UC San Diego
UC San Diego Previously Published Works

Title

Machine Learning-Based Single Cell and Integrative Analysis Reveals That Baseline mDC Predisposition Correlates With Hepatitis B Vaccine Antibody Response

Permalink

Authors

Aevermann, Brian D
Shannon, Casey P
Novotny, Mark
et al.

Publication Date

2021

DOI

Peer reviewed

Machine Learning-Based Single Cell and Integrative Analysis Reveals That Baseline mDC Predisposition Correlates With Hepatitis B Vaccine Antibody Response

Brian D. Aevermann[1†], Casey P. Shannon[2,3†], Mark Novotny[1†], Rym Ben-Othman[4,5†], Bing Cai[4], Yun Zhang[1], Jamie C. Ye[2,3], Michael S. Kobor[4], Nicole Gladish[4], Amy Huei-Yi Lee[6], Travis M. Blimkie[7], Robert E. Hancock[7], Alba Llibre[8], Darragh Duffy[8], Wayne C. Koff[9], Manish Sadarangani[4,10‡], Scott J. Tebbutt[2,3,11‡], Tobias R. Kollmann[4,5‡] and Richard H. Scheuermann[1,12,13*‡]

[1] Department of Informatics, J. Craig Venter Institute, La Jolla, CA, United States, [2] Prevention of Organ Failure (PROOF) Centre of Excellence, St. Paul's Hospital, Vancouver, BC, Canada, [3] The University of British Columbia (UBC) Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, BC, Canada, [4] Department of Pediatrics, University of British Columbia, Vancouver, BC, Canada, [5] Telethon Kids Institute, Perth Children's Hospital, University of Western Australia, Nedlands, WA, Australia, [6] Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada, [7] Department of Microbiology and Immunology, Life Sciences Institute, University of British Columbia, Vancouver, BC, Canada, [8] Translational Immunology Lab, Institut Pasteur, Paris, France, [9] Human Vaccines Project, New York, NY, United States, [10] Vaccine Evaluation Center, BC Children's Hospital Research Institute, Vancouver, BC, Canada, [11] Department of Medicine, Division of Respiratory Medicine, University of British Columbia, Vancouver, BC, Canada, [12] Department of Pathology, University of California, San Diego, San Diego, CA, United States, [13] Division of Vaccine Discovery, La Jolla Institute for Immunology, La Jolla, CA, United States

Vaccination to prevent infectious disease is one of the most successful public health interventions ever developed. And yet, variability in individual vaccine effectiveness suggests that a better mechanistic understanding of vaccine-induced immune responses could improve vaccine design and efficacy. We have previously shown that protective antibody levels could be elicited in a subset of recipients with only a single dose of the hepatitis B virus (HBV) vaccine and that a wide range of antibody levels were elicited after three doses. The immune mechanisms responsible for this vaccine response variability is unclear. Using single cell RNA sequencing of sorted innate immune cell subsets, we identified two distinct myeloid dendritic cell subsets (NDRG1-expressing mDC2 and CDKN1C-expressing mDC4), the ratio of which at baseline (pre-vaccination) correlated with the immune response to a single dose of HBV vaccine. Our results suggest that the participants in our vaccine study were in one of two different dendritic cell dispositional states at baseline – an NDRG2-mDC2 state in which the vaccine elicited an antibody response after a single immunization or a CDKN1C-mDC4 state in which the vaccine required two or three doses for induction of antibody responses. To explore this correlation further, genes expressed in these mDC subsets were used for feature selection prior to the construction of predictive models using supervised canonical correlation

machine learning. The resulting models showed an improved correlation with serum antibody titers in response to full vaccination. Taken together, these results suggest that the propensity of circulating dendritic cells toward either activation or suppression, their "dispositional endotype" at pre-vaccination baseline, could dictate response to vaccination.

INTRODUCTION

Vaccination as a general strategy to prevent infectious disease has been one of the most effective public health measures since its conceptualization and implementation by Edward Jenner in the 18th century, and has resulted in the complete eradication of smallpox, the near elimination of polio, and a dramatic reduction in the incidences of measles, mumps and other common diseases. In contrast to these successes, several notable failures in the development of effective vaccines against other common infectious diseases, including AIDS, tuberculosis, and malaria suggest that the current empirical approach to vaccine design is not effective in eliciting protective immunity in many cases (1). This variability in vaccine effectiveness highlights the need to better understand the fundamental principles of human immune responses, the "rules of immunity", and how this understanding could be used to develop vaccination strategies that are consistently effective and result in durable immunity.

Recently, several groups have applied high-throughput multi-omics assays to produce a comprehensive systems-level evaluation of vaccine responses, so-called "systems vaccinology" (reviewed in (2–5). One of the key questions addressed is - can baseline (pre-vaccine) signatures of the immune system predict vaccine responses and differentiate between responders vs. non-responders and, if so, what can these signatures tell us about the mechanisms for eliciting protective immunity (6). The general concept that specific baseline immune signatures can predict vaccine responses has been explored in large cohort studies in the context of hepatitis B virus (HBV), influenza and malaria vaccines (7–12). However, the connection between these molecular signatures and the underlying immunological mechanisms remains tenuous. Further, the application of large-scale multi-omics assessment of large vaccination cohorts is cost-prohibitive, raising the question of whether advanced computational and machine learning methods may allow for the discovery of predictive mechanistic signatures in studies with smaller sample sizes (13).

The hepatitis B virus (HBV) vaccine is an ideal platform to explore these questions. First, serum anti-HBV antibody levels, which can be easily measured in participant samples, are a well-established correlate of protection (14). Second, the response to the HBV vaccine is highly variable, providing a broad range of responses, which is useful for identifying correlates and predictors (15). Third, around 10% of subjects respond with protective antibody titers following a single dose (16). We recently applied a series of validated multi-omics assays to

measure the full range of cellular and molecular components of the immune system, including immune cell composition, DNA methylation, gene expression, protein abundance, and fecal 16S microbiome, to provide an exhaustive picture of the immune response to the HBV vaccine (17). Multi-omics integrative analysis on these data sets identified a number of candidate baseline predictors of vaccine response using serum antibody titers to the HBV surface antigen following three vaccine doses as the quantitative endpoint in a relatively small cohort of 15 vaccine recipients (13). While these candidate predictive signatures could be identified using this systems-level approach in a relatively small cohort, a unifying mechanistic driver did not emerge. Furthermore, multi-omic analysis of whole blood failed to reveal features predictive of the variable protective responses following only a single vaccination dose.

In this report, we sought to determine if a more granular approach, namely single cell RNA sequencing in the context of an integrated multi-omic analysis, could identify the relevant cellular phenotypes and functions associated with vaccine responses. Using machine learning to identify the most discriminative gene expression features for dimensionality reduction to optimize performance of canonical correlation analysis (CCA), a truly integrative machine learning approach emerged that helps to overcome small sample sizes through a hypothesis-generation-and-hypothesis-testing-in-orthologous-datasets workflow.

METHODS

Descriptions of all methods not detailed below have been published recently in (13, 17).

Cohort and Sampling Description

A prospective, observational study (ClinicalTrials.gov: NCT03083158) of immune responses to the HBV vaccine (ENGERIX®-B) was undertaken, with recruitment occurring at the Vaccine Evaluation Center (VEC), British Columbia Children's Hospital Research Institute in Vancouver, Canada. Briefly, a total of 15 eligible individuals aged 44 – 73 were enrolled in the study. One ml (20 µg) of ENGERIX®-B vaccine was administered via intramuscular deltoid injection at three different times during the study (Day 0, Day 28 and Day180). HBV titres were measured at screening, Day 28 after the initial vaccine dose (immediately prior to second dose), Day 180 (immediately prior to the third dose), and Day 208 (one

month after the final dose). For the purposes of this study, participants were categorized as "dose 1 responders" if their anti-HBV serum antibody titer was >10 mIU/ml at Day 28 after a single dose of HBV vaccine, (a response considered as protective), "dose 1 marginal responder" if they had detectable anti-HBV serum antibody titer above baseline (3.1 mIU/ml) but <10 mIU/ml at Day 28, or "dose 1 non-responders" if they had no detectable anti-HBV serum antibodies at Day 28. Note that due to funding constraints, only a subset of participants and samples were used for some of the mechanistic assays.

Various omics studies were performed as described (13, 17). Briefly, peripheral whole blood cells were profiled by flow cytometry, genome-wide DNA methylation (Illumina Infinium MethylationEPIC BeadChip), transcript abundance (bulk RNA-Seq), and proteome-wide protein abundance (mass spectrometry) at various time points. Additionally, the bacterial composition (microbiome) of the gut was assessed by 16S rRNA microbiome profiling pre- (Day -14 and Day 0) and post-vaccination (Day 14). Procedures for the collection and processing of PBMC samples for single cell RNA sequencing are described below.

Single Cell RNA Sequencing

Four innate immune cell subsets (monocytes, natural killer (NK) cells, myeloid dendritic cells (mDCs), and plasmacytoid dendritic cells (pDCs)), were single cell sorted for RNA sequencing as described (17, 18) (**Supplementary Figures 1**, **2** and **Supplementary Table 1**). Briefly, 1.5 ml blood samples were stained and single cells sorted for each cell population of interest before performing subsequent single cell RNA sequencing. 20 mM of EDTA (Fisher #BP120-500) was diluted 1:10 in 1.5 ml blood and red blood cells lysed by adding RBC lysis buffer (eBiosciences, cat #00-4333-57) per manufacture's recommendations. After 10 min at room temperature, PBS (Gibco # 14190) was added and the cell suspension separated by centrifugation at 500 × g for 5 min. The cell pellet was resuspended in an antibody mixture diluted in PBS and 0.5% BSA (bovine serum albumin, Sigma Aldrich, cat #A7906) according to the manufacture's recommendations. APC-eFluor 780 (eBioscineces, cat #65-0865-14) viability dye was added to cells prior to staining to sort viable cells. The cell mixture was incubated at room temperature in the dark for 30 min, then washed once in PBS, and resuspend in 3 ml of PBS for immediate single cell sorting into wells of a 96-well microtiter plate chilled on ice using a BD FACS Aria. Innate immune cell subsets were sorted from the APC-eFluor 780- viable cell gate as follows: NK cells (CD45+CD66-CD14-HLA-DR-CD3-CD16+), monocytes (CD45+CD66-CD14+), mDCs (CD45+CD66-CD14-HLA-DR+CD11c+), and pDCs (CD45+CD66-CD14-HLA-DR+CD11c-CD123+) (**Supplementary Figure 2**). Prior to sorting, 96-well plates were pre-loaded with 2µl lysis buffer (0.2% Triton X-100, (Sigma Aldrich, cat #9002-93-1), 2 Units/µl RNase inhibitor (Applied Biosystems, cat #N8080119), 1:2,000,000 dilution of ERCC spike-in RNAs (Life Technologies, cat #4456740) and centrifuged at 300 x g at 4°C for 1 minute to distribute liquid in the bottom of the well. After sorting, each plate containing the single cell lysates was immediately sealed, frozen on dry ice, and stored at −80°C.

Processing of the frozen 96-well plates containing single cell lysates was performed as previously described (19) with modifications to accommodate an Agilent BioCel automated liquid handling platform (20). Briefly, single cell lysates contained in the 96-well sorted plates were processed in batches of eight plates, with each plate containing wells reserved for 10 pg Universal Human RNA (Clontech Cat#636538) as a positive control, an ERCC-only control, and water as a negative control. Smart-seq2 cDNA synthesis, reverse transcription, and PCR were carried out in a reduced volume (12.5 µL) and with ERCC internal controls spiked-in at a reduced concentration (55 million-fold dilution of the ERCC stock in the first strand cDNA synthesis step). Amplified cDNAs from the eight 96-well plates were consolidated to two 384-well plates and purified with Ampure magnetic particles. A 10-fold diluted portion of each cDNA was assessed for expression of the human beta-actin (ACTB) housekeeping gene by qPCR for quality control of the amplified cDNAs. A total of 14,592 sample wells were processed through cDNA synthesis and ACTB qPCR on the automated platform.

A cycle threshold (Ct) of ≤35 for ACTB amplification was used as a cutoff for the selection of 3,072 cDNAs (768 per cell type) for library preparation and sequencing. A Star liquid handling platform (Hamilton) was used to consolidate cDNAs selected for Illumina Nextera XT library preps (Illumina cat# FC-131-1096) into 384-well plates. An automated 1/8th Nextera XT reaction was carried out on 125 pg of the selected cDNAs for the Tn5 tagmentation step, with limited 15 cycle PCR followed by AmPure XP (Beckman Coulter Cat# A63881) bead purification. Nextera XT PCR was carried out with a combination of 384 barcode pairs using Nextera XT Index Kit V2 barcode sets A and D (Illumina cat# FC-131-2001 and -2004). Concentrations of the purified Nextera XT reactions were normalized to 1 ng/µL and combined into a 2ng pool of 384 dual-barcoded samples. RNA-seq was carried out with a total of eight 384 barcoded pools loaded across 16 lanes of an Illumina HiSeq 2500 according to manufacturer's specifications for a total of 3,072 samples sequenced, including controls. A HiSeq SBS V4 250 cycle kit (Illumina cat# FC-401-4003) and a Paired End V4 Cluster Kit (Illumina cat# PE-401-4001) was used for an estimated 2 million reads per sample.

Sequence Data Processing

Single cell RNA-seq data was processed according to published methods (19, 21). Briefly, raw fastq sequencing files were demultiplexing using Illumina barcodes. Sequencing primers and low-quality bases were removed using the Trimmomatic software package (22). Trimmed reads were then aligned using HISAT (23) in two steps: first to a reference of ERCC sequences, and then to GRCh38 (Ensembl). StringTie (23) was used to assemble the resulting alignments into transcript structures using GENCODE v25 annotation (Ensembl 87; 10-2016) and gene expression values (TPM) estimated. HTSeq-count (24) was used to generate raw gene alignment counts.

Quality control analysis was performed using sequencing and laboratory metrics, including average Phred score, percent duplicate reads, and transcript isoform count, to classify cell samples as pass or fail using a Random Forest quality control classification model previously described (21). Expression values for the top 2500 genes ranked based on variance from cell

samples that passed quality control classification were fed into Scanpy (25) for principal component analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP)-based non-linear dimensionality reduction and visualization (26, 27). Unsupervised clustering was performed for the entire dataset, while additional supervised clustering guided by flow cytometry marker panels was performed to investigate within cell type variation. Lastly, cell type marker determination was performed using the Louvain unsupervised clustering results and the NS-Forest algorithm (28, 29). The end result of this computational pipeline produced a set of unbiased cell type clusters, a gene expression matrix with the expression levels of genes in individual single cells grouped into cell type clusters, and a set of sensitive and specific marker genes for each cell type cluster (**Supplementary Table 2**) for use in downstream quantitative PCR assays and semantic representations (30).

qRT-PCR

Aliquots (2 µL) of the Smart-seq2 cDNAs from single sorted myeloid dendritic cells (mDCs) were diluted 10-fold in low TE (10mM Tris, 0.1mM EDTA) and 2.5 µL of the diluted cDNAs were subjected to 10 µL Taqman[TM] qPCR assays for the human beta actin (ACTB) housekeeping gene (ThermoFisher Hs01060665_g1 FAM-MGB) using 5 µL of a 2X PerfeCTa qPCR SuperMix ROX (Quantabio cat# 95050-500) as an initial screen for endogenous gene expression. Thermocycling conditions were completed on a Quantstudio 6 qPCR instrument (Applied Biosystems) using the following thermocycling profile: initial 95°C activation for 2 minutes followed by 45 cycles of 95°C for 10 seconds and 60°C for 30 seconds. Positive reactions - cycle threshold (Ct) of less than 35 for ACTB amplification - were identified and their corresponding cDNAs screened using two additional marker genes selected from the NS-Forest analysis, CDKN1C (ThermoFisher Hs00175938_m1 FAM-MGB) and NDRG2 (ThermoFisher Hs01045114_g1 FAM-MGB), using the same thermocyling conditions as for ACTB.

In Vitro Whole Blood Stimulation

Pre-vaccination (baseline) blood samples were stimulated *in vitro* (Milieu Interieur) with LPS (Invivogen - tlrl-3pelps), poly I:C (Invivogen - vac-pic), or SEB (Kindly given by Bernard Nocht Institute) with appropriate negative controls and incubated in TruCulture tubes within 15 minutes of blood collection, inserted into a dry block incubator, and maintained at 37°C (± 1°C) for 22 hours as described (13, 17). Cell fractions were collected and lysed in Trizol for RNA extraction. cDNA was prepared using the SmartSeq 2 protocol as described above. Quantitative PCR (qPCR) was performed in triplicates for each sample targeting CDKN1CC and NDRG2 using ACTB as a housekeeping gene. The data were analyzed using the standard delta-delta Ct method ($2^{-\delta\delta Ct}$) in order to generate fold difference in gene expression values.

mDC Functional Assessment

To assess the ability of mDCs to induce T cell activation, 50 ml of whole blood was drawn from healthy adult donors and PBMC isolated as previously described (17). Cells were stained using cocktails of surface marker antibodies (FCER1A, CD11c, CD1C, CD14, CD3, CD123, CD16A and HLADR) to specifically sort mDC2, mDC4 and T cells (**Supplementary Figures 3**, **4** and **Supplementary Table 3**). Cell sorting was performed using the BD Aria (II 85µm nozzle) in cold chambers. Sorted cells were spun down at 600g for 10 min, resuspended in 1ml PBS (Gibco # 14190), counted, and seeded in 96 well plates pre-filled with either LPS (100 ng/ml, Invivogen tlrl-3pelps) or polyI:C (20µg/ml, Invivogen Tlrt-pact) or medium as a negative control. The autologous T cells were labelled with Cell Trace [TM] Oregon Green diluted in PBS according to the manufacturer's instructions (Invitrogen # C34555), and then rested in AIM V medium (Gibco Cat# 12055-091) with 2% heat inactivated human AB serum (e.g., Gemini BioProducts) at 37°C for 24 hr. The next day, DC cultures were briefly treated with 20 mM of EDTA to detach adherent cells, all cells harvested, washed three times in complete medium to remove the TLR ligands, and counted. Dendritic cells (mDC2 or mDC4) were then mixed with labelled T cells at a ratio of 1:5 in 125µl complete AIM V medium and incubated at 37°C for 5 days. As a positive control, Oregon green labelled T cells were stimulated with 1µg/ml of anti CD3/28 antibody (Invitrogen #16-0037/#16-0288). On day 5, the cells were detached using EDTA, washed and stained with a cocktail of antibodies (**Supplementary Table 4**) to assess the proliferation of specific T cells using BD LSRII flow cytometer. All flow cytometry data were analyzed using Flowjo version 10 (Flowjo, Ashland, OR).

Diablo Supervised sGCCA

DIABLO, part of the mixOmics framework, is a data-driven, hypothesis-free multi-omics integration approach that has been successfully applied, by us and others, to derive novel, robust biomarkers, and increase our understanding of the molecular regulatory mechanisms that underlie health and disease (31–34). DIABLO extends sparse Generalized Canonical Correlation Analysis (sGCCA) into a supervised multi-omics data integration framework (35, 36). DIABLO performs multivariate dimensionality reduction and selects correlated variables from several datasets by maximizing the covariance between linear combinations of variables (latent component), across both multi-omics datasets (blocks) and an outcome (response) variable, in this case anti-HBV serum antibody titers. The data are then projected into a smaller dimensional subspace spanned by the latent components for classification. Here we used DIABLO to identify correlates of vaccine response (anti-HBV IgG level measured at the final follow-up visit, Day 208), from the multi-omics profiles in baseline pre-vaccination samples in an integrative fashion.

RESULTS

Variable Antibody Response to HBV Vaccination

A detailed description of the study design and sample collection strategy has been reported previously (17). Briefly, fifteen participants were given the standard three-dose HBV vaccination regimen and blood samples were collected before vaccination (Visit 3) and 28 days after each of the three vaccine doses (Visits 8, 10, and 12). As observed in previous studies (13),

a subset of three participants showed measurable Ab titers following a single HBV vaccine dose, with two of the three participants achieving protective Ab levels of >10 mIU/ml (**Figure 1**). Fourteen of the fifteen participants achieved protective antibody titers after the second and third doses, with a >2 log range in antibody titers after the third dose.

mDC Subsets With Distinct Gene Expression

Single cell RNA sequencing (scRNAseq) of innate immune cell subsets sorted from whole blood was used to define their transcriptional phenotypes with relationship to HBV vaccination responses. In order to ensure the capture of any distinct phenotype that might correlate with vaccine response, blood samples were collected on Day 0 pre-vaccination and Day 1, 3, 7, and 14 post-vaccination, from a subset of participants, including the two dose 1 responders (GR01, GR04), the one dose 1 marginal responder (GR15), and three dose 1 non-responders (GR13, GR17, GR19) according to the HBV-specific antibody titers measured at 28 days after the first vaccine dose.
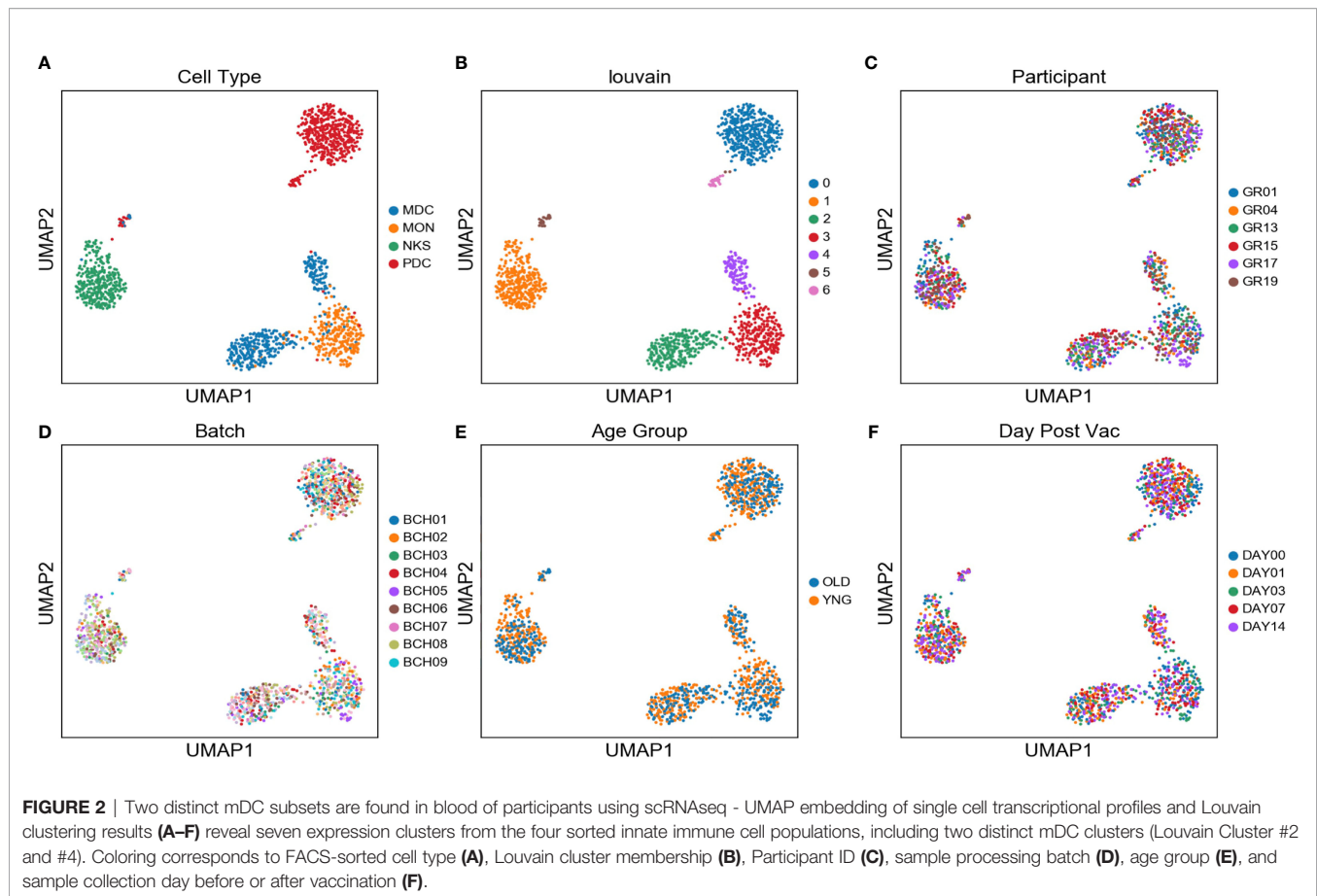
Single monocytes (MON), myeloid dendritic cells (mDC), plasmacytoid dendritic cells (pDC), and natural killer cells (NK) were sorted into microtiter plate wells and single cell cDNAs showing positive ACTB expression by qPCR analyzed by scRNAseq and UMAP embedding of the gene expression data. Each of the four major innate immune cell subsets were well

segregated in the UMAP plot (**Figure 2A**). In addition, lower abundance outlier clusters we also detected for the mDC and pDC sorted populations, indicating some level of subtype heterogeneity. Unsupervised clustering produced seven distinct transcriptome clusters (**Figure 2B**), including the lower abundance mDC and pDC outliers. No obvious cluster-specific enrichment of cells from individual participants, processing batch, age group or sample collection date was observed (**Figures 2C–F**).

Each unsupervised cell cluster showed distinct differential gene expression patterns identified using both logistic regression (**Figures 3A–C**) and NS-Forest-based marker gene selection (**Figures 3D, E** and **Supplementary Table 2**). The main mDC subset (Louvain cluster #2) appeared to exclusively express the p57 kip2 cyclin-dependent kinase inhibitor gene CDKN1C and expressed relatively high levels of LINC01272 in comparison with other innate cell subsets (**Figures 3A–E**). These cells also expressed high levels of the Fc gamma receptor gene FCGR3A (**Figures 3A–C**). In contrast, the outlier mDC cluster (Louvain cluster #4) exclusively expressed the n-myc regulated gene NDRG2 and expressed relatively high levels of the Fc epsilon receptor gene FCER1A, the MHC class II gene HLA-DQA1 (**Figures 3A–E**), and other MHC class II genes (not shown). The high-level expression of MHC class II genes suggests that the outlier mDC subset is activated, whereas the expression of the p57 kip2 CDK inhibitor suggests that the main mDC subset is resting. Expression of FCGR3A in the main mDC cluster and FCER1A in the outlier
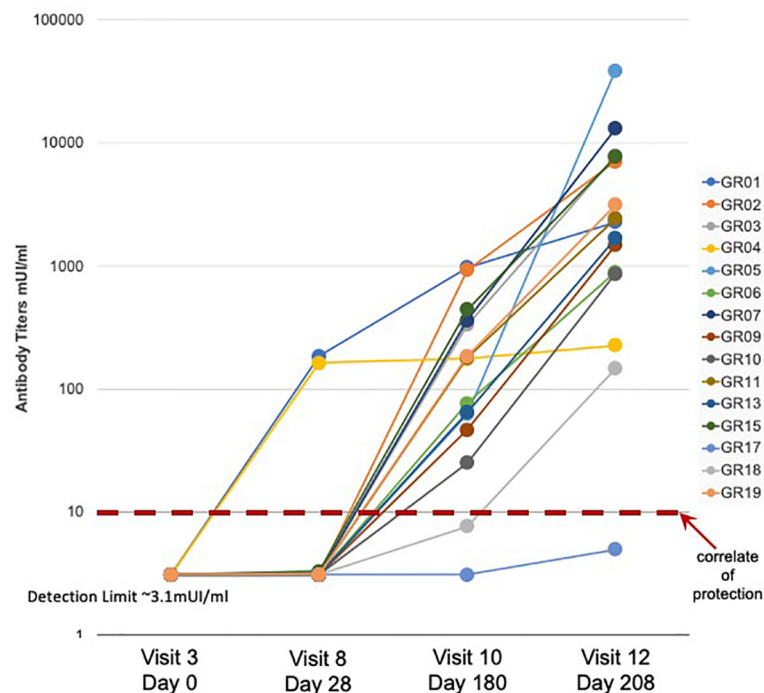
FIGURE 1 | Serum antibody response to HBV vaccination - Serum antibody titers were measured in samples collected from fifteen study participants (GR01 – GR19) before (Visit 3) and after (Visit 8, 10, and 12) vaccination. Vaccine doses were administered on Visit 3, Visit 8 and Visit 10 (after blood collection) for a total of three doses of HBV vaccine in all fifteen participants. Limit of detection of 3.1 mUI/ml and correlate of protection of 10 mUI/ml are indicated. These data were also previously used to create Figure 1C in (13).

FIGURE 2 | Two distinct mDC subsets are found in blood of participants using scRNAseq - UMAP embedding of single cell transcriptional profiles and Louvain clustering results **(A–F)** reveal seven expression clusters from the four sorted innate immune cell populations, including two distinct mDC clusters (Louvain Cluster #2 and #4). Coloring corresponds to FACS-sorted cell type **(A)**, Louvain cluster membership **(B)**, Participant ID **(C)**, sample processing batch **(D)**, age group **(E)**, and sample collection day before or after vaccination **(F)**.

mDC cluster suggests that these two subsets correspond to the DC4 and DC2 dendritic cell types defined previously (37) and will be referred to as CDKN1C-expressing mDC4 and NDRG2-expressing mDC2.

In order to determine if the NDRG2-expressing mDC2 and CDKN1C-expressing mDC4 phenotypes had been observed in previous studies, the NDRG2 and CDKN1 marker genes were used to search for expression modules in the MSigDB database. Four MSigDB modules included both of these marker genes:

- GSE17721_0.5H_VS_24H_POLYIC_BMDC_UP,
- GSE17721_0.5H_VS_8H_POLYIC_BMDC_DN,
- GSE17721_CTRL_VS_POLYIC_12H_BMDC_DN,
- GSE17721_POLYIC_VS_PAM3CSK4_12H_BMDC_UP),

which were all derived from dendritic cells stimulated with TLR3 agonists.

Relative Abundance of mDC Subsets Correlate With Vaccine Response

Exclusive expression of NDRG2 in mDC2 and CDKN1C in mDC4 suggested that these two markers could be used to distinguish these mDC subsets. Indeed, qPCR amplification showed mutually exclusive expression of these two genes in
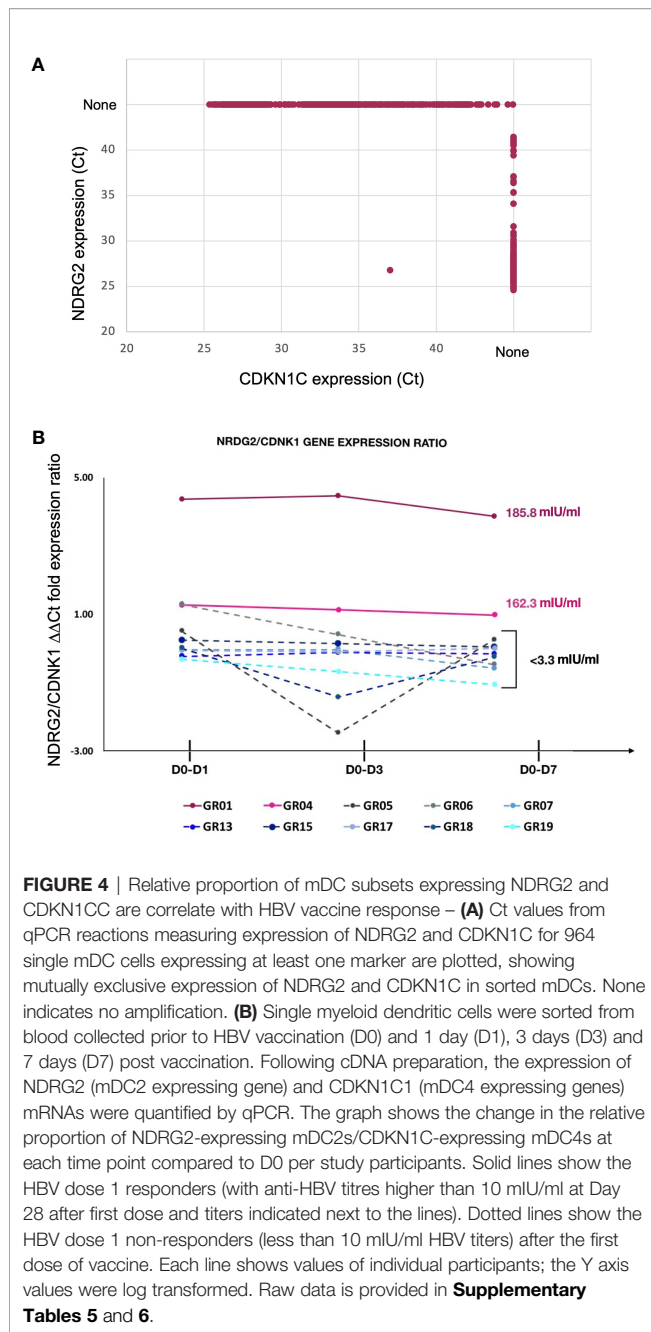
sorted mDCs (**Figure 4A**). Thus, qPCR for NDRG2 and CDKN1C was used to identify and quantify these two mDC subsets in ten of the participants across the entire time course of the study (**Supplementary Tables 5** and **6**). The relative proportions of NDRG2-expressing mDC2/CDKN1C-expressing mDC4 were found to be dynamic and vary between individuals (**Figure 4B**). Interestingly, a relatively high ratio of NDRG2-expressing mDC2/CDKN1C-expressing mDC4 was found in the two dose 1 responders (GR01 and GR04) at baseline Day 0 (**Supplementary Table 6**). Indeed, the average NDRG2-expressing mDC2/CDKN1C-expressing mDC4 ratio in dose 1 responders on Day 1 was 3.13 and in non-responders was 0.46. Interestingly, while the ratio of NDRG2-expressing mDC2/CDKN1C-expressing mDC4 dropped dramatically following vaccination of the two dose 1 responders, the ratio was relatively static or increased in non-responders (**Figure 4B** and **Supplementary Table 6**). While these findings in-and-of themselves are not adequately powered to draw definitive conclusions regarding these correlations due to the small sample size of this pilot, they did produce a hypothesis regarding the dispositional state of dendritic cells that could be explored in orthologous data.

Given that NDRG2-expressing mDC2 and CDKN1C-expressing mDC4 marker genes in the MSigDB database had
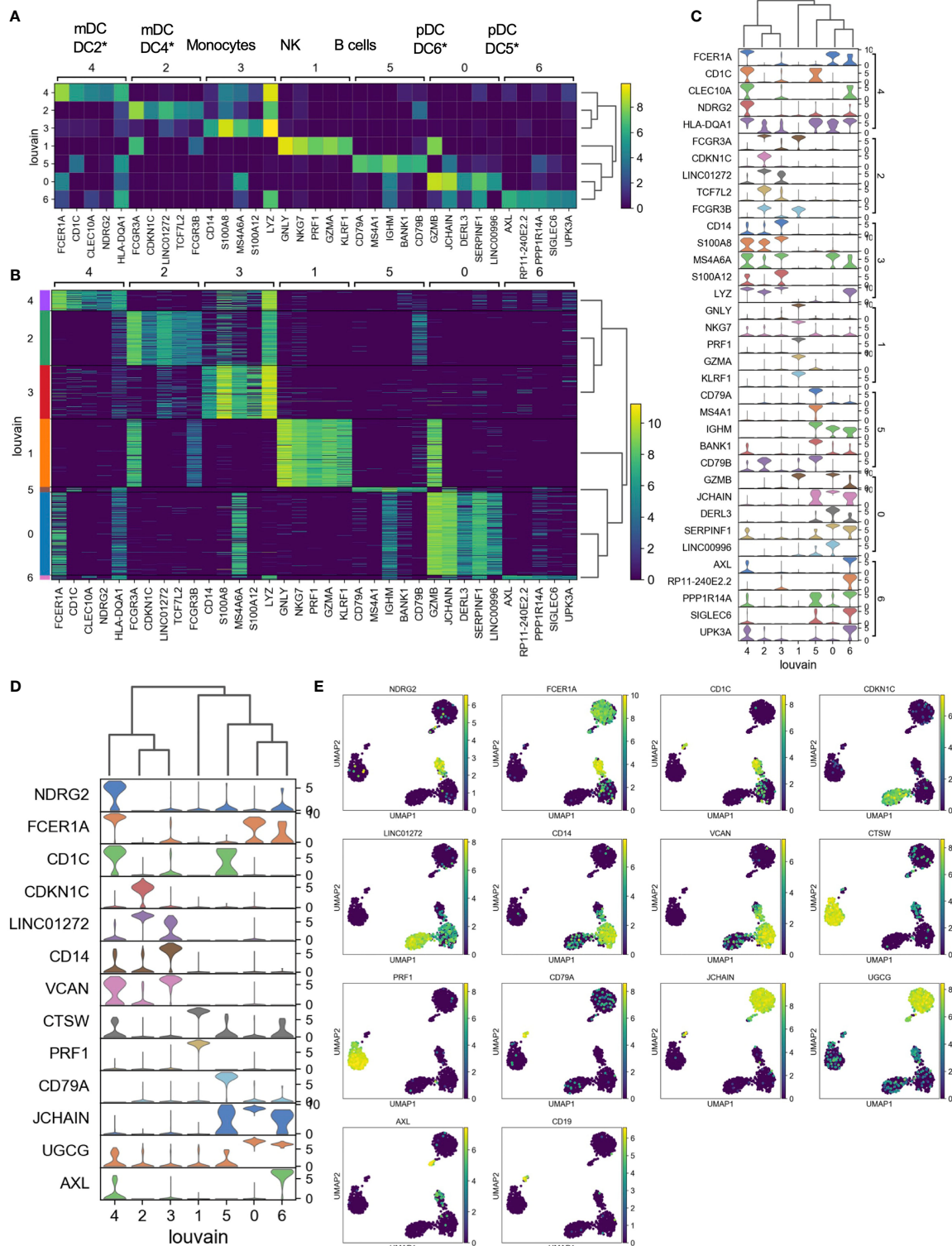
FIGURE 3 | Expression cluster marker genes – (A) The top five marker genes for each cluster was determined by logistic regression. Median expression of marker genes in cells within each cluster is shown. *Dendritic cell types reported in Villani et al. (37) were identified based on marker gene expression. (B) Expression of logistic regression marker genes in each individual cell within each cluster. (C) Violin plots showing logistic regression marker gene expression distributions. (D) Violin plots showing gene expression distributions for the minimum set of necessary and sufficient marker genes as determined using the NS-Forest algorithm. (E) Expression of NS-Forest marker genes in UMAP Louvain clusters.

FIGURE 4 | Relative proportion of mDC subsets expressing NDRG2 and CDKN1CC are correlate with HBV vaccine response – **(A)** Ct values from qPCR reactions measuring expression of NDRG2 and CDKN1C for 964 single mDC cells expressing at least one marker are plotted, showing mutually exclusive expression of NDRG2 and CDKN1C in sorted mDCs. None indicates no amplification. **(B)** Single myeloid dendritic cells were sorted from blood collected prior to HBV vaccination (D0) and 1 day (D1), 3 days (D3) and 7 days (D7) post vaccination. Following cDNA preparation, the expression of NDRG2 (mDC2 expressing gene) and CDKN1C1 (mDC4 expressing genes) mRNAs were quantified by qPCR. The graph shows the change in the relative proportion of NDRG2-expressing mDC2s/CDKN1C-expressing mDC4s at each time point compared to D0 per study participants. Solid lines show the HBV dose 1 responders (with anti-HBV titres higher than 10 mIU/ml at Day 28 after first dose and titers indicated next to the lines). Dotted lines show the HBV dose 1 non-responders (less than 10 mIU/ml HBV titers) after the first dose of vaccine. Each line shows values of individual participants; the Y axis values were log transformed. Raw data is provided in **Supplementary Tables 5** and **6**.

been identified as malleable to polyI:C stimulation, we further explored how TLR3 activation might lead to differential responses in these DC subsets, we examined the changes in expression of CDKN1C and NDRG2 in whole blood collected from participants prior to HBV vaccination and stimulated with a TLR3 agonist. Interestingly, the two participants that showed the highest Ab responses to the first vaccine dose – GR01 and GR04 – showed preferential up-regulation of NDRG2 in their blood cells, whereas cells from non-responding participants showed preferential up-regulation of CDKN1C suggesting they were predisposed to a more resting/inhibitory mDC4-like phenotype (**Supplementary Figure 5**).

mDC Subsets Differ in Their Functional Predisposition

To explore if there are any functional differences between these mDC subsets, NDRG2-expressing mDC2 and CDKN1C-expressing mDC4 subsets were sorted and incubated with sorted T cells from the same (autologous) donor, with or without pre-stimulation with TLR3 (pIC) or TLR4 (LPS) agonists prior to co-culture to assess the impact of TLR pre-stimulation on the ability to induce T cell proliferation as a proxy read out for immune activation as described (37). When unstimulated CDKN1C-expressing mDC4 cells were incubated with unstimulated autologous T cells, proliferation of 13.0% of CD4 and 15.8% of CD8 was induced (**Figures 5A, B**). However, if CDKN1C-expressing mDC4 cells were first stimulated with pI:C or LPS, this baseline T cell proliferation was inhibited 3.9-fold (mean n = 4) for both CD4 and CD8 T cells (paired t-test p = 0.030 and 0.036 respectively). In contrast, NDRG2-expressing mDC2 cells did not induce T cell proliferation with or without LPS stimulation (**Figure 5C**). These results suggest CDKN1C-expressing mDC4 cells exhibit tonic T cell activating ability and that this activating ability is suppressed following TLR3 or TLR4 stimulation, whereas NDRG2-expressing mDC2 do not activate T cell proliferation either with or without TLR stimulation.

Machine Learning Model Using mDC TLR3 Feature Selection Improves Serum Antibody Response Predictions

In our previous study, we found that multi-omics data could be used to produce predictive models of antibody titers based on the supervised sparse generalized canonical correlation analysis implemented in the Diablo algorithm (13), even given the relatively large feature space provided by the transcriptomic and CpG methylation data (**Figure 6A**). Based on the scRNAseq and functional studies described above, we hypothesized that if the relative abundance of the NDRG2-expressing mDC2 and CDKN1C-expressing mDC4 subsets were indeed mechanistically linked to vaccine responses, selecting features specific to these cell subsets might produce improved correlation with serum antibody levels.

Using just the bulk gene expression and DNA methylation data from baseline pre-vaccination samples associated with the MSigDB marker genes derived from dendritic cells stimulated with TLR3 agonists to build the Diablo predictive model for all fifteen participants, a significant improvement in model performance was obtained (**Figure 6B**). For example, the Spearman's rank correlation of the gene expression model improved from 0.62 to 0.87 and the median error improved from 15.66% to 7.23%. Similar improvements were observed with the model produced using the selected CpG DNA methylation features. And while cross-validation did not show a significant improvement, likely due to the small sample size, we went from ~50,000 transcripts and ~800,000 CpG sites to only a few hundred transcripts and a few thousand CpG sites as a result of this feature selection step demonstrating a substantial enrichment of informative features.
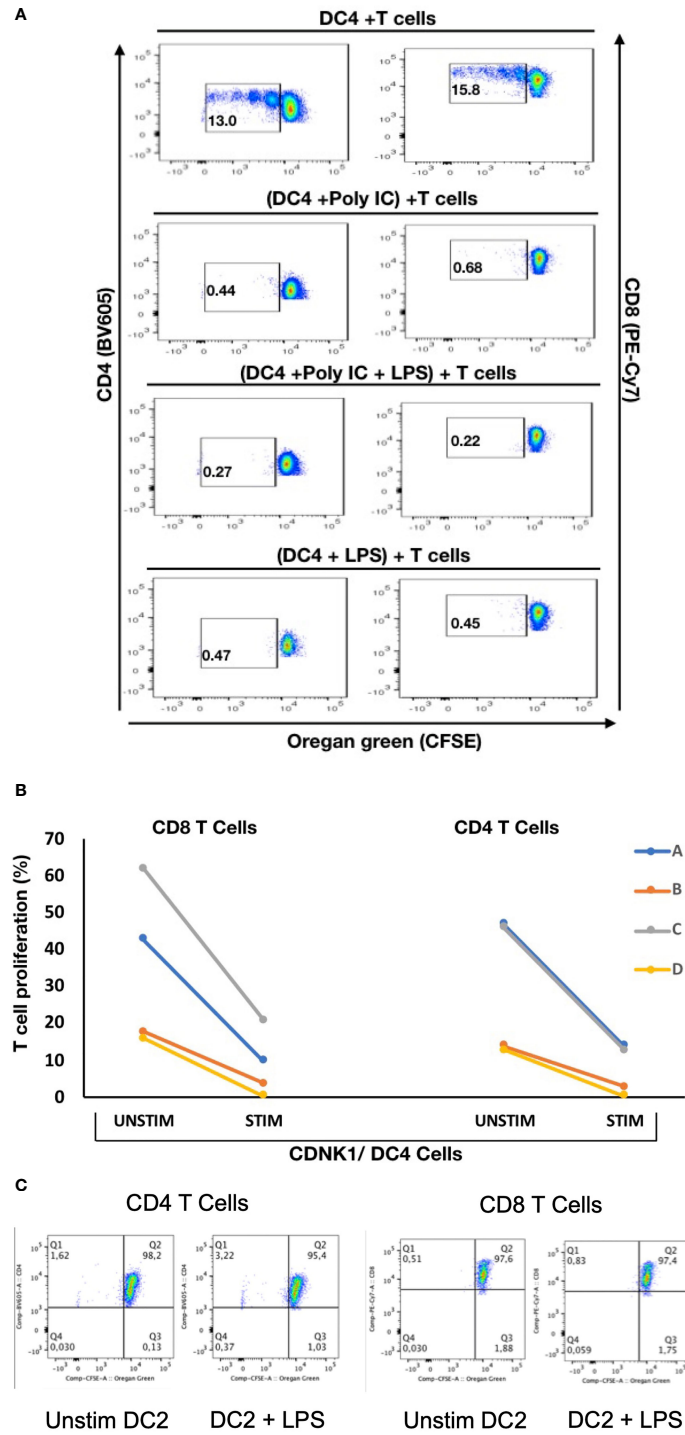
FIGURE 5 | Stimulation of CDKN1C-expressing mDC4 cells with TLR agonists suppresses T cell activation function – **(A)** CDKN1C/mDC4 cells were sorted as viable, singlets, CD14[-], CD19[-], CD3[-], HLA-DR[int-to-high], CD11c[+], CD16[+] cells, stimulated with polyI:C and/or LPS and mixed with autologous T cells labelled with Oregon Green (OG) from the same patient. (Note that CD16 is encoded by FCGR3A expressed in CDKN1C-expressing mDC4.) Following 5 days, samples were analyzed by flow cytometry. Proliferating T cells (lower OG staining) were gated. The pseudocolor dot plots are representative of 4 different experiments. **(B)** Percent of CD4 and CD8 T cells proliferating using sorted cells from four different individuals **(A–D)** with or without stimulated. Proliferation was assessed using OG staining after co-culture of autologous T cells with CDKN1C-expressing mDC4 cells for 5 days induced by mDC4 either unstimulated or pre-stimulated with the TLR3 agonist polyI:C. **(C)** mDC2 stimulation of autologous T cells - T cell proliferation assays with unstimulated or LPS-stimulated mDC2 dendritic cells.
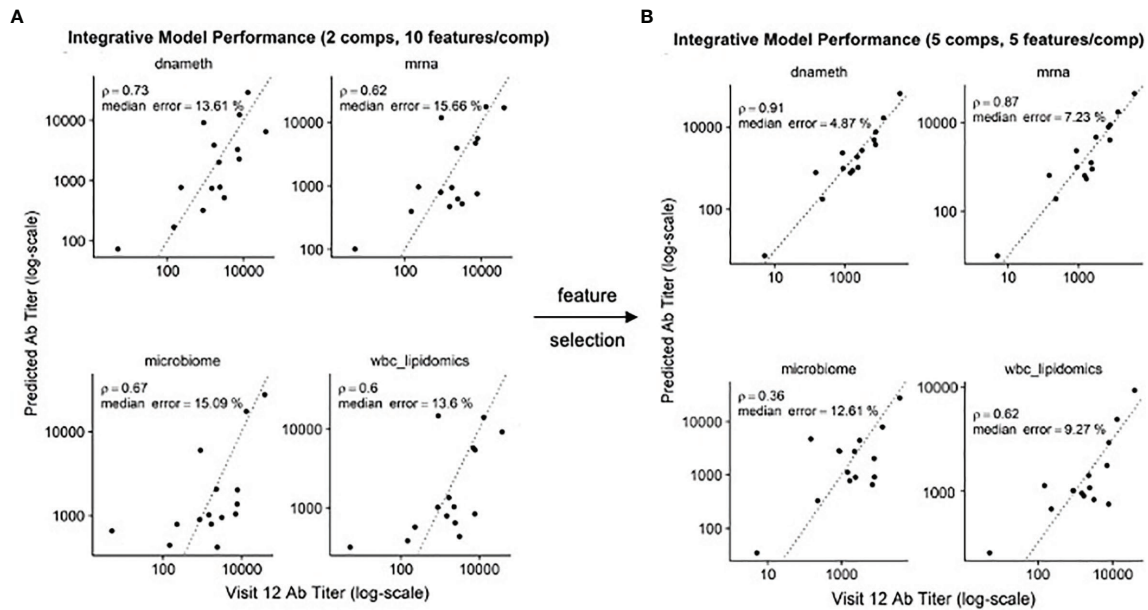
FIGURE 6 | Improved correlation of Diablo models with serum antibody (Ab) responses to HBV vaccination – Actual Ab titers at Visit 12 (after third dose) (x-axis) vs. predicted Ab titers (y-axis) in models derived from different assay platforms are visualized. Dotted line is the identity line representing perfect prediction. Rho is Spearman's rank correlation when comparing actual Ab titers to predicted Ab titers. **(A)** Optimal Diablo models were produced using 2 components and 10 features/component and all available assay variables. **(B)** Optimal Diablo models were produced using 5 components and 5 features/component and only selected variables related to 735 dendritic cell TLR3-response genes in the DNA methylation (dnameth) and bulk transcriptomic (mrna) data. CpG sites were assigned to dendritic cell TLR3-response genes as described previously (38, 39). Note that although the microbiome and wbc-lipidomics data are identical in the two sets, the models they produce (features retained and their coefficients) in the generalized canonical correlation framework are different due to different correlation characteristics with the different dnameth and mrna models. In both cases, the number of components and number of features/component were selected to maximize model performance.

DISCUSSION

In a relatively small cohort of HBV vaccine recipients, we identified two distinct mDC subsets using single cell transcriptomics analysis, the ratio of which at baseline (i.e., before vaccination) correlated with vaccine response to a single dose of the HBV vaccine. These two mDC subsets were distinguishable by the differential expression of a number of genes that allowed for their putative matching to dendritic cell subsets identified previously (37), designated here as NDRG2-expressing mDC2 and CDKN1C-expressing mDC4.

Three pieces of evidence suggest that these dendritic cell subsets contribute to the immune state at baseline prior to vaccination that influences vaccine responses. First, the two individuals who generated serum antibody responses to a single dose of HBV vaccine had a high ratio of NDRG2-expressing mDC2/CDKN1C-

These results indicate that feature selection based on TLR3-induced dendritic cell genes produced machine learning models that provided better correlation with serum antibody responses, suggesting that the relative contributions of the NDRG2-expressing mDC2 and CDKN1C-expressing mDC4 subsets influences this vaccine response.

expressing mDC4 in their peripheral blood prior to vaccination. Second, whole blood from these responding individuals showed preferential upregulation of NDRG2 when stimulated with a TLR3 agonist. Third, machine learning developed to predict serum antibody titers after the third vaccine dose in all fifteen participants and built using baseline pre-vaccination sample data and genes differentially expressed in dendritic cells stimulated with TLR3 agonists outperformed models built without this feature selection step. Thus, the dispositional state of these dendritic cell subsets at baseline appears to provide improved correlation with HBV vaccine responses.

The mechanism of how NDRG2-expressing mDC2 and CDKN1C-expressing mDC4 impact the vaccine response to the HBV vaccine is not yet clear. But, interestingly, *ex vivo* CDKN1C-expressing mDC4 cells were able to induce autologous CD4 and CD8 T cell proliferation yet mDC2 did not; and TLR3 or TLR4 stimulation of the mDC4 subset inhibited this T cell stimulation. Such a high state of functional plasticity in the mDC4 subset, and the longitudinal variation over time in the ratio of NDRG2-expressing mDC2 and CDKN1C-expressing mDC4 in whole peripheral blood may at least partially explain the variability in immune responses to the HBV vaccine.

Several other groups have recently explored the identification of baseline predictors of vaccine responses using multi-omics assays with results that differ from those reported here. Tsang

et al. used multi-omics assays to explore responses to seasonal influenza vaccine in healthy adults (8). While neither Day 0 gene expression nor pathway analysis alone was predictive of vaccine responses in their study, twelve cell populations assessed by flow cytometry, including memory, naive, and transitional B cells, CD4 effector memory T cells, IFNa+ myeloid dendritic cells (mDC), and several activated T cell populations correlated with mean fold change in antibody titers. However, the responses in these healthy influenza vaccine cohorts likely represent a recall response to antigenically-similar prior exposure as opposed to the primary HBV response in our study. Fourati et al. constructed a naive Bayes classifier based on the top 15 differentially-expressed genes between PBMCs from responders and poor-responders to the HBV vaccines, which included B cell markers (e.g., CD20, IGHG1), downstream targets of B-cell receptor signaling (e.g., BANK1) and molecules known to have functional interactions with IgG (e.g., C1, FCGR3B), with a predictive accuracy of ~63% (9). However, the use of bulk transcriptomic analysis of PBMCs may have obscured the contribution of the minor dendritic cell components evaluated in our scRNAseq assays, emphasizing the value of using scRNAseq to assess the contribution of rare cell subsets. Bartholomeus et al. found that the GRN and IFITM1 genes were significantly downregulated in responders while upregulated in non-responders by whole blood transcriptomics analysis, and absolute granulocytes numbers were significantly higher in non-responders at Day 0 prior to vaccination with ENGERIX®-B (11). However, the role of the dendritic cell component was not evaluated.

Our results suggest that the participants in our vaccine study existed in one of two different dendritic cell dispositional states at baseline – an NDRG2-mDC2 state in which the vaccine elicited an early antibody response or a CDKN1C-mDC4 state in which the vaccine response may have been actively suppressed. While the possibility that challenge with a foreign antigen (e.g., a vaccine) would be immunosuppressive seems counterintuitive, a healthy immune system has to strike a delicate balance between responsiveness and non-responsiveness under many circumstances. Furthermore, ample data now supports the conclusion that for some vaccines, including ENGERIX®-B, reducing general immune activation and inflammation may in fact increase the antigen-specific response to the vaccine (9, 40).

From a biochemical perspective, foreign antigens are not that different from self-antigens. In order to avoid autoimmunity, the immune system must carefully assess whether an antigen is truly foreign or not. Evidence suggests that the setting in which the naïve immune system experiences antigen may play an important role and is only activated when some type of tissue injury signal accompanies antigen exposure, the so-called "danger hypothesis" (41). In the absence of this danger signal, activation of adaptive immune cells with Signal 1 from antigen without Signal 2 from antigen presenting cell help would be tolerated and not result in activation. And even when the immune system responds to a truly foreign antigen, an overexuberant response (e.g., the cytokine storm) can do more harm than good. Thus, in addition to mechanisms designed to activate an immune response, the immune system has also developed mechanisms for dampening the response, with the regulatory T cell being a classic example. Perhaps the CDKN1C-mDC4 cell is an example of a suppressive type of regulatory dendritic cell. The extent to which these CDKN1C-mDC4s may be similar to the myeloid-derived suppressor cells (MDSCs) observed in some pathological conditions, such as inflammation, chronic infection or cancer (42–45) remains to be determined.

So, what are the implications of this hypothesis for improving vaccination outcomes? We found that the NDRG2-mDC2/CDKN1C-mDC4 ratio differed between individuals as well as within the same individual over time. This suggests that perhaps the dendritic cell dispositional state could be modulated to establish an activatable predisposition prior to vaccination. Since adjuvant effects appear to function through the innate immune system, perhaps prior exposure to adjuvant *before* antigen could establish the appropriate activatable predisposition. While there is some evidence that preconditioning injection sites with TLR agonists can enhance dendritic cell migration (46) and protection against pathogen infection (47) in animal models, to our knowledge this has never been formally assessed in humans (6). This possibility could lead to a more precision-medicine approach for vaccines by determining at baseline who will respond well or not to specific vaccines or who might need just a single dose (48). This approach could readily follow the model of point-of-care testing currently used in infectious disease settings [e.g (49)]. In fact, the field of vaccinology already does assign different vaccines based on individual characteristics, e.g., different flu vaccines are given to different people based on age.

Given the relatively small cohort size of this pilot study, it was not possible to draw definite conclusions about the ability to predict dose 1 vaccine responses. There are many ways to follow up on the findings described here. One approach would be a direct expansion of the study. First, we would need to recruit a larger cohort of patients (100-200) and follow the visitation strategy and HBV serum Ab testing outlined in Ben-Othman (17) and Shannon (13). Pre-vaccination, we would obtain both HPV serum antibody levels and FACS sorted mDC populations for all patients. Next these mDC populations could be profiled by qPCR using probes to the marker genes discovered during this investigation: NDRG2 and CDKN1C. Lastly, response to vaccination would be measured by the antibody titers as performed in this study. This experimental design would now be possible from a cost perspective as profiling many cells from many patients by qPCR is relatively cheap, while using single cell RNAseq on this many patients would be cost prohibitive.

Finally, we described a novel machine learning approach to multi-omics data integration. The results reported by Shannon et al. (13) suggested that because different sources of background noise and technical confounders would contribute to the results from different omics assays, focusing on the consensus information related to outcome using the canonical correlation analysis approach to multi-omics data integration could reduce overfitting and result in more robust and generalizable models (32, 34), even in studies where p>>n. However, the number of

parameter features available in these systems biology studies makes it impossible to complete an exhaustive search of all available parameter combinations and makes the L1 penalized or LASSO regression implemented in Diablo ineffective at mitigating the effects of noisy features (50). Using single cell RNA sequencing in this study, we were able to identify relatively rare dendritic cell populations whose abundance and activation disposition appear to correlate with vaccine responses. By using this finding to guide feature selection for those genes expressed in these dendritic cell subsets or those CpG sites that are involved in establishing cell type identity, the correlations of the vaccine response predictive models were dramatically improved, demonstrating the value of directed feature selection prior to machine learning model production to further circumvent the p>>n problem.

In conclusion, the machine learning approaches for informative feature selection based on NS-Forest and multi-omics data integration based on supervised canonical correlation analysis not only produced better correlation with vaccine response but also revealed the possible cellular mechanisms responsible. These results suggest that vaccine recipients exist as different dispositional endotypes that dictate their response to vaccination. With a hypothetical mechanism in hand, developing strategies to adjust these dispositional endotypes in preference of dendritic cell-mediated activation rather than suppression could lead to the development of more effective precision vaccination strategies to achieve protective immunity from single vaccine doses, which are of critical importance in resource-limited settings.

DATA AVAILABILITY STATEMENT

Raw and expression matrix for the Single cell RNAseq data associated with this study are available at dbGaP accession number phs002508.v1.p1, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002508.v1.p1.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of British Columbia Children's & Women's Research Ethics Board (Ref: H17-00175). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BA and YZ performed data analysis and interpretation of results. CS, MN, RB-O, BC, JY, MK, NG, AHL, TB, RH, AL, and DD participated in planning and performing the experiments. WK, MS, ST, TK, and RS conceived and planned the experiments. All authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.690470/full#supplementary-material

REFERENCES

1. Koff WC, Burton DR, Johnson PR, Walker BD, King CR, Nabel GJ, et al. Accelerating Next-Generation Vaccine Development for Global Disease Prevention. *Science* (2013) 340(6136):1232910. doi: 10.1126/science.1232910
2. Raeven RHM, van Riet E, Meiring HD, Metz B, Kersten GFA. Systems Vaccinology and Big Data in the Vaccine Development Chain. *Immunology* (2019) 156(1):33–46. doi: 10.1111/imm.13012
3. Sharma M, Krammer F, García-Sastre A, Tripathi S. Moving From Empirical to Rational Vaccine Design in the 'Omics' Era. *Vaccines (Basel)* (2019) 7 (3):89. doi: 10.3390/vaccines7030089
4. Kollmann TR, Marchant A, Way SS. Vaccination Strategies to Enhance Immunity in Neonates. *Science* (2020) 368(6491):612–5. doi: 10.1126/science.aaz9447
5. Wimmers F, Pulendran B. Emerging Technologies for Systems Vaccinology - Multi-Omics Integration and Single-Cell (Epi)Genomic Profiling. *Curr Opin Immunol* (2020) 65:57–64. doi: 10.1016/j.coi.2020.05.001
6. Tsang JS, Dobaño C, VanDamme P, Moncunill G, Marchant A, Othman RB, et al. Improving Vaccine-Induced Immunity: Can Baseline Predict Outcome? *Trends Immunol* (2020) 41(6):457–65. doi: 10.1016/j.it.2020.04.001
7. Warimwe GM, Fletcher HA, Olotu A, Agnandji ST, Hill AV, Marsh K, et al. Peripheral Blood Monocyte-to-Lymphocyte Ratio at Study Enrollment Predicts Efficacy of the RTS,S Malaria Vaccine: Analysis of Pooled Phase II Clinical Trial Data. *BMC Med* (2013) 11:184. doi: 10.1186/1741-7015-11-184
8. Tsang JS, Schwartzberg PL, Kotliarov Y, Biancotto A, Xie Z, Germain RN, et al. Global Analyses of Human Immune Variation Reveal Baseline Predictors of Postvaccination Responses. *Cell* (2014) 157(2):499–513. doi: 10.1016/j.cell.2014.03.031.Erratumin:Cell
9. Fourati S, Cristescu R, Loboda A, Talla A, Filali A, Railkar R, et al. Pre-Vaccination Inflammation and B-Cell Signalling Predict Age-Related Hyporesponse to Hepatitis B Vaccination. *Nat Commun* (2016) 7:10369. doi: 10.1038/ncomms10369
10. HIPC-CHI Signatures Project Team and HIPC-I Consortium. Multicohort Analysis Reveals Baseline Transcriptional Predictors of Influenza Vaccination Responses. *Sci Immunol* (2017) 2(14):eaal4656. doi: 10.1126/sciimmunol.aal4656
11. Bartholomeus E, De Neuter N, Meysman P, Suls A, Keersmaekers N, Elias G, et al. Transcriptome Profiling in Blood Before and After Hepatitis B Vaccination Shows Significant Differences in Gene Expression Between Responders and Non-Responders. *Vaccine* (2018) 36(42):6282–9. doi: 10.1016/j.vaccine.2018.09.001
12. Qiu S, He P, Fang X, Tong H, Lv J, Liu J, et al. Significant Transcriptome and Cytokine Changes in Hepatitis B Vaccine non-Responders Revealed by Genome-Wide Comparative Analysis. *Hum Vaccin Immunother* (2018) 14 (7):1763–72. doi: 10.1080/21645515.2018.1450122
13. Shannon CP, Blimkie TM, Ben-Othman R, Gladish N, Amenyogbe N, Drissler S, et al. Multi-Omic Data Integration Allows Baseline Immune Signatures to

Predict Hepatitis B Vaccine Response in a Small Cohort. *Front Immunol* (2020) 11:578801. doi: 10.3389/fimmu.2020.578801

14. Jack AD, Hall AJ, Maine N, Mendy M, Whittle HC. What Level of Hepatitis B Antibody Is Protective? *J Infect Dis* (1999) 179(2):489–92. doi: 10.1086/314578

15. Tsang JS. Utilizing Population Variation, Vaccination, and Systems Biology to Study Human Immunology. *Trends Immunol* (2015) 36(8):479–93. doi: 10.1016/j.it.2015.06.005

16. Schillie SF, Murphy TV. Seroprotection After Recombinant Hepatitis B Vaccination Among Newborn Infants: A Review. *Vaccine* (2013) 31 (21):2506–16. doi: 10.1016/j.vaccine.2012.12.012

17. Ben-Othman R, Cai B, Liu AC, Varankovich N, He D, Blimkie TM, et al. Systems Biology Methods Applied to Blood and Tissue for a Comprehensive Analysis of Immune Response to Hepatitis B Vaccine in Adults. *Front Immunol* (2020) 11:580373. doi: 10.3389/fimmu.2020.580373

18. Smolen KK, Cai B, Kollmann TR. OMIP-038: Innate Immune Assessment With a 14-Color Flow Cytometry Panel. *Cytometry A* (2017) 91(10):966–8. doi: 10.1002/cyto.a.23109

19. Krishnaswami SR, Grindberg RV, Novotny M, Venepally P, Lacar B, Bhutani K, et al. Using Single Nuclei for RNA-Seq to Capture the Transcriptome of Postmortem Neurons. *Nat Protoc* (2016) 11(3):499–524. doi: 10.1038/nprot.2016.015

20. McLean JS, Lombardo MJ, Ziegler MG, Novotny M, Yee-Greenbaum J, Badger JH, et al. Genome of the Pathogen Porphyromonas Gingivalis Recovered From a Biofilm in a Hospital Sink Using a High-Throughput Single-Cell Genomics Platform. *Genome Res* (2013) 23(5):867–77. doi: 10.1101/gr.150433.112

21. Aevermann B, McCorrison J, Venepally P, Hodge R, Bakken T, Miller J, et al. Production of a Preliminary Quality Control Pipeline for Single Nuclei RNA-Seq and Its Application in the Analysis of Cell Type Diversity of Post-Mortem Human Brain Neocortex. *Pac Symp Biocomput* (2017) 22:564–75. doi: 10.1142/9789813207813_0052

22. Bolger AM, Lohse M, Usadel B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* (2014) 30(15):2114–20. doi: 10.1093/bioinformatics/btu170

23. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-Level Expression Analysis of RNA-Seq Experiments With HISAT, StringTie and Ballgown. *Nat Protoc* (2016) 11(9):1650–67. doi: 10.1038/nprot.2016.095

24. Anders S, Pyl PT, Huber W. HTSeq–A Python Framework to Work With High-Throughput Sequencing Data. *Bioinformatics* (2015) 31(2):166–9. doi: 10.1093/bioinformatics/btu638

25. Wolf FA, Angerer P, Theis FJ. SCANPY: Large-Scale Single-Cell Gene Expression Data Analysis. *Genome Biol* (2018) 19(1):15. doi: 10.1186/s13059-017-1382-0

26. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* (2018). doi: 10.21105/joss.00861

27. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP. *Nat Biotechnol* (2018) 37:38–44. doi: 10.1038/nbt.4314

28. Aevermann BD, Novotny M, Bakken T, Miller JA, Diehl AD, Osumi-Sutherland D, et al. Cell Type Discovery Using Single-Cell Transcriptomics: Implications for Ontological Representation. *Hum Mol Genet* (2018) 27(R1): R40–7. doi: 10.1093/hmg/ddy100

29. Aevermann BD, Zhang Y, Novotny M, Keshk M, Bakken TE, Miller JA, et al. A Machine Learning Method for the Discovery of Minimum Marker Gene Combinations for Cell-Type Identification From Single-Cell RNA Sequencing. *Genome Res* (2021) gr.275569.121. doi: 10.1101/gr.275569.121

30. Bakken T, Cowell L, Aevermann BD, Novotny M, Hodge R, Miller JA, et al. Cell Type Discovery and Representation in the Era of High-Content Single Cell Phenotyping. *BMC Bioinf* (2017) 18(Suppl 17):559. doi: 10.1186/s12859-017-1977-1

31. Langenberg MCC, Hoogerwerf MA, Koopman JPR, Janse JJ, Kos-van Oosterhoud J, Feijt C, et al. A Controlled Human Schistosoma Mansoni Infection Model to Advance Novel Drugs, Vaccines and Diagnostics. *Nat Med* (2020) 26(3):326–32. doi: 10.1038/s41591-020-0759-x

32. Lee AH, Shannon CP, Amenyogbe N, Bennike TB, Diray-Arce J, Idoko OT, et al. Dynamic Molecular Changes During the First Week of Human Life Follow a Robust Developmental Trajectory. *Nat Commun* (2019) 10(1):1092. doi: 10.1038/s41467-019-08794-x

33. Rohart F, Gautier B, Singh A, Lê Cao KA. Mixomics: An R Package for 'Omics Feature Selection and Multiple Data Integration. *PLoS Comput Biol* (2017) 13 (11):e1005752. doi: 10.1371/journal.pcbi.1005752

34. Singh A, Shannon CP, Gautier B, Rohart F, Vacher M, Tebbutt SJ, et al. DIABLO: An Integrative Approach for Identifying Key Molecular Drivers From Multi-Omics Assays. *Bioinformatics* (2019) 35(17):3055–62. doi: 10.1093/bioinformatics/bty1054

35. Lê Cao KA, Boitard S, Besse P. Sparse PLS Discriminant Analysis: Biologically Relevant Feature Selection and Graphical Displays for Multiclass Problems. *BMC Bioinf* (2011) 12:253. doi: 10.1186/1471-2105-12-253

36. Tenenhaus A, Philippe C, Guillemot V, Le Cao KA, Grill J, Frouin V. Variable Selection for Generalized Canonical Correlation Analysis. *Biostatistics* (2014) 15(3):569–83. doi: 10.1093/biostatistics/kxu001

37. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-Cell RNA-Seq Reveals New Types of Human Blood Dendritic Cells, Monocytes, and Progenitors. *Science* (2017) 356(6335):eaah4573. doi: 10.1126/science.aah4573

38. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical Evaluation of the Illumina MethylationEPIC BeadChip Microarray for Whole-Genome DNA Methylation Profiling. *Genome Biol* (2016) 17 (1):208. doi: 10.1186/s13059-016-1066-1

39. Price ME, Cotton AM, Lam LL, Farré P, Emberly E, Brown CJ, et al. Additional Annotation Enhances Potential for Biologically-Relevant Analysis of the Illumina Infinium HumanMethylation450 BeadChip Array. *Epigenet Chromatin* (2013) 6(1):4. doi: 10.1186/1756-8935-6-4

40. Alter G, Sekaly RP. Beyond Adjuvants: Antagonizing Inflammation to Enhance Vaccine Immunity. *Vaccine* (2015) 33(Suppl 2):B55–9. doi: 10.1016/j.vaccine.2015.03.058

41. Matzinger P. Tolerance, Danger, and the Extended Family. *Annu Rev Immunol* (1994) 12:991–1045. doi: 10.1146/annurev.iy.12.040194.005015

42. Palucka K, Ueno H, Fay J, Bancherceau J. Dendritic Cells and Immunity Against Cancer. *J Intern Med* (2011) 269(1):64–73. doi: 10.1111/j.1365-2796.2010.02317.x

43. Sendo S, Saegusa J, Morinobu A. Myeloid-Derived Suppressor Cells in non-Neoplastic Inflamed Organs. *Inflamm Regener* (2018) 38:19. doi: 10.1186/s41232-018-0076-7

44. Dorhoi A, Du Plessis N. Monocytic Myeloid-Derived Suppressor Cells in Chronic Infections. *Front Immunol* (2018) 8:1895. doi: 10.3389/fimmu.2017.01895

45. Karin N. The Development and Homing of Myeloid-Derived Suppressor Cells: From a Two-Stage Model to a Multistep Narrative. *Front Immunol* (2020) 11:557586. doi: 10.3389/fimmu.2020.557586

46. Tripp CH, Ebner S, Ratzinger G, Romani N, Stoitzner P. Conditioning of the Injection Site With CpG Enhances the Migration of Adoptively Transferred Dendritic Cells and Endogenous CD8+ T-Cell Responses. *J Immunother* (2010) 33(2):115–25. doi: 10.1097/CJI.0b013e3181b8ef5f

47. Lipford GB, Sparwasser T, Zimmermann S, Heeg K, Wagner H. CpG-DNA-Mediated Transient Lymphadenopathy Is Associated With a State of Th1 Predisposition to Antigen-Driven Responses. *J Immunol* (2000) 165(3):1228–35. doi: 10.4049/jimmunol.165.3.1228

48. Poland GA, Ovsyannikova IG, Kennedy RB. Personalized Vaccinology: A Review. *Vaccine* (2018) 36(36):5350–7. doi: 10.1016/j.vaccine.2017.07.062

49. Pennisi I, Rodriguez-Manzano J, Moniri A, Kaforou M, Herberg JA, Levin M, et al. Translation of a Host Blood RNA Signature Distinguishing Bacterial From Viral Infection Into a Platform Suitable for Development as a Point-of-Care Test. *JAMA Pediatr* (2021) 4:e205227. doi: 10.1001/jamapediatrics.2020.5227

50. Debashis P, Bair E, Hastie T, Tibshirani R. "Preconditioning" for Feature Selection and Regression in High-Dimensional Problems. *Ann Stat* (2008) 36 (4):1595–618. doi: 10.1214/009053607000000578

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in