyl ether (CFSE) was added at a concentration of 1 µg/ml. Lymphocytes were incubated at 37°C for 30 minutes. Then the cells were washed 4 times with cold RPMI-1640, diluted with RPMI-1640 medium supplemented with 10% calf fetal serum, 2 mM L-glutamine and gentamicin to a concentration of 2×10^6 cells and placed in the wells of a 4-well panel in a volume of 1 mL per well (Nunc). The first well was a negative control, in the second phytohemagglutinin (PHA) 5 µg/ml was added as a positive control, in the third—measles antigens, in the fourth—rubella antigens. Panels were incubated for 8 days at 37°C in an atmosphere of 5% CO_2 and 100% humidity. At the end of the incubation, 200 µl of supernatants were taken from each well for analysis of the cytokine profile. Then the cells were washed in PBS and fixed with 1% formaldehyde in PBS.

2.7. Flow cytometry

All samples prepared for FACS analysis were analyzed within 24 h of staining on an FACSCalibur cytometer using CellQuest software (BD).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 January 2023 doi:10.20944/preprints202301.0027.v1

2.8. Detection of cytokines

The level of 13 cytokines: IL-1 (IL-1β), IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12, IL-17A, IFN-γ, TGF-β and TNF-α were evaluated both in serum and in culture supernatants using a two-laser automated analyzer (Bio-plex Protein Assay System, Bio-Rad) with commercial test systems (determined dynamic range 0.2-3200 pg/mL) in accordance with the manufacturer's instructions. All reactions were performed in a 96-well plate format. The amount of cytokines in the test samples was determined using standard calibration dilutions, cytokine concentrations were calculated automatically using the Bio-Plex Manager software.

2.9. Specific immunoglobulin ELISA

Specific IgG antibodies to measles, rubella and mumps were determined in the serum samples by enzyme-linked immunosorbent assay (ELISA) (Euroimmun). Values of less than 0.01 IU/mL for measles antigens and less than 15 IU/mL for rubella and mumps antigens were considered negative. The amount of IgG antibodies considered to be protective was greater than 0.25 IU/mL for measles viruses and 25 IU/mL for rubella and mumps viruses. To identify specific IgA antibodies to the studied viruses in the same test systems, the anti-IgG conjugate was replaced with an anti-IgA conjugate (clone 14A/1H9). The level of total antibodies in blood serum classes G, M, and A was determined by turbidimetry.

2.10. Data source and ethics statement

The work was approved by the Ethics Committee of G.N. Gabrichevsky Research Institute for Epidemiology and Microbiology, parents signed an informed consent for the participation of children in the research program.

2.11. Data analysis

2.11.1. Exploratory analysis and data preprocessing

The baseline characteristics of the immune status of all subjects at day 0 (baseline) and day 7 post-vaccination are presented in Figure 1B,C, respectively. The response variables are summarized in Figure 1D. All variables were tested for normality, the resulting D'Agostino's K-squared omnibus tests didn't show consistent normality among either variables or log-transformed variables (data not shown). Before training predictive models, we log-transformed the variables and then standardized design matrices and centered the responses (models with log-transformed variables showed better performance). The missing data are indicated in Figure 1E. The variables of baseline immune status $CD3^+CD45R0^+CD4^+CD161^+$, TGF-β and IL-17A were excluded from the sets of potential predictors within feature selection strategies (see below) due to large number of missing values.

2.11.2. Feature selection strategies for building predictive models

We utilized three feature selection strategies to select potential predictors of the post-vaccination responses among baseline variables: (a) screening of the variables based on the correlation analysis, (b) supervised screening based on the information of changes of baseline variables at day 7, (c) implicit feature selection is performed using regularization-based sparse regression (Figure 2). The following filtering criteria were used for screening using strategies (a) and (b). In strategy (a), the variables were selected with Spearman correlation p-value < 0.2; among them, up to 5 variables that correlate with each other as low as possible were selected. In strategy (b), the variables with significant and coherent changes on day 7 were selected in substrategy (b1) as determined by the paried t-test p-value < 0.05 and the Cohen's d-value effect size > 1; the variables with unsignificant changes on day 7 (which can be either due to a coherent lack of change among variables or due to changes in different directions

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 January 2023 doi:10.20944/preprints202301.0027.v1

among variables) were selected in substrategy (b2) as determined by the paired t-test p-value > 0.5 and the Cohen's d-value effect size < 0.05.

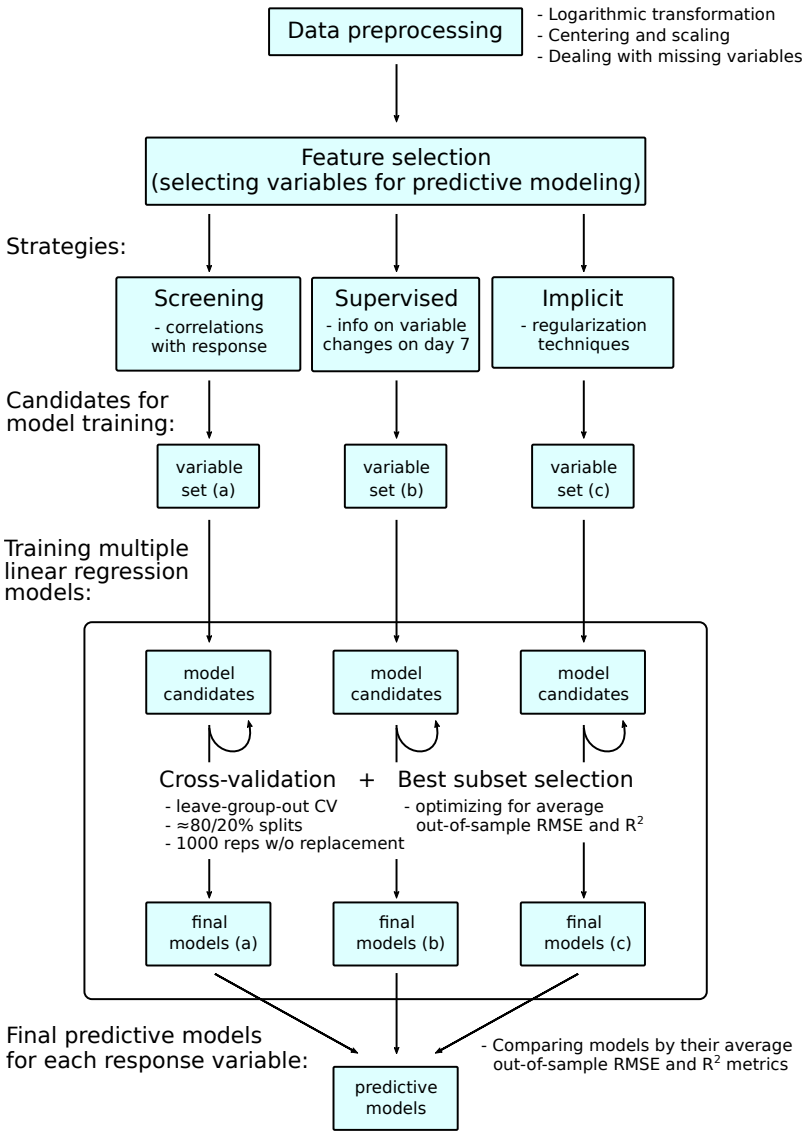

Figure 2. The workflow for building predictive models of post-vaccination immune response, which includes screening for potential predictors with three different screening strategies and comparing the candidate models using cross-validation.

2.11.3. Correlation analysis

For the analysis of pairwise baseline correlations and correlations between baseline variables and response variables, the Spearman's rank correlation coefficients were used as nonparametric measures of monotonic relationships. The p-values were obtained using Fisher transformation and Student's t-distribution.

2.11.4. Analysis of changes from baseline to day 7 postvaccination data

For analysis of the effect size and statistical significance of variable changes from day 0 to day 7, we applied both the paired t-test and Wilcoxon signed-rank test. The p-values were adjusted using Benjamini-Hochberg multiple testing correction procedure to control for the false discovery rate. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 January 2023 doi:10.20944/preprints202301.0027.v1

absolute value of Cohen's d-values are reported as a measure of effect size for t-tests, the r-values ($|r| = |z|/\sqrt{n}$, where z is z-score) are reported for Wilcoxon tests.

2.11.5. Sparse regularization-based regression models

Given the number of potential predictors $p = 56$ is much larger than the number of observations with non-missing data $n \leq 19$, we applied various regularization procedures for the linear regression models to deal with the respective ill-posed problems. The value of regularization parameter determines the variable selection, i.e., the number of baseline parameters of immune status included in the model. There are various ways to regularize the linear regression, differing by the penalty function that is used to prevent the values of model coefficients from becoming too large (and thus screening out variables with zero coefficients). We used the minimax concave penalty (MCP) method implemented in R package ncvreg [21], which has two tuning parameters γ and λ, and converges to popular lasso regression [22] with convex l_1-norm penalty as $\gamma \to \infty$. Contrary to the lasso regression, MCP-penalized regression is less biased because MCP applies less shrinkage to the large nonzero coefficients. Moreover, MCP-penalized regression is one of the sparse regression methods which have the oracle property, i.e., asymptotically, they recover the true variables and screen out the noise variables. The tuning parameters can be chosen using the cross-validation (CV) methods as those that provide the minimal out-of-sample prediction error estimate. The same out-of-sample error can be estimated using the information criteria (IC). We used AICc, the corrected for small sample sizes Akaike IC, which is a second-order estimate of the out-of-sample error. To determine the regularization parameters, we utilized the hybrid strategy combining the convexity diagnostics and AICc as described in [21]. For a given value of γ, we chose the value of λ using the AICc criterion and compared this value with the critical value λ^*, below which the objective function becomes locally nonconvex. If $\lambda < \lambda^*$, we increased γ to make the penalty more convex, and for $\lambda \gg \lambda^*$, we decreased γ to reduce the bias without fear of reaching the unstable region. This procedure was iterated for each model until the γ value reflected the balance between parsimony and convexity. The value of λ was chosen based on AICc. We also looked for the diagnostics for choosing the final value of λ for each model which are based on marginal false discovery rates (mFDR) of selected variables, calculated using the permutation of residuals [23]. The final value of λ was chosen to leave larger set of candidate variables to be used later for best-subset selection (see below).

2.11.6. Choosing final predictive models of post-vaccination responses

For each set of candidate variables obtained using three feature selection strategies, multiple candidate linear regression models were built using all possible subsets of variables as independent variables of the model. The comparison of candidate models were performed using repeated leave-group-out cross-validation by splitting the datasets into training and validation datasets randomly in approximately 80/20% proportions 1000 times without replacement. The best models were selected for each candidate set by choosing the model with minimal average out-of-sample (computed on validation datasets) residuals, i.e, with minimal RMSE. The average out-of-sample coefficient of determination R^2 of models are reported. The final predictive models for each post-vaccination response variable are chosen by comparing the final models obtained for each set of model candidates (Figure 2). The overall workflow for building predictive models of post-vaccination immune response is summarized in Figure 2.

3. Results

3.1. Correlation-based selection of predictor variables

Screening of the predictor variables using the correlation analysis between the vaccine-induced response and the baseline immune status parameters resulted in the following selection

For measles:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 January 2023 doi:10.20944/preprints202301.0027.v1

- **IgA:** $CD127^{-}CD4^{+}CD25^{+}$ lymphocytes, $CD45RA^{+}CD45R0^{+}$ lymphocytes, $CD5^{+}$ lymphocytes, TNF-α, IL-17A.
- **IgG:** $CD3^{+}CD8^{+}CD38^{+}$ lymphocytes, $CD5^{+}$ lymphocytes, IL-7, TGF-β, $CD8^{+}CD122^{+}$ lymphocytes, $CD3^{+}CD16/56^{+}$ lymphocytes, IL-1, IL-2, IL-12.
- **Proliferation:** $CD3^{+}CD16/56^{+}$ lymphocytes, IL-7, IL-8, $CD8^{+}CD122^{+}$ lymphocytes, TGF-β.
- **$CD8^{+}CD107a^{+}$:** $CD38^{+}$ lymphocytes, $CD3^{+}CD8^{+}CD122^{+}$ lymphocytes, IL-6, IL-8, IL-1, TNF-α.
- **IFN-γ:** $CD3^{+}CD8^{-}CD122^{+}$ lymphocytes, IL-2, TGF-β, $CD127^{+}$ lymphocytes, $CD19^{+}CD27^{+}$ lymphocytes.

For rubella:

- **IgA:** $CD3^{+}CD38^{+}$ lymphocytes, $CD25^{+}CD127^{+}$ lymphocytes, $CD8^{+}CD45R0^{+}CD62L^{-}$ lymphocytes, $CD127^{-}CD4^{+}CD25^{+}$ lymphocytes, $CD27^{+}$ lymphocytes.
- **IgG:** $CD45RA^{+}$ lymphocytes, IL-10, IFN-γ, $CD3^{-}CD8^{-}CD38^{+}HLADR^{+}$ lymphocytes, $CD3^{+}CD8^{-}CD122^{+}$ lymphocytes, $CD4^{+}CD45R0^{+}CD62L^{-}$ lymphocytes, $CD19^{+}CD27^{+}$ lymphocytes.
- **Proliferation:** $CD8^{+}CD122^{+}$ lymphocytes, $CD3^{+}CD16/56^{+}$ lymphocytes, IL-7, IL-8, TGF-β.
- **$CD8^{+}CD107a^{+}$:** IL-8, IL-12, $CD3^{+}CD8^{-}CD122^{+}$ lymphocytes, IL-6, $CD27^{+}$ lymphocytes.
- **IFN-γ:** $CD3^{+}38^{+}$ lymphocytes, $CD3^{+}CD8^{-}CD122^{+}$ lymphocytes, $CD4^{+}CD45R0^{+}CD62L^{+}$ lymphocytes, $CD4^{+}CD45R0^{+}CD62L^{-}$ lymphocytes, $CD27^{+}$ lymphocytes.

For mumps:

- **IgA:** $CD8^{+}CD38^{+}$ lymphocytes, $CD3^{+}CD122^{+}$ lymphocytes, $CD127^{+}$ lymphocytes, $CD8^{+}CD45RA^{+}CD62L^{-}$ lymphocytes, $CD19^{+}CD27^{+}$ lymphocytes, IL-1.
- **IgG:** $CD3^{+}CD8^{-}CD122^{+}$ lymphocytes, $CD19^{+}CD27^{+}$ lymphocytes, $CD4^{+}CD45R0^{+}CD62L^{-}$ lymphocytes, IL-2, IL-17A, $CD25^{+}CD127^{+}$ lymphocytes, $CD4^{+}CD45R0^{+}CD62L^{+}$ lymphocytes.

Additionally, the formally selected predictor variables were critically reviewed/approved taking into account their immunological function.

3.2. Mode of change-based selection of predictor variables

Our second approach to selection of potential predictors of the response to vaccination variables is based on the analysis of changes from day 0 to day 7 of the baseline immune status parameters, i.e. the effect size and statistical significance. In the first strategy (b1), the variables with significant and coherent changes on day 7 were selected. These are:

- $CD3^{+}HLADR^{+}$ lymphocytes,
- $CD3^{+}CD8^{+}CD38^{+}$ lymphocytes,
- $CD3^{+}CD8^{-}CD122^{+}$ lymphocytes,
- $CD3^{-}CD8^{+}CD122^{+}$ lymphocytes,
- $CD3^{-}CD8^{-}CD38^{+}HLADR^{+}$ lymphocytes,
- $CD3^{+}CD8^{+}CD122^{+}$ lymphocytes,
- $CD3^{+}CD45R0^{+}CD4^{+}CD161^{+}$ lymphocytes,
- $CD4^{+}CD25^{+}$ lymphocytes,
- IL-17A.

Following a second strategy, the predictor variables with nonsignificant changes on day 7 were selected as follows:

- $CD8^{+}CD45R0^{+}CD62L^{+}$ lymphocytes,
- $CD3^{-}CD16/56^{+}$ lymphocytes,
- IgM,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 January 2023 doi:10.20944/preprints202301.0027.v1

- $CD3^+$ lymphocytes,
- $CD122^+$ lymphocytes,
- IL-4
- $CD3^+CD16/56^+$ lymphocytes,
- $CD45RA^+$ lymphocytes.

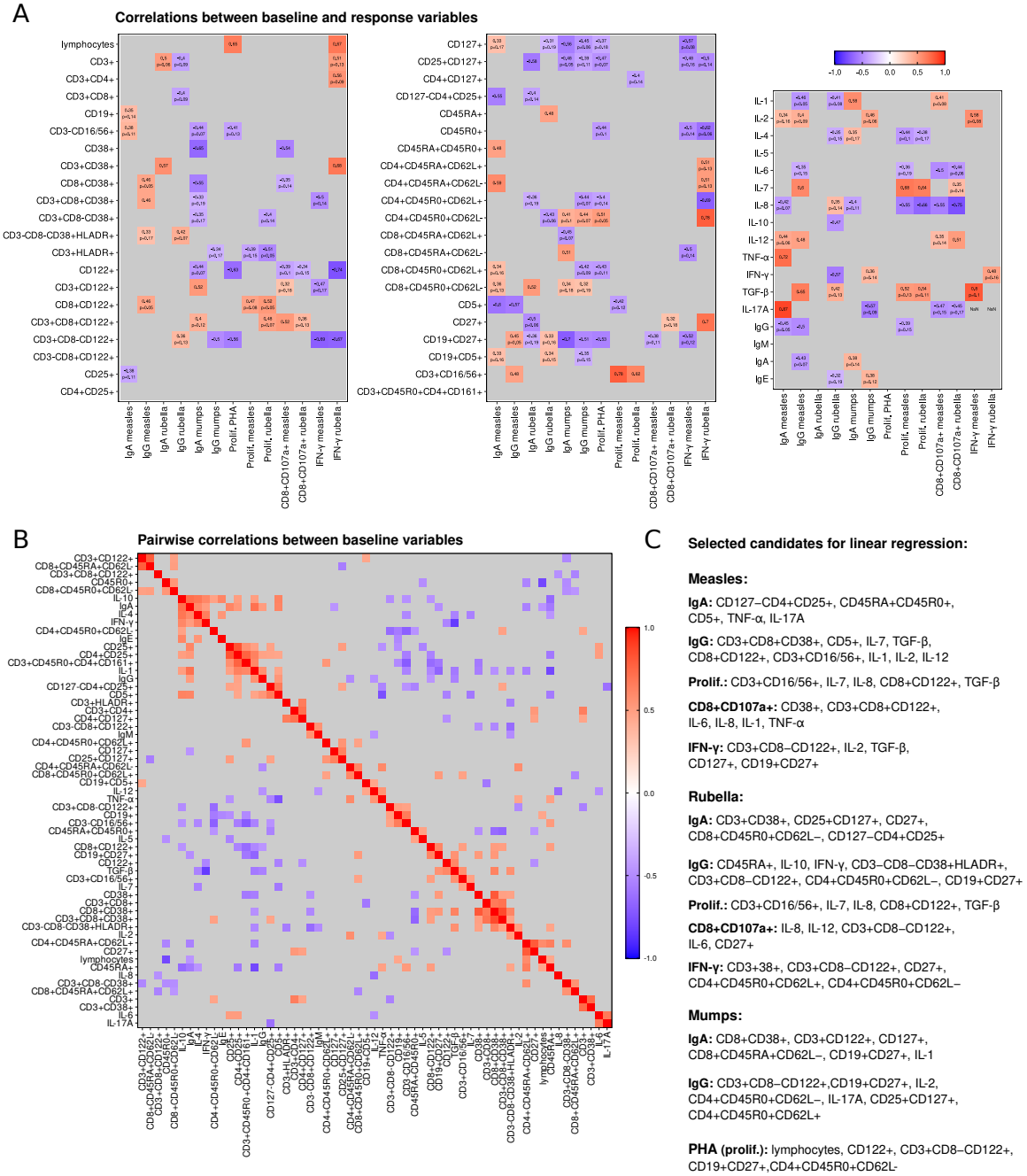

Figure 3. Correlation analysis. (**A**) Correlations between baseline immune status parameters and vaccine-specific immune response variables. Three levels of statistical significance are indicated: significant correlation coefficients ($p < 0.05$) are shown, p-values are additionally written for correlations with $0.20 < p < 0.05$, and correlations with $p > 0.20$ are marked as gray cells. (**B**) Matrix of pairwise correlations for baseline immune status parameters sorted with hierarchical clustering ordering method. Correlations with significance level $p \geq 0.5$ are marked as gray cells. (**C**) Baseline variables selected by correlation analysis as potential predictors to construct candidate predictive models of post-vaccination immune response.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 January 2023 doi:10.20944/preprints202301.0027.v1

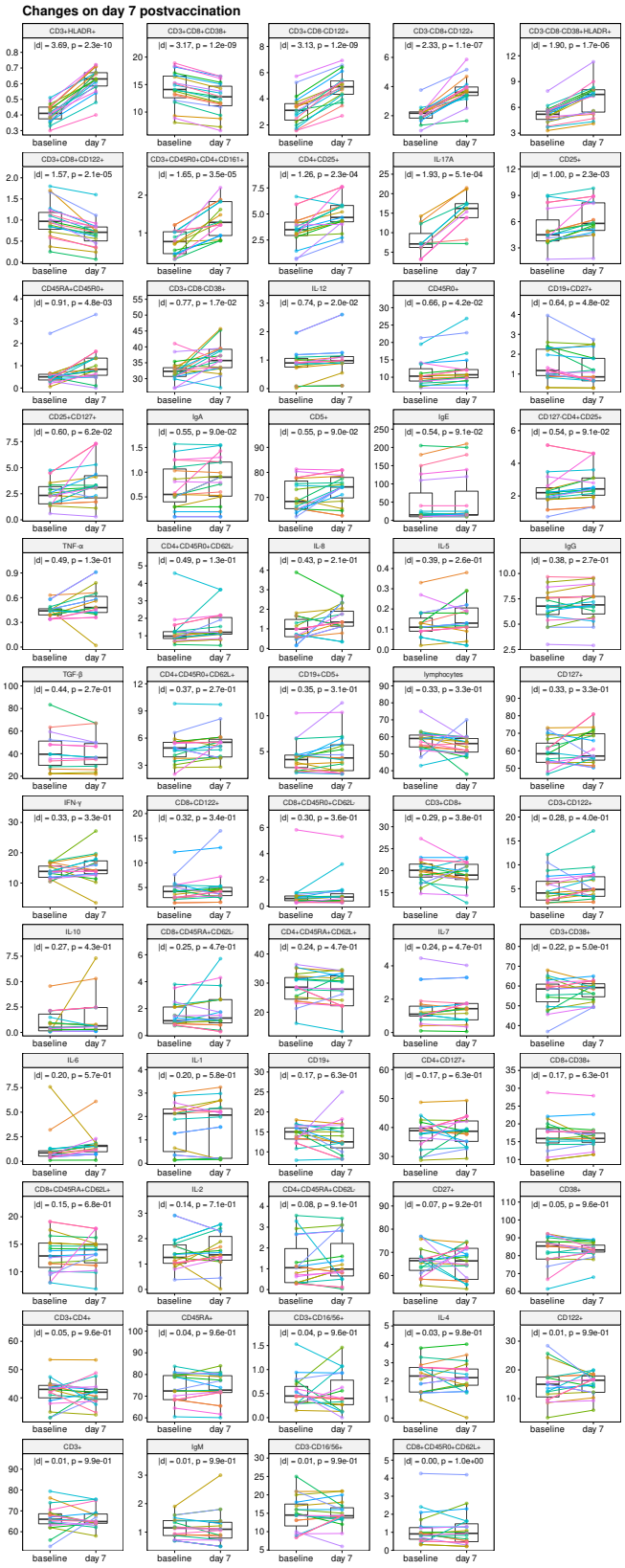

Figure 4. Analysis of the variable changes from baseline (day 0) to day 7 post-vaccination using the paired t-test. The effect size measured as the absolute value of the Cohen's d-value and p-value adjusted to control over false discovery rate are indicated. The variables are sorted by the effect size. The first 9 variables were used as potential predictor candidates in strategy (b1), the last 8 variables—in strategy (b2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 January 2023 doi:10.20944/preprints202301.0027.v1

3.3. Regularization-based selection of predictor variables

The third approach to the selection of the baseline predictor variables for the vaccine-induced immune responses was based on the application regularization procedure linked to the linear multivariate regression building and is implicit in its nature. The finally identified set of the immune status parameter which are the best predictors of the post-vaccination immune response variables are presented in Table 1.

Table 1. Summary of the final multilinear regression models predicting post-vaccination immune response variables: baseline predictor variables, cross-validation based consistency metrics (mean out-of-sample error (RMSE), coefficient of determination (R^2_{out})), model performance fitted on the complete dataset (adjusted R^2_{adj}, F-test p-value).

Response variable Y	**Baseline predictor variables** x_i	**Str.**[*]	**RMSE**	$\mathbf{R^2_{out}}$	$\mathbf{R^2_{adj}}$	**p-value**
Measles:						
IgA	TNF-α, $CD45RA^+CD45R0^+$	(a)	0.68	0.78	0.33	0.015
IgG	IL-12, IL-1, IL-2	(a)	0.6	0.75	0.36	0.02
Cell prolif.	$CD3^+CD16/56^+$, $CD3^-CD16/56^+$, $CD3^+$, $CD45RA^+$	(b2)	0.07	0.81	0.73	0.001
$CD8^+CD107a^+$	$CD38^+$, $CD3^+CD8^+CD122^+$	(a)	0.21	0.37	0.24	0.04
IFN-γ	$CD3^+CD8^-CD122^+$, IL-2	(a)	14.2	1	0.52	0.03
Mumps:						
IgA	$CD4^+CD45RA^+CD62L^-$, IL-1, $CD27^+$	(c)	0.44	0.74	0.77	$6 \cdot 10^{-5}$
IgG	$CD19^+CD27^+$, IL-2	(a)	0.34	0.52	0.53	0.001
Rubella:						
IgA	$CD27^+$, $CD25^+CD127^+$	(a)	10.41	0.81	0.56	0.003
IgG	$CD45RA^+$, IFN-γ, $CD127^+$, IL-10, $CD3^+CD8^-CD38^+$	(c)	0.3	0.89	0.7	$6 \cdot 10^{-4}$
Cell prolif.	IL-7, $CD3^+CD16/56^+$	(a)	0.07	0.89	0.75	$9 \cdot 10^{-5}$
$CD8^+CD107a^+$	IL-8, $CD27^+$	(a)	0.32	0.85	0.24	0.04
IFN-γ	$CD27^+$	(a)	9.07	1	0.67	0.002
PHA (unspecific):						
Cell prolif.	$CD3^-CD8^-CD38^+HLADR^+$, $CD3^+HLADR^+$, $CD3^+CD8^+CD122^+$, $CD3^+CD45R0^+CD4^+CD161^+$	(b1)	0.06	0.61	0.52	0.04

[*] Screening strategy that produced the best model: (a) correlation-based, (b) selecting variables that change on day 7 coherently (b1) or uncoherently (b2), (c) implicit regularization-based method.

3.4. Predictive models for post-vaccination antibody and cellular responses

The general form of the multiple linear regression models reads

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i \times x_i, \tag{1}$$

where Y stands for the response variable and x_i are the predictor (or explanatory) variables. The identified predictors are presented in Table 1. The constants β_i are called partial regression coefficients and β_0 is know as intercept.

During data pre-processing, the baseline variables x_i are standardized, i.e., centered around their mean values μ_i and scaled by their standard deviations σ_i. We denote the standardization transformation as function z: $z(x, \mu, \sigma) = (x - \mu)/\sigma$. Before standardization, variables can be optionally logarithmically transformed (which is done if this transformation improves the model performance metrics). We denote the initial transformation as function $g(x)$, which can be an identical

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 January 2023 doi:10.20944/preprints202301.0027.v1

function if logarithmic transformation is not applied ($g_{id}(x) = x$, $g_{id}^{-1}(x) = x$) or a function of natural logarithm ($g_{log}(x) = \log(x)$, $g_{log}^{-1}(x) = \exp(x)$). Thus, the final form of the predictive models reads

$$Y = g^{-1}\left(\beta_0 + \sum_{i=1}^{p} \beta_i \times z(g(x_i), \mu_i, \sigma_i)\right) \quad (2)$$

The estimated values of the regression coefficients β_i, intercepts β_0 and scaling coefficients μ_i, σ_i, as well as the form of function g are summarized in Table 2.

Table 2. Summary of the parameters of the final multilinear regression models predicting post-vaccination immune response variables: regression coefficients β_i, intercepts β_0 and scaling coefficients μ_i, σ_i, the form of transformation function g.

Y	g	β_0	x_i	β_i	μ_i	σ_i
Measles:						
IgA	g_{log}	-0.52	TNF-α	0.45	-0.8365	0.1859
			$CD45RA^+CD45R0^+$	0.27	-0.8186	0.6961
IgG	g_{id}	2.03	IL-12	0.33	0.8679	0.5424
			IL-1	-0.32	1.5974	1.0041
			IL-2	0.12	1.4395	0.6621
Cell prolif.	g_{log}	3.75	$CD3^+CD16/56^+$	0.10	-0.7672	0.5367
			$CD3^-CD16/56^+$	-0.08	2.6545	0.3263
			$CD3^+$	-0.05	4.1927	0.0950
			$CD45RA^+$	0.04	4.2964	0.0895
$CD8^+CD107a^+$	g_{log}	1.36	$CD38^+$	-0.11	4.4108	0.1089
			$CD3^+CD8^+CD122^+$	0.06	-0.1064	0.4953
IFN-γ	g_{id}	37	$CD3^+CD8^-CD122^+$	-15	3.1568	1.1324
			IL-2	11	1.4395	0.6621
Mumps:						
IgA	g_{log}	3.45	IL-1	0.76	0.0789	1.1037
			$CD4^+CD45RA^+CD62L^-$	0.35	-0.1043	0.9017
			$CD27^+$	0.06	4.1852	0.0923
IgG	g_{log}	3.71	$CD19^+CD27^+$	-0.26	0.2228	0.6622
			IL-2	0.21	0.2634	0.4756
Rubella:						
IgA	g_{id}	20.4	$CD27^+$	-8.3	65.975	6.1217
			$CD25^+CD127^+$	-6.9	2.5274	1.1903
IgG	g_{log}	4.24	$CD45RA^+$	0.31	4.2964	0.0895
			IFN-γ	-0.24	2.6070	0.1597
			$CD127^+$	-0.19	4.0726	0.1334
			$CD3^+CD8^-CD38^+$	-0.16	3.4746	0.1004
			IL-10	-0.07	-0.6446	1.2909
Cell prolif.	g_{log}	3.72	IL-7	0.08	0.1041	0.8241
			$CD3^+CD16/56^+$	0.05	-0.7672	0.5367
$CD8^+CD107a^+$	g_{log}	1.59	IL-8	-0.22	-0.1909	0.8052
			$CD27^+$	0.06	4.1852	0.0923
IFN-γ	g_{id}	21	$CD27^+$	13	65.975	6.1217
PHA (unspecific):						
Cell prolif.	g_{log}	4.04	$CD3^-CD8^-CD38^+HLADR^+$	-0.10	1.6134	0.1998
			$CD3^+HLADR^+$	-0.05	-0.8991	0.1372
			$CD3^+CD8^+CD122^+$	0.04	-0.1064	0.4953
			$CD3^+CD45R0^+CD4^+CD161^+$	-0.01	-0.3762	0.4725

 doi:10.20944/preprints202301.0027.v1

We have identified the set of multiple linear regression models for the vaccine-induced immune responser to measles, rubella, mumps. The corresponding predictor variables and coefficients are presented in Table 2. The number of predictors ranges form one to five.

We have identified the set of multiple linear regression models for the vaccine-induced immune responser to measles, rubella, mumps. The corresponding predictor variables and coefficients are presented in Table 2. The number of predictors ranges form one to five. They appear to differen essentially between the response variables of interest, i.e. the antibody and cellular vaccine-induced immunity.

4. Discussion

Prediction of the vaccine-induced immune using immune signatures measured at baseline prior to vaccination is a problem of fundamental and medical importance [26]. We presented a systematic analysis of the original data coming from study involving 19 children aged 1 to 2 years vaccinated against measles, rubella and mumps with the Priorix vaccine (GlaxoSmithKline). We have identified optimal multivariate liner regression models for predicting the effectiveness of vaccination against measles-mumps-rubella using the baseline immune status parameters. It turned out that the sufficient number of predictor variables ranges from one to five depending on the response variable of interest. The results of our study provide a quantitative computational tool for predicting the effectiveness of vaccines on a personalized basis.

Dimensionality reduction is a procedure required for the analysis of high dimensional data collected for a small number of samples. It is recognized that an exhaustive exploration of high-dimensional data space is practically impossible [28]. One needs to explore various techniques to reduce the dimensionality. However, the building of a low dimensional models which approximate high-dimensional data is based upon assumption that a manifold with small intrinsic dimension exists [27]. Our study suggest that this is likely to be the case in building low-dimensional multivariate liner regression models for predicting the vaccine-induced immunity from a plethora of baseline immune status parameters.

Understanding the cellular and molecular mechanisms that control the ability of immune system to mount a protective immune response against various pathogens is a central problem in immunology. Modern research in immunology is characterized by an unprecedented level of detail that has progressed towards viewing the immune system as numerous components that function together as a whole network [10]. There exist significant difficulties in analyzing the data being generated from high-throughput technologies for understanding immune system dynamics and functions. This calls for the application of mathematical modelling to complement the clinical studies with the aim to describe, analyze and predict the observable characteristics of infections [9]. Mathematical models describing the dynamics of virus infections is a rapidly developing theoretical area of mathematical immunology [11]. Mathematical models considering the spread of infections are broadly used in the evaluation of vaccination strategies at the population level [12,13]. However, practically relevant mathematical studies concerning the personalized predictive modelling of the immune responses to vaccines are still rather rare. The results of our study provide a basis for the development of the mechanistic models which could help to derive a better understanding of the mechanisms by which vaccines protect and with a targeted modulation of immune baseline parameters in order to improve vaccine outcomes [24].

Overall, systems approaches combining a high-dimensional characterization of the individual's responses to antigenic perturbation with the information extraction power of computational modeling should pave the way to a rapid and transformative advances in modern systems vaccinology [25].

Author Contributions: Conceptualization, A.T., D.G. and G.B.; methodology, A.T., D.G. and G.B.; software, D.G.; validation, A.T., D.G. and G.B.; investigation, A.T., D.G. and G.B.; resources, A.T.; data curation, A.T.; writing—original draft preparation, A.T., D.G. and G.B.; writing—review and editing, A.T., D.G. and G.B.; visualization, D.G.; supervision, A.T. and G.B.; funding acquisition, G.B. All authors have read and agreed to the published version of the manuscript.

Funding: The reported study was funded by the Russian Foundation for Basic Research, project number 20-01-00352. D.G and G.B. were partly supported by Moscow Center of Fundamental and Applied Mathematics (agreement with the Ministry of Education and Science of the Russian Federation No. 075-15-2022-286).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of G.N. Gabrichevsky Research Institute for Epidemiology and Microbiology (protocol №41, 14.02.2018), parents signed an informed consent for the participation of children in the research program.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MMR	measles-mumps-rubella
CFSE	5- and 6-carboxyfluorescein diacetate-succinylmidyl ether
PBS	phosphate buffered saline
PBMCs	Peripheral blood mononuclear cells
ELISA	enzyme-linked immunosorbent assay
MCP	minimax concave penalty
AIC	Akaike Information Criterion
CV	cross-validation
mFDR	marginal false discovery rates
RMSE	Root mean square error

References

1. Toptygina A.P., Aziattceva V.V., Savkin I.A., Kislitsyn A.A., Semikina E.L., Grebennikov D.S., Alioshkin V.A., Sulimov A.V., Sulimov V.B., Bocharov G.A. The prediction of specific humoral immune responses using the baseline immune status parametres in children, vaccinated with measles-mumps-rubella vaccine. *Immunologiya* **2015**, *36(1)*, 22–30.
2. World Health Organization Regional Office for Europe (WHO/Europe). Fifth Meeting of the European Regional Verification Commission for Measles and Rubella Elimination (RVC) 24-26 October 2016, Copenhagen, Denmark. Copenhagen: WHO/Europe. Available online: http://www.euro.who.int/__data/assets/pdf_file/0005/330917/5th-RVC-meeting-report.pdf (accessed on 22 Dec 2022).
3. European Centre for Disease Prevention and Control (ECDC). Epidemiological update: Measles - monitoring European outbreaks, 7 July 2017. Stockholm: ECDC. Available online: https://ecdc.europa.eu/en/news-events/epidemiological-update-measles-monitoring-european-outbreaks-7-july-2017 (accessed on 22 Dec 2022).
4. Filia, A.; Bella, A.; Manso, M.D.; Baggieri, M.; Magurano, F.; Rota, M.C. Ongoing Outbreak with Well over 4,000 Measles Cases in Italy from January to End August 2017 – What Is Making Elimination so Difficult? *Eurosurveillance* **2017**, *22*, 30614, doi:10.2807/1560-7917.ES.2017.22.37.30614.
5. Ovsyannikova, I.G.; Ryan, J.E.; Vierkant, R.A.; Pankratz, V.S.; Jacobson, R.M.; Poland, G.A. Immunologic Significance of HLA Class I Genes in Measles Virus-Specific IFN-γ and IL-4 Cytokine Immune Responses. *Immunogenetics* **2005**, *57*, 828–836, doi:10.1007/s00251-005-0061-6.
6. Haralambieva, I.H.; Ovsyannikova, I.G.; Kennedy, R.B.; Vierkant, R.A.; Pankratz, V.S.; Jacobson, R.M.; Poland, G.A. Associations between Single Nucleotide Polymorphisms and Haplotypes in Cytokine and Cytokine Receptor Genes and Immunity to Measles Vaccination. *Vaccine* **2011**, *29*, 7883–7895, doi:10.1016/j.vaccine.2011.08.083.
7. Bursac, Z.; Gauss, C.H.; Williams, D.K.; Hosmer, D.W. Purposeful Selection of Variables in Logistic Regression. *Source Code for Biology and Medicine* **2008**, *3*, 17, doi:10.1186/1751-0473-3-17.

8. Haralambieva, I.H.; Kennedy, R.B.; Ovsyannikova, I.G.; Schaid, D.J.; Poland, G.A. Current Perspectives in Assessing Humoral Immunity after Measles Vaccination. *Expert Review of Vaccines* **2019**, *18*, 75–87, doi:10.1080/14760584.2019.1559063.
9. Bocharov, G.; Volpert, V.; Ludewig, B.; Meyerhans, A. Editorial: Mathematical Modeling of the Immune System in Homeostasis, Infection and Disease. *Frontiers in Immunology* **2020**, *10*.
10. Bocharov, G.; Volpert, V.; Ludewig, B.; Meyerhans, A. *Mathematical Immunology of Virus Infections*; Springer International Publishing: Cham, 2018; ISBN 9783319723167.
11. Handel, A.; La Gruta, N.L.; Thomas, P.G. Simulation Modelling for Immunologists. *Nat Rev Immunol* **2020**, *20*, 186–195, doi:10.1038/s41577-019-0235-3.
12. Béraud, G. Mathematical Models and Vaccination Strategies. *Vaccine* **2018**, *36*, 5366–5372, doi:10.1016/j.vaccine.2017.10.014.
13. Lanzieri, T.M.; Gastañaduy, P.A.; Gambhir, M.; Plotkin, S.A. Review of Mathematical Models of Vaccination for Preventing Congenital Cytomegalovirus Infection. *The Journal of Infectious Diseases* **2020**, *221*, S86–S93, doi:10.1093/infdis/jiz402.
14. Querec, T.D.; Akondy, R.S.; Lee, E.K.; Cao, W.; Nakaya, H.I.; Teuwen, D.; Pirani, A.; Gernert, K.; Deng, J.; Marzolf, B.; et al. Systems Biology Approach Predicts Immunogenicity of the Yellow Fever Vaccine in Humans. *Nat Immunol* **2009**, *10*, 116–125, doi:10.1038/ni.1688.
15. Tsang, J.S.; Schwartzberg, P.L.; Kotliarov, Y.; Biancotto, A.; Xie, Z.; Germain, R.N.; Wang, E.; Olnes, M.J.; Narayanan, M.; Golding, H.; et al. Global Analyses of Human Immune Variation Reveal Baseline Predictors of Postvaccination Responses. *Cell* **2014**, *157*, 499–513, doi:10.1016/j.cell.2014.03.031.
16. Li, S.; Rouphael, N.; Duraisingham, S.; Romero-Steiner, S.; Presnell, S.; Davis, C.; Schmidt, D.S.; Johnson, S.E.; Milton, A.; Rajam, G.; et al. Molecular Signatures of Antibody Responses Derived from a Systems Biology Study of Five Human Vaccines. *Nat Immunol* **2014**, *15*, 195–204, doi:10.1038/ni.2789.
17. Jaqaman, K.; Danuser, G. Linking Data to Models: Data Regression. *Nat Rev Mol Cell Biol* **2006**, *7*, 813–819, doi:10.1038/nrm2030.
18. Tarca, A.L.; Carey, V.J.; Chen, X.; Romero, R.; Drăghici, S. Machine Learning and Its Applications to Biology. *PLoS Comput Biol* **2007**, *3*, e116, doi:10.1371/journal.pcbi.0030116.
19. Tan, Y.; Tamayo, P.; Nakaya, H.; Pulendran, B.; Mesirov, J.P.; Haining, W.N. Gene Signatures Related to B-Cell Proliferation Predict Influenza Vaccine-Induced Antibody Response: Clinical Immunology. *Eur J Immunol* **2014**, *44*, 285–295, doi:10.1002/eji.201343657.
20. Poland, G.A.; Ovsyannikova, I.G.; Kennedy, R.B. Personalized Vaccinology: A Review. *Vaccine* **2018**, *36*, 5350–5357, doi:10.1016/j.vaccine.2017.07.062.
21. Breheny, P.; Huang, J. Coordinate Descent Algorithms for Nonconvex Penalized Regression, with Applications to Biological Feature Selection. *Ann Appl Stat* **2011**, *5*, doi:10.1214/10-AOAS388.
22. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **1996**, *58*, 267–288, doi:10.1111/j.2517-6161.1996.tb02080.x.
23. Breheny, P.J. Marginal False Discovery Rates for Penalized Regression Models. *Biostatistics* **2019**, *20*, 299–314, doi:10.1093/biostatistics/kxy004.
24. Tsang JS, Dobaño C, VanDamme P, Moncunill G, Marchant A, Othman RB, Sadarangani M, Koff WC, Kollmann TR. Improving Vaccine-Induced Immunity: Can Baseline Predict Outcome? *Trends Immunol.* **2020** *41*(6):457-465. doi: 10.1016/j.it.2020.04.001.
25. Diray-Arce J, Miller HER, Henrich E, Gerritsen B, Mulè MP, Fourati S, Gygi J, Hagan T, Tomalin L, Rychkov D, Kazmin D, Chawla DG, Meng H, Dunn P, Campbell J; Human Immunology Project Consortium (HIPC), Sarwal M, Tsang JS, Levy O, Pulendran B, Sekaly R, Floratos A, Gottardo R, Kleinstein SH, Suárez-Fariñas M. The Immune Signatures data resource, a compendium of systems vaccinology datasets. *Sci Data.* **2022** *9*(1):635. doi: 10.1038/s41597-022-01714-7.
26. Wu Yunan, Wang Lan. A Survey of Tuning Parameter Selection for High-Dimensional Regression. *Annual Review of Statistics and Its Application* **2020** 71:209-226, doi: 10.1146/annurev-statistics-030718-105038

27. Stolz BJ, Tanner J, Harrington HA, Nanda V. Geometric anomaly detection in data. *Proc Natl Acad Sci U S A*. **2020** *117*(33):19664-19669. doi: 10.1073/pnas.2001741117.
28. Nguyen LH, Holmes S. Ten quick tips for effective dimensionality reduction. *PLoS Comput Biol*. **2019** *15*(6):e1006907. doi: 10.1371/journal.pcbi.1006907.