



## Research Paper

## Nasopharyngeal Protein Biomarkers of Acute Respiratory Virus Infection



Thomas W. Burke<sup>a</sup>, Ricardo Henao<sup>a,d</sup>, Erik Soderblom<sup>f</sup>, Ephraim L. Tsalik<sup>a,b,c</sup>, J. Will Thompson<sup>f</sup>, Micah T. McClain<sup>a,c,e</sup>, Marshall Nichols<sup>a</sup>, Bradley P. Nicholson<sup>b</sup>, Timothy Veldman<sup>a</sup>, Joseph E. Lucas<sup>a,d</sup>, M. Arthur Moseley<sup>a,f</sup>, Ronald B. Turner<sup>g</sup>, Robert Lambkin-Williams<sup>h</sup>, Alfred O. Hero III<sup>i</sup>, Christopher W. Woods<sup>a,c,e,\*\*</sup>, Geoffrey S. Ginsburg<sup>a,\*</sup>

<sup>a</sup> Center for Applied Genomics and Precision Medicine, Department of Medicine, Duke University, Durham, NC 27708, USA

<sup>b</sup> Durham Veteran's Affairs Medical Center, Durham, NC 27705, USA

<sup>c</sup> Division of Infectious Diseases and International Health, Department of Medicine, Duke University, Durham, NC 27710, USA

<sup>d</sup> Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA

<sup>e</sup> Section for Infectious Diseases, Medicine Service, Durham Veteran's Affairs Medical Center, Durham, NC 27705, USA

<sup>f</sup> Proteomics and Metabolomics Shared Resource, Duke University Medical Center, Durham, NC 27708, USA

<sup>g</sup> School of Medicine, University of Virginia, Charlottesville, VA 22908, USA

<sup>h</sup> hVIVO, London, United Kingdom

<sup>i</sup> Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA

## ARTICLE INFO

## Article history:

Received 12 October 2016

Received in revised form 13 February 2017

Accepted 15 February 2017

Available online 21 February 2017

## Keywords:

Infectious disease

Influenza

Human rhinovirus

Proteomics

Diagnostic biomarker

## ABSTRACT

Infection of respiratory mucosa with viral pathogens triggers complex immunologic events in the affected host. We sought to characterize this response through proteomic analysis of nasopharyngeal lavage in human subjects experimentally challenged with influenza A/H3N2 or human rhinovirus, and to develop targeted assays measuring peptides involved in this host response allowing classification of acute respiratory virus infection. Unbiased proteomic discovery analysis identified 3285 peptides corresponding to 438 unique proteins, and revealed that infection with H3N2 induces significant alterations in protein expression. These include proteins involved in acute inflammatory response, innate immune response, and the complement cascade. These data provide insights into the nature of the biological response to viral infection of the upper respiratory tract, and the proteins that are dysregulated by viral infection form the basis of signature that accurately classifies the infected state. Verification of this signature using targeted mass spectrometry in independent cohorts of subjects challenged with influenza or rhinovirus demonstrates that it performs with high accuracy (0.8623 AUROC, 75% TPR, 97.46% TNR). With further development as a clinical diagnostic, this signature may have utility in rapid screening for emerging infections, avoidance of inappropriate antibacterial therapy, and more rapid implementation of appropriate therapeutic and public health strategies.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Acute respiratory viral (ARV) infections are among the most common reasons for patient visits in primary and acute care settings (Hong et al., 2004; Johnstone et al., 2008). Many viruses cause such acute respiratory illness including human rhinovirus

(HRV), respiratory syncytial virus (RSV) and influenza. These viruses can be associated with a range of clinical severity from asymptomatic to mild, self-limited illness to respiratory failure and death. Influenza alone causes 25 to 50 million infections annually in the USA, resulting in several hundred thousand hospitalizations and 20–40,000 deaths (Thompson et al., 2010).

Despite viral etiologies driving most cases of acute respiratory infection, definitive diagnostic tools for these syndromes are lacking. Even highly sensitive pathogen-specific tests such as PCR are dependent upon proper sampling technique and inclusion of virus-type-specific reagents and processing methods. Moreover, detection of a specific microbe in a clinical sample does not necessarily indicate the cause of the acute clinical syndrome. For example, it has been reported that HRV has been detected in up to 44% of asymptomatic individuals (Byington et al., 2015; Johnston et al., 1993). Therefore, better tools that help providers define the etiology

\* Correspondence to: Geoffrey S. Ginsburg, Center for Applied Genomics and Precision Medicine, Department of Medicine, Duke University, 101 Science Drive, Rm 2111, CIEMAS Building, Durham, NC 27708, USA.

\*\* Correspondence to: Christopher W. Woods, Center for Applied Genomics and Precision Medicine, Department of Medicine, Duke University, 310 Trent Drive, Trent Building – Room 205, Durham, NC 27708, USA.

E-mail addresses: [chris.woods@duke.edu](mailto:chris.woods@duke.edu) (C.W. Woods), [geoffrey.ginsburg@duke.edu](mailto:geoffrey.ginsburg@duke.edu) (G.S. Ginsburg).

of a suspected infectious syndrome in a safe, rapid, accurate, and cost-effective manner are of paramount importance for both individual and public health as recently noted by the Presidential Advisory Council on Combating Antibiotic-Resistant Bacteria (House, 2014), and others (O'Neill, 2015; Organization, 2015). A complementary diagnostic strategy to pathogen detection could focus on utilizing the varied (but pathogen-class specific) host-response to infection (Ramilo and Mejias, 2009; Zaas et al., 2014). This approach discriminates between infection and colonization. It is pathogen-agnostic and therefore circumvents another limitation of pathogen detection assays, which due to technical limitations are only capable of detecting a limited subset of microorganisms. Furthermore, categorizing infection based on host response provides additional insights into the mechanisms of infection and disease response, and may offer new targets, pathways, or strategies for therapeutic intervention.

We recently identified gene expression patterns in peripheral whole blood capable of differentiating (Zaas et al., 2009; Woods et al., 2013; McClain et al., 2016; Tsalik et al., 2016; Huang et al., 2011) individuals with symptomatic infection due to influenza H3N2, HRV, or RSV from uninfected individuals with >90% accuracy. Moreover, this ARV signature was validated in an independent population of patients with influenza A infection, demonstrating an ability to distinguish from bacterial respiratory infections (93% accuracy) and healthy controls (100% accuracy) (Zaas et al., 2009). Thus, host derived biomarkers are capable of making these types of distinction. However, considering the technical challenges inherent in developing peripheral blood host gene expression classifiers as a diagnostic tool – including semi-invasive venipuncture, RNA instability, processing complexity, relatively high cost of RNA profiling, and time to result – we sought to extend this host response paradigm for ARV diagnosis to an alternative and potentially more suitable sample matrix and analyte class.

Upon contact with the respiratory epithelium, respiratory viruses incite activation of type I interferons (IFNs) and pro-inflammatory cytokines, orchestrate proliferation of inflammatory cells and the innate immune response, and regulate induction of adaptive immunity (Yoneyama and Fujita, 2010; Koyama et al., 2008; Bhoj et al., 2008). Based on the prominent role of the nasopharyngeal epithelium in mediating ARV infections, we hypothesized that nasopharyngeal lavage (NPL) would reflect the in situ host response and serve as a potential target for diagnostic development. Furthermore, the NPL protein fraction represents an accessible sample matrix, providing a highly tractable diagnostic analyte class. Multiple reaction monitoring (MRM), a quantitative mass spectrometry (MS) platform for facile development of multiplexed, quantitative assays for measuring specific protein levels in biologic fluids and is routinely used for biomarker verification in clinical cohorts (Kiyonami et al., 2011; Gerszten et al., 2010; Boja and Rodriguez, 2011). In addition to being customizable for nearly any target protein, MRM assays provide a more specific quantitation of individual proteins and protein isoforms by targeting multiple unique peptides per protein target. Combined with internal stable-isotope labeled (SIL) peptide standards, these assays match or exceed the quantitative precision of ELISA assays with low femtomole limits of quantitation and analytical precision coefficient of variation < 10% across clinically sized cohorts (Addona et al., 2009; Aebersold et al., 2013).

Using human viral challenge cohorts for influenza A/H3N2 and HRV, we have discovered and independently verified multiple NPL protein biomarkers capable of classifying human influenza A and HRV infection from uninfected individuals. This work reinforces the important concept that host response to infection, particularly in the NPL proteome, serves as a potential basis for diagnostic testing. It also sheds light on the complex interactions of host and pathogen in two of the most common infectious diseases in humans.

## 2. Materials and Methods

### 2.1. Study Design

All pathogen exposures were approved by the relevant Institutional Review Boards and conducted according to the Declaration of Helsinki. All volunteers provided informed consent. The objective of these experimental challenge studies was to generate clinico-molecular classifiers of ARV infection through the development and characterization of high-density sample and data sets across the course of respiratory virus exposure, infection, and resolution. A description of methods used in each challenge study can be found in Supplementary materials and have been described previously (Liu et al., 2016; McClain et al., 2016; Woods et al., 2013; Zaas et al., 2009). Briefly, healthy volunteers underwent extensive pre-enrollment health screening and were excluded for positive baseline antibody titers of the strain of virus used in each challenge (Influenza A H3N2 A/Wisconsin/67/2005 or HRV serotype 39). Following 24–48 h in quarantine, we instilled viral inoculum into bilateral nares of subjects using standard methods. At predetermined intervals, biological samples and clinical and symptom data were collected. NPL sampling was performed daily for each participant. The H3N2 #2 cohort included an early (36 h post-inoculation) oseltamivir treatment arm, while HRV #2 included a blinded “sham” inoculation (saline only) control group. NPL analyses included baseline and time T samples from all individuals in each challenge study with complete and unambiguous symptomatology and microbiology data, and available NPL samples. Sample phenotype labels were blinded for MRM analysis, but were assayed in a manner to ensure that samples from an individual, and from within a challenge study, would be processed in same batch and assayed in close temporal proximity, to minimize batch effects between distinct phenotypes.

### 2.2. Case Definitions

Self-reported symptoms were recorded at predetermined intervals prior to inoculation and at least twice daily throughout the time-course of infection and resolution as reported previously (Zaas et al., 2009; Jackson et al., 1958) and described in Supplementary materials. This modified Jackson score requires subjects to rank 8 symptoms of upper respiratory infection (headache, sore throat, rhinorrhea, rhinitis, sneezing, coughing, myalgia, malaise) on a standardized scale of 0 (no symptoms) to 3 (high symptoms). Symptom scores were tabulated for each study participant to assign symptom status as symptomatic or asymptomatic (Supplementary Table S2A). For each symptomatic subject, time T was identified as time of maximal symptoms. The average time T was then defined for that cohort, which served as the time chosen for asymptomatic subjects (Table 1). Participants were tested for virus shedding based on quantitative culture assays as described previously

**Table 1**

Description of experimental ARV challenge cohorts. Four experimental HRV challenge cohorts (two influenza A/H3N2 and two HRV) are described, including adjudicated phenotype summary data for each. Individuals with discordant<sup>a</sup> symptom and shedding labels, i.e. symptomatic non-shedders, or asymptomatic shedders, are shown. Sx = symptomatic; Asx = asymptomatic; mean Sx and Asx time T represents the average time of maximal self-reported symptoms among subject included in NPL analysis.

Cohort	H3N2 #1	H3N2 #2	HRV #1	HRV #2
Subjects included in challenge	17	21	20	30
Subjects included in NPL analysis	15	21	20	24
Symptomatic	9	13	12	15
Shedding positive	10	10	12	15
Number discordant <sup>a</sup>	1	7	4	9
Mean Sx time T (in hrs)	69	68	66	76
Corresponding Asx time T (in hrs)	69.5	71	72	72

<sup>a</sup> Individuals with discordant symptom and shedding labels, i.e. symptomatic non-shedders, or asymptomatic shedders. NPL = Nasopharyngeal lavage. Sx = Symptomatic. Asx = Asymptomatic. Time T represents the time of maximal symptoms.

(Zaas et al., 2009) and outlined in Supplementary Table S2B. For the purpose of the current analysis, we differentiated between “symptomatic” and “shedding”. A symptomatic subject shedding virus was labeled as “infected”. Asymptomatic non-shedders were “uninfected”. Discordance between symptom and shedding status were reconciled by measuring a previously published peripheral blood gene expression score, as described in Supplementary materials. This tiebreaker was a gene expression analysis (GEA) representing the host peripheral blood response to viral infection (Supplementary Table S2C) (unpublished data; N. Arzouni, T. Burke, M. McClain, A. Hero). GEA was only applied as a tiebreaker and not to subjects where the symptom and shedding status agreed.

### 2.3. Sample Collections

Sample collections for subjects in H3N2 cohort #1 and HRV cohort #1 have been described previously (Zaas et al., 2009). Samples for H3N2 cohort #2 and HRV cohort #2 were collected, processed, and analyzed similarly. Briefly, biological samples were collected prior to inoculation (baseline) and at predetermined intervals throughout the course of infection and resolution. Nasopharyngeal lavage procedures were performed using sterile 0.9% saline solution (5 ml into each nares), as described in Supplementary Materials, prior to inoculation and at 24-hour intervals post-inoculation and stored in aliquots at  $-80^{\circ}\text{C}$ . This study focused on comparison of pre-inoculation (baseline) samples with samples taken at or shortly following time T.

### 2.4. Unbiased 2D-LC-MS/MS

NPL samples were processed for proteomic analyses as described in Supplementary materials. For the pooled 2D-LC-MS/MS discovery analysis, four sample pools were created representing the four groups in H3N2 #1 challenge - Uninf-BL ( $n = 4$ ), Uninf-T ( $n = 6$ ), Inf-BL ( $n = 6$ ), and Inf-T ( $n = 8$ ) - with equal protein mass (2  $\mu\text{g}$ ) from each participant. Normalized pooled samples (2D-LC-MS/MS) and individual participant samples (MRM) were reduced, alkylated, and digested with trypsin as described in Supplementary materials. Prior to analysis, all samples were spiked with ADH1\_YEAST digest (Massprep standard, Waters Corporation) as an internal technical standard. Unbiased proteomic discovery analysis was performed using Nano-scale Capillary UPLC-MS/MS as described in Supplementary Materials. Briefly, quantitative 10-fraction 2D-LC-MS/MS was performed on duplicate injections for each sample pool, providing accurate mass and intensity (abundance) acquisitions with qualitative identification of the resulting peptide fragments via searching against a SwissProt\_Human ([www.uniprot.org](http://www.uniprot.org)) database that also contained a reversed-sequence “decoy” database for false positive rate determination. Analytical reproducibility of the label-free 2D-LC-MS/MS method was assessed by calculating the variation in measured abundance of the spiked ADH1\_YEAST standard, demonstrating a coefficient of variation of 10.6% across all eight injections. For quantitative processing, peptide quantities across all ten 2D-LC fractions were summed and the dataset was intensity-scaled to the robust mean (excluding highest and lowest 10% of detected features) across all quantitative acquisitions. The final quantitative dataset for NPL was based on 3285 peptides and contained 438 unique proteins.

### 2.5. Targeted MRM

Following the selection of 25 candidate protein targets from the unbiased discovery data, all individual samples were subjected to a targeted MRM assay as detailed in Supplementary materials. MRM assay development and transition selection was performed within the open-access Skyline (MacCoss Laboratory, Univ of Washington) software. Initially, up to five unique peptides were selected from each candidate protein based on average precursor ion intensity. Five transitions for each precursor ion were selected based on 1) qualitative DDA

discovery MS/MS data, 2) other discovery datasets for which the same peptide sequence was identified or 3) from the PeptideAtlas ([www.PeptideAtlas.org](http://www.PeptideAtlas.org)) public repository. Following deployment of the initial MRM assay on a healthy human control NPL pool, the MRM method was optimized to choose three transitions from the two most robust peptides per protein.

Custom SIL peptides were ordered for each candidate peptide to be assayed, and were spiked into each individual digested NPL participant sample at one of four ratios relative to endogenous peptide as described in Supplementary Materials. Each of the four patient cohorts (two H3N2 and two HRV) was run as individual run blocks and samples within a cohort were randomized in injection order across the cohort. Single MRM assays were performed on spiked NPL samples with a target quantity of up to 1  $\mu\text{g}$  on-column. Four samples (3 from H3N2 #1 and 1 from H3N2 #2) were prepared and MRM quantification was attempted, but did not have sufficient protein material to generate robust quantitative data and therefore were excluded from subsequent analyses. To assess analytical variation, an equal portion of each patient's SIL spiked NPL sample was used to generate a QC pool, which was run approximately every 12 h across the entire cohort. In addition, all samples were spiked with five SIL peptides from yeast\_ADH as an internal technical control. MRM assay reproducibility metrics are shown in Supplementary Fig. S1.

### 2.6. Data Processing

We examined 51 human peptide analytes (plus 5 yeast ADH) from 26 different human proteins. Eighty subjects from the four viral challenge studies were assayed although four subjects had insufficient NPL material at one time point resulting in 156 samples. Four peptides with  $>2$  missing values were excluded from analysis. Missing values and zeroes (11 sample-analytes) were imputed with half the observed minimum value of a given peptide, and expression levels from the remaining 47 peptides were log transformed and carried forward for further analysis. Simple batch (study) correction was performed by removing study-wise mean values from each peptide. Final analysis included 47 human peptides measured in 156 samples.

### 2.7. Statistical Analysis

Univariate testing was performed using two-sided *t*-tests with Benjamini-Hochberg FDR corrected *p*-values. For classification we used sparse logistic regression, in particular a Least Absolute Shrinkage and Selection Operator (LASSO) generalized linear model with binomial likelihood (Friedman et al., 2010). Performance metrics and model parameters were obtained via nested leave-one-out cross-validation (LOOCV). As classification performance metrics we consider area under receiver operating curve (AUROC) (Fawcett, 2006), true positive rate (TPR), and true negative rate (TNR). A 17 protein (30 peptide) relaxed classifier ( $\alpha = 0.1$  rather than 1.0) was subjected to pathway association analysis using DAVID 6.7 Functional Annotation Tool (Huang et al., 2009) using UniProt accession identifiers and human background.

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request. Computational scripts were written in MATLAB using the GLMnet toolbox ([web.stanford.edu/~hastie/glmnet\\_matlab/](http://web.stanford.edu/~hastie/glmnet_matlab/)) and can be accessed at [https://bitbucket.org/rhenao/npl\\_ebm](https://bitbucket.org/rhenao/npl_ebm).

## 3. Results

### 3.1. Influenza A/H3N2 and Human Rhinovirus Challenge Cohorts

Four independent viral challenge studies were conducted (Table 1) - two challenges with influenza A/H3N2 and two with HRV, and were described previously (McClain et al., 2016; Woods et al., 2013; Zaas et al., 2009). Clinical and self-reported symptom data, along with



corresponding samples, were collected multiple times per day across the course of exposure, infection, and resolution. For each challenge participant, standardized algorithms were applied to assign phenotype labels describing symptomatic, shedding, and “infected” status (Supplementary Tables S2A, S2B, S2C, respectively) across the time-course. Since symptom and shedding status were not always congruent, we required that both be present to define a patient as “infected”. When both were absent, that subject was labeled “uninfected”. When discordant, we applied a tiebreaker based on gene expression analysis (GEA) representing the host peripheral blood response to viral infection (Supplementary Table S2C) (unpublished data; N. Arzouni, T. Burke, M. McClain, A. Hero). Of the 80 challenge participants included in this analysis, just over half (42/80) of the participants were adjudicated as becoming infected (H3N2 #1: 9 of 15 infected; H3N2 #2: 10 of 21; HRV #1: 10 of 20; and HRV #2: 13 of 24).

### 3.2. Proteomics Discovery and Verification Strategy

An overview of the proteomic discovery and candidate biomarker verification strategy is depicted in Fig. 1. Despite being a readily accessible matrix, one technical challenge of studying NPL is the variability in protein yield from subject to subject. To overcome this limitation, we utilized a two-phase biomarker discovery strategy: The first phase used pooled NPL samples from a single H3N2 challenge for in-depth MS-based discovery proteomics to assist in identification and prioritization of candidate biomarkers. In the second phase, targeted assays were developed for these candidates using the more sensitive and quantitatively reproducible MRM approach. Verification of the utility of the markers was then performed in the original and three independent challenge cohorts using targeted MRM quantitation.

### 3.3. Identification of Candidate Protein Biomarkers Using Pooled NPL from H3N2 Challenge Study

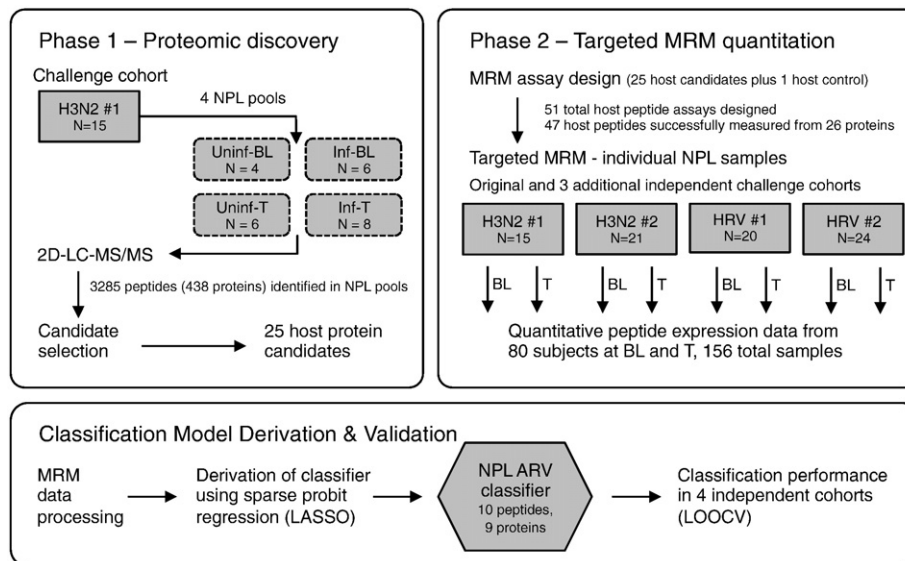
Discovery proteomics was performed on the H3N2 #1 challenge cohort using open-platform, 2-dimensional liquid chromatography, tandem MS (2D-LC-MS/MS) analysis of four sets of pooled NPL samples (uninfected and infected individuals at baseline and time of maximal symptoms, time T) using equal protein mass per participant sample.

Across the four unique NPL sample pools, a total of 3285 peptides corresponding to 438 unique proteins were identified at a 1.0% peptide-level false discovery rate (FDR).

We next investigated the variable expression of NPL proteins in infected and uninfected viral challenge pooled samples. Three criteria were used to prioritize suitable peptides for subsequent MRM quantification from the entire collection of 438 identified NPL proteins (Supplementary Table S3). First, we sought a minimum two-fold change in expression between baseline and time T in the infected pool in at least 2 peptides per protein; this criterion was used to classify proteins that would increase or decrease as a function of infection within the same individuals over time. Levels of 107 proteins increased at least two-fold, while 61 proteins decreased at least two-fold from baseline to time T. The second criterion was a greater than two-fold difference between infected and time-matched uninfected subjects; this criteria was to enable specificity of the proteins by comparing test and control subjects at a time where symptoms are present. This included 36 proteins with higher expression and 33 proteins with lower expression. The third criterion excluded proteins that might reflect general nasal trauma stemming from repeated collections. To address this, we calculated the expression change between the time of maximal symptoms for infected subjects (Inf-T) relative to similar times for uninfected subjects (Uninf-T), and prioritized candidates with unique response to infection at time T. Additionally, four proteins met 2 of 3 criteria and also had a reported association with infection (IC1) (Zaas et al., 2009) and inflammation (APOA1, APOA2, and APOA4) (Pirillo et al., 2015) and were included in the verification phase. Based on these criteria, 25 proteins were selected for subsequent MRM assay development - 13 had increased and 12 had decreased expression in infected participants.

### 3.4. MRM Quantification of Candidate Biomarkers in Individual NPL Samples

MRM is a quantitative LC-MS/MS method utilizing synthetic, SIL peptides as internal standards, and provides absolute specificity for the target analyte and relative abundance measurements (Kiyonami et al., 2011; Addona et al., 2009). A total of 51 unique peptide MRM assays were designed to target two unique peptides for each of the 25 prioritized candidate biomarker proteins, plus human serum albumin and

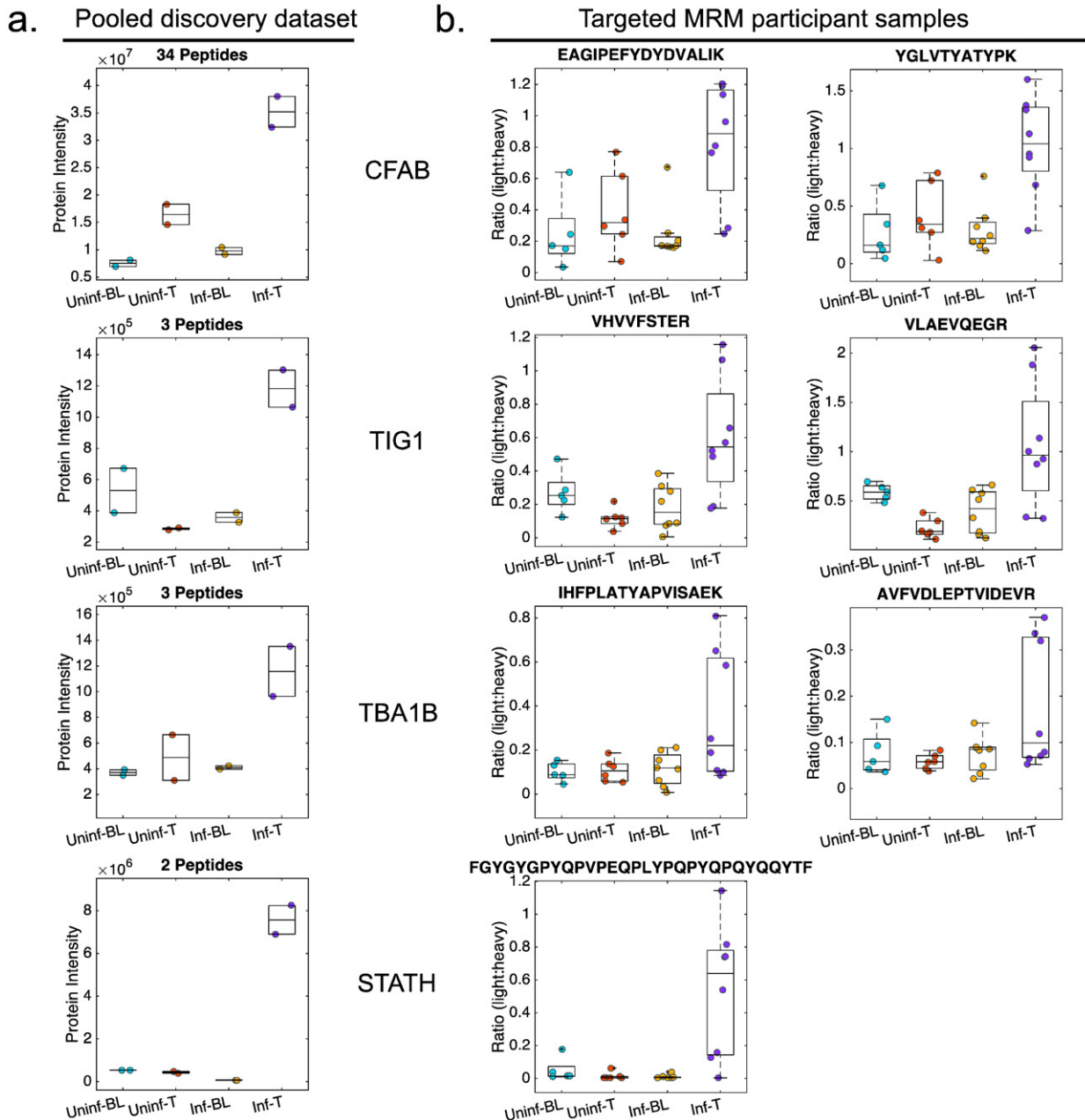


**Fig. 1.** Study design and experimental workflow. A two-phased strategy was employed to identify and characterize candidate protein biomarkers of ARV infection from NPL samples collected from participants in four experimental ARV challenge cohorts. For phase 1 discovery analysis, four NPL pools were prepared from H3N2 #1 cohort and analyzed using unbiased 2D-LC-MS/MS. The numbers of subject (N) with samples included in each pool are shown (Uninf = uninfected individuals; Inf = infected individuals; BL = baseline; T = time of maximal symptoms). For phase 2, the original and three additional independent challenge cohorts were assayed by targeted MRM. Quantitative peptide expression data from 80 individuals and 156 total samples were used in the derivation of an NPL ARV classifier, and classification performance was assessed in independent challenge cohorts using LOOCV.

yeast alcohol dehydrogenase as a spiked exogenous control. Two proteins yielded only a single peptide that was suitable for an MRM assay – Statherin (only one viable SIL peptide could be synthesized) and Filaggrin (one of two SIL peptides was insoluble). A third protein, Calcyphosin, yielded a third suitable peptide assay from the publicly available PeptideAtlas database.

Following development, MRM assays were performed on 156 individual NPL samples from 80 subjects across four viral challenge study cohorts. This included the original influenza H3N2 cohort (H3N2 #1), a second H3N2 cohort (H3N2 #2), and two HRV challenge cohorts (HRV #1 and #2). From the original set of 51 human peptide assays designed, 47 peptides representing 26 proteins were successfully measured.

For the H3N2 #1 cohort, we observed that both the direction and magnitude of change in protein expression was consistent between the initial pooled discovery measurements (Fig. 2A) and the targeted MRM peptide measurements (Fig. 2B). This was expected, since the pooled sample is the biological average of all samples included in the pool. For example, retinoic acid receptor responder protein TIG1 was measured across three unique peptides in the pooled discovery data set, with an average fold change from uninfected to infected time T of 3.3-fold. Targeted MRM assays developed for two TIG1 peptides demonstrated average increased expression of 2.3- and 2.0-fold when assayed in individual NPL samples (Table 2). Similarly, two complement factor B (CFAB) peptides displayed 2.0- and 2.0-fold increased expression when assayed in individual H3N2 #1 samples by MRM, compared to a 3.6-fold



**Fig. 2.** Expression levels of candidate protein biomarkers as measured by unbiased LC/LC-MS/MS and targeted MRM. Four representative proteins (CFAB, TIG1, TBA1B, STATH) in H3N2 #1 challenge cohort were measured by (A) LC/LC-MS/MS unbiased discovery methods in sample pools (duplicate measures of multi-peptide meta-protein is shown), and (B) targeted MRM analysis of unique peptides (peptide sequence is shown above each plot) from individual participant samples from the same H3N2 #1 cohort. MRM data are expressed as relative ratio (endogenous “light” peptide to spiked SIL “heavy” peptide). Box-and-whisker plots for each group indicate intra-quartile range (box) with upper and lower (whisker) values representing 99% data coverage. Colored circles represent Uninf-BL (blue), Uninf-T (red), Inf-BL (yellow), and Inf-T (purple).

**Table 2**

Candidate biomarker protein relative expression ratios. Rows represents the 26 host NPL proteins for which MRM assays were developed. UniProt gene symbol and protein names are shown for each candidate protein, with number of peptides measured for both unbiased pool and targeted MRM analyses. Unbiased pool ratios are shown for selection criteria 1, 2, and 3, as described in Supplementary Table S3. Fold expression changes from BL to T for infected individuals, as measured by peptide-specific MRM assay, across all four ARV challenge studies are shown with two-sided *t*-test Benjamini-Hochberg FDR-adjusted *p*-values (in parentheses) for each peptide, respectively.

UniProt gene symbol/protein name	Unbiased pool peptides	Unbiased (pool) H3N2 #1			Targeted MRM (individual)			Notes
		Criterion <sup>a</sup>			Targeted MRM peptides	H3N2 #1/2 and HRV #1/2 Fold change [T/BL] <sub>Inf</sub> (FDR) <sup>b</sup>		
		1	2	3				
CFAB Complement factor B	34	3.6	2.1	1.6	2	2.0, 2.0 (1.8e−4, 2.2e−4)		
IC1 Plasma protease C1 inhibitor	8	6.4	1.8	1.0	2	1.3, 1.8 (0.089, 3.9e−4)		c
A1AT Alpha-1-antitrypsin	100	5.1	2.1	1.3	2	1.7, 1.7 (7.8e−4, 5.0e−4)		
A1AG1 Alpha-1-acid glycoprotein 1	19	4.1	2.6	1.1	2	1.8, 1.8 (3.9e−4, 3.9e−3)		
ALB Albumin	805	2.4	1.4	1.0	2	1.5, 1.6 (7.8e−4, 7.8e−4)		d
TIG1 Retinoic acid receptor responder prot 1	3	3.3	4.2	5.7	2	2.3, 2.0 (6.2e−4)		
LCN15 Lipocalin-15	6	−2.5	−3.7	−2.9	2	−3.2, −3.3 (1.1e−3, 1.1e−3)		
TBA1B Tubulin alpha-1B chain	3	2.8	2.6	2.2	2	1.8, 1.9 (1.0e−3, 8.6e−4)		
CYTA Cystatin-A	5	−2.6	−2.8	−2.4	2	−3.5, −2.8 (8.6e−4, 6.7e−4)		
APOA1 Apolipoprotein A-I	57	5.1	−1.3	−1.8	2	2.1 (1.2e−3)		c
ANGT Angiotensinogen	6	4.9	2.3	1.5	2	1.4, 1.7 (0.018, 8.4e−3)		
SPR1A Cornifin-A	2	−2.7	−2.0	−1.2	2	−2.2, −1.8 (9.0e−3, 3.9e−3)		
CASPE Caspase-14	10	−3.8	−2.1	−2.2	2	−3.1, −2.7 (6.7e−4, 1.0e−3)		
APOA4 Apolipoprotein A-IV	7	11.4	1.1	1.0	2	2.2, 2.0 (1.3e−3, 1.4e−3)		c
AMY1 Alpha-amylase 1	7	−2.4	−2.9	−1.7	2	−2.1 (2.9e−3)		
VMO1 Vitelline membrane outer layer prot 1	3	−2.0	−2.9	−1.4	2	−2.2, −2.6 (0.016, 8.7e−3)		
ANXA2 Annexin A2	8	−2.0	−2.0	−1.6	2	−1.5, −1.6 (5.7e−3, 4.7e−3)		
CAYP1 Calcyphosin	4	7.5	2.5	2.1	3	2.6, 2.7, 2.4 (2.9e−3, 1.6e−3, 2.4e−3)		
GGCT Gamma-glutamylcyclotransferase	2	−4.6	−2.0	−1.7	2	−1.7, −2.1 (0.029, 9.9e−4)		
DCD Dermcidin	3	−7.4	−2.5	−4.3	2	−2.3, −2.2 (6.4e−3, 6.4e−3)		
SPRR3 Small proline-rich protein 3	2	2.4	2.1	1.2	2	−1.6, −1.9 (0.023, 7.8e−3)		
APOA2 Apolipoprotein A-I	9	8.7	1.3	1.3	2	1.7, 1.5 (0.016, 0.07)		c
STATH Statherin	2	112.6	17.5	140.0	1	5.4 (1.0e−3)		
PROL4 Proline-rich protein 4	66	−2.6	−2.0	−2.3	2	−1.1 (0.802)		
GSTA2 Glutathione S-transferase A2	3	9.1	4.2	1.6	2	1.0 (0.902)		
FILA Filaggrin	2	−7.3	−3.8	−3.9	1	−1.6 (0.061)		

<sup>a</sup> Fold change in pooled analysis for each of 3 prioritization criteria.

<sup>b</sup> Average fold change per peptide across all cohorts (*p*-values adjusted for false discovery rate).

<sup>c</sup> Included due to potential biological significance.

<sup>d</sup> Included as potential endogenous control.

increase for the summed 24-peptide CFAB protein composite assayed in the pooled H3N2 #1 discovery dataset. Also as expected, the pooled discovery dataset shows much more narrow standard deviation (Fig. 2A), as this deviation is due only to technical variability, whereas the variability in the targeted MRM experiment of all individuals (Fig. 2B) is able to discern the actual biological variation. These combined results provided confidence that the pooled-sample 2D-LC-MS/MS phase 1 discovery strategy gives a good measure of the average protein expression level in situations where the matrix of interest is extremely sample-limited, and that the phase 2 targeted MRM assays reproduce protein expression changes candidates selected using the pooled data, while also providing biological variance with higher sensitivity.

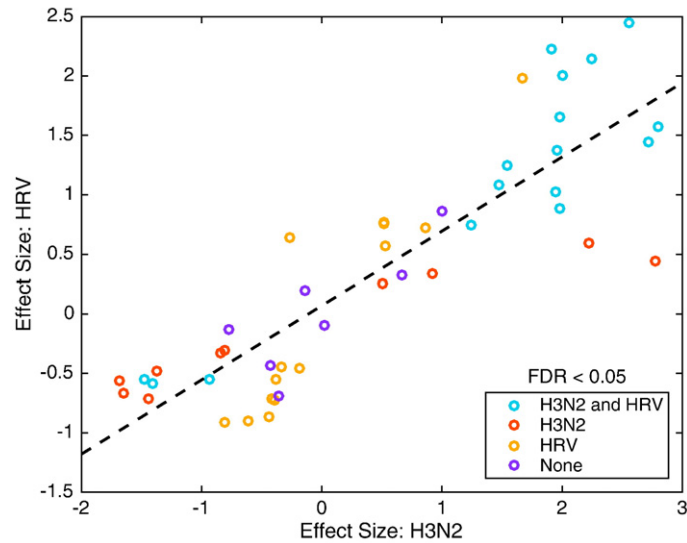
### 3.5. Independent Verification in Additional Viral Challenge Cohorts

We then sought to independently verify candidate peptide expression changes in individual participant samples from a second H3N2 challenge cohort, as well as two additional challenge studies using a second common ARV, human rhinovirus. Individual participant NPL samples from the 3 additional cohorts were processed and measured using the same panel of experimental and control peptide MRM assays derived from the H3N2 #1 cohort. In a combined analysis across the four independent ARV cohorts, we found that median sample-wise peptide expression intensity was relatively constant across all samples regardless of infection status, though 26 of 47 peptide analytes measured to be differentially expressed (Benjamini-Hochberg FDR < 0.05) at the time of maximal symptoms in the infected H3N2 subjects. Likewise, 30 peptide analytes were differentially expressed (Benjamini-Hochberg FDR < 0.05) in the HRV subjects at time T, with 16 peptides overlapping

between the two H3N2 and two HRV cohorts. The direction and magnitude of change for all 40 peptide analytes (10 H3N2 only, 14 HRV only, 16 in both) that were significantly differentially expressed in either the H3N2 or HRV cohorts are highly correlated ( $r = 0.871$ ) between viral groups (Fig. 3), indicating that these analytes are likely not virus type-specific or useful in differentiating between H3N2 and HRV. This finding provided a basis for investigating the utility of these protein assays in a multi-analyte classification model of ARV infection.

### 3.6. Classification Model of ARV Infection

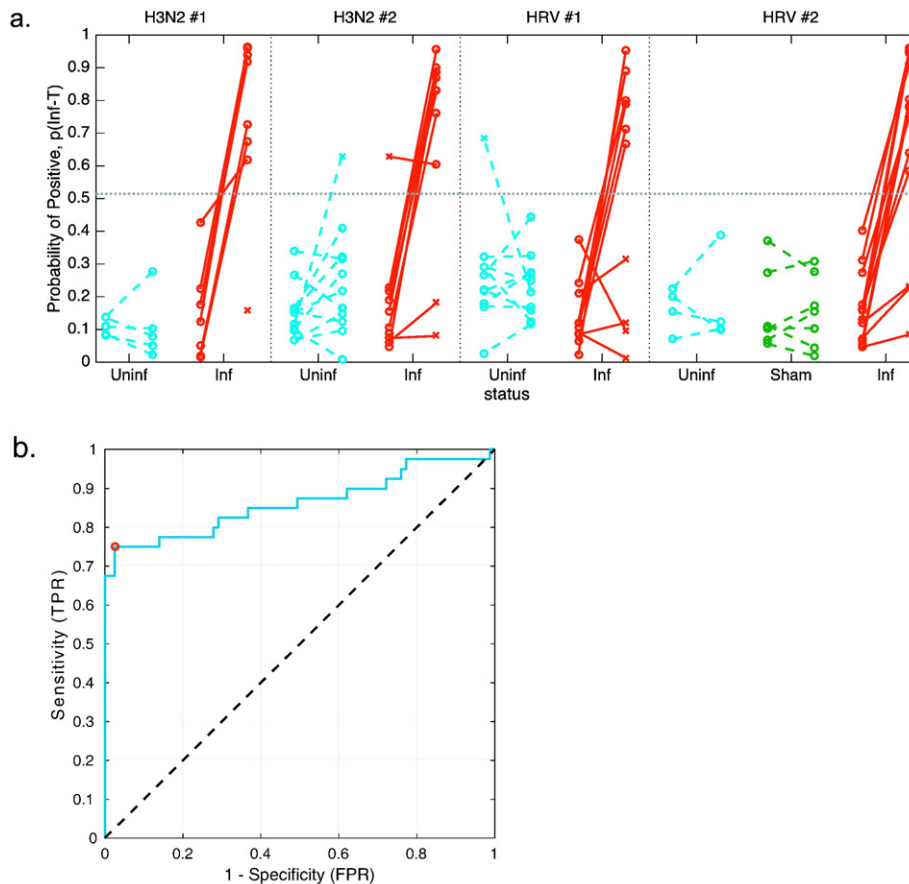
To build an NPL protein classifier that distinguishes infected from uninfected subjects we utilized LASSO sparse logistic regression (Friedman et al., 2010) to build a list of potential logistic regression models. The classifier was trained to discriminate samples from infected individuals at time T (Inf-T) from paired baseline measures (Inf-BL) and from uninfected individuals at time T (Uninf-T). Fig. 4A shows the performance of a NPL classification model that selected 10 peptides (9 unique proteins), in the four independent challenge cohorts, including discordant individuals adjudicated using GEA status as a tiebreaker. For this analysis we ignored the paired nature of samples in the training and classification tasks in order to better estimate the accuracies in a situation in which baseline samples are not available. Overall classifier performance (Fig. 4B) in LOOCV yielded an AUROC of 0.8623 (95% CI: 0.7538–0.9315, bootstrapped 10 K samples) with a 75% TPR and 97.46% TNR. Baseline samples from the asymptomatic cohort were withheld from model training since this was not a phenotype the model was trying to identify. However, they represent a cohort of asymptomatic individuals available for validation. Applying the model



**Fig. 3.** Correlation of the peptide expression changes for all four H3N2 and HRV cohorts ( $r = 0.871$ ). Graph shows peptides with significant expression changes (Benjamini-Hochberg FDR < 0.05) in H3N2 (red circle, 10 peptides), HRV (yellow circle, 14 peptides), both H3N2 and HRV (blue circle, 16 peptides), or neither cohort (purple circle) as measured by MRM.

to these 37 samples revealed only one misclassification error (2.7%). LOOCV confusion matrix and identities of the 10 peptide classifier with weights are shown in Table 3A and B, respectively. We also

performed analyses separately on infection status of individuals independently of the GEA status (i.e., including only symptomatic shedders and asymptomatic shedders) with no surprisingly increased



**Fig. 4.** Performance of a 10-peptide classification model in independent ARV cohorts. The model was fit to discriminate Inf-T from Inf-BL and Uninf-T. (A) Probability of positive is plotted (x-axis) for individual participant samples from 4 cohorts, with threshold for positive classification set at >0.5. Lines connect paired samples from individuals at baseline (left) and time T (right). Solid red line = infected; dashed blue line = uninfected; dashed green line = uninfected sham inoculation; open circles = classifier agrees with phenotype label; X = classifier disagrees with phenotype label. (B) Receiver operating curve for LOOCV with the 10 peptide (9 unique proteins) classifier. The optimal threshold on the curve (open red circle) produces 0.8623 AUROC, 75% TPR, 97.46% TNR, with the confusion matrix and peptide weights shown in Table 3A and B, respectively.



**Table 3**

10-peptide classification model performance. (A) Classifier performance on individual samples from all four ARV cohorts as represented by LOOCV confusion matrix, with (B) identity and contribution (weight) for 10 peptides, with amino acid sequence in one-letter notation. Average peptide length is 11.8 amino acids, with range between 8 and 30 residues. Negative weight values indicate down-regulation upon infection.

A: 10-peptide classification model performance on ARV challenge cohorts		
Actual/predicted	Others <sup>a</sup>	Inf at T
Others <sup>a</sup>	77	2
Inf at T	10	30

<sup>a</sup>Inf-BL and Uninf-T; Confusion matrix based on LOOCV using 10 peptide (9 protein) classification model.

B: Identity and contribution of peptides in classification model		
Weight <sup>a</sup>	Gene symbol	Peptide sequence
0.927	A1AG1	EQLGEFYEALDCLR
0.850	TIG1	VLAEVQEGR
−0.266	ANXA2	TNQLQEINR
−0.260	LCN15	VPALGYLDVR
0.249	ANGT	ADSQAQLLLSTVVGFTAPGLHLK
0.228	CFAB	EAGLPEFYDYDVALIK
0.198	TBA1B	IHFPLATYAPVISA EK
0.154	CAYP1	EAVIAAAFAK
0.126	STAT	FGYGYGPYQVPPEQPLYQPYPQYQYQYTF
0.081	CFAB	YGLVITYATYPK

<sup>a</sup>Peptide weight contribution to classification model (negative value indicates down-regulated upon infection).

classification performance, 0.8821 LOOCV-AUC (vs. 0.8623), and on all individuals using shedding as outcome (ignoring symptoms) with 0.8610 LOOCV-AUC.

### 3.7. Symptomatic but Uninfected Sham Controls

The experimental design of the HRV #2 challenge included a “sham” inoculation control group (Fig. 4A, panel HRV #2 in green), with several participants receiving intranasal inoculation of saline without HRV. Four of the seven sham controls reported symptoms sufficient to be classified as symptomatic, despite receiving no HRV inoculum and showing no microbiological evidence of infection. All four subjects were classified as negative by NPL MRM analysis, consistent with absence of HRV shedding and no observed GEA host response in the blood mRNA ARV infection pathway (Huang et al., 2011; Woods et al., 2013; McClain et al., 2016) for these subjects.

### 3.8. Pathway Analysis

Since the classifier tested in Fig. 4 includes only 9 unique proteins, we repeated the analysis using a more relaxed variable selection parameter to better describe the biological pathways involved (LASSO regularization parameter,  $\alpha = 0.1$  rather than 1.0). The new classifier has nearly identical performance (0.8765 AUROC, 75% TPR, and 97.46 TNR) but selected 30 of 47 peptides (17 unique proteins). Pathway association analysis was performed using DAVID 6.7 Functional Annotation Tool (Huang da et al., 2009) using the 17 proteins selected by this expanded classifier as being informative in differentiating infected from uninfected individuals (results in Supplementary Table S5). Significant annotation clusters were primarily driven by 5 proteins (A1AG1, CFAB, A1AT, IC1, and APOA4), and were associated with Gene Ontology (GO) biological process terms acute inflammatory response (GO:0002526;  $p = 1.6e-4$ ), innate immune response (GO:0045087;  $p = 9.9e-3$ ), inflammatory response (GO:0006954;  $p = 5.0e-3$ ), and defense response (GO:0006952;  $p = 3.9e-3$ ). Similar results were seen when expanding the query to include all 26 protein candidates used in the phase 2 MRM analyses.

## 4. Discussion

We conducted unbiased and targeted protein analyses on human NPL samples collected from four independent ARV challenge cohorts (two influenza A H3N2, two HRV) to define host protein expression patterns characteristic of response to viral respiratory infection. The results demonstrate that robust changes in secreted proteins occur in the NPL of infected individuals, and that a subset of proteins is capable of accurately classifying the infected state of the individual.

Despite changes to the assay approach and methods in transitioning from pooled phase 1 discovery with unbiased 2D-LC-MS/MS proteomics to individual measurements in phase 2 discovery with MRM, we observed good reproduction of both the direction and magnitude of peptide expression measurements between the methods. This provides confidence in the specificity and quantitative performance of both methods. Importantly, this approach enabled accurate selection of candidate biomarkers in a pooled phase 1 discovery where sample amounts were severely limited on individual samples, and precluded the ability to analyze all samples using unbiased 2D-LC-MS/MS. Though some biomarker candidate attrition was expected because the biological variance is not available from the pooled data, with appropriate strategic filtering of the candidates we were able to have a high success rate for statistical validation between phase 1 (pooled discovery) and phase 2 (individual MRM measurements).

Investigations by our group and others have shown differential expression of host genes at the RNA level in peripheral blood in response to ARV infections (Ramilo et al., 2007; Zaas et al., 2009; Woods et al., 2013; Zaas et al., 2013), with heavy representation of genes in the IFN-signaling canonical pathway and innate immune response signaling. Analysis of differentially expressed NPL proteins in infected individuals demonstrates involvement in several biological pathways critical to mounting a host defense to virus infection, including innate immune responses, acute inflammatory responses, and defense response pathways. The inclusion of three members of the complement system (CFAB, A1AT, and IC1) is particularly consistent with an innate immune response, as the complement system enhances the ability of antibodies and phagocytic cells to clear pathogens from the infected site (Ricklin et al., 2010). Despite the apparently related pathways involved, there does not appear to be extensive overlap between the differentially expressed nasal proteins identified in our H3N2 #1 challenge study, and the previous peripheral blood RNA signatures of ARV infection characterized in the same challenge cohort (Zaas et al., 2009). One gene product that is increased at both the NPL peptide and peripheral blood RNA level upon influenza infection (Cameron et al., 2008; Cameron et al., 2007; Zaas et al., 2009) is IC1 (*SERPINC1* gene). IC1 (also called C1-inhibitor) is a peptidase inhibitor belonging to the serpin superfamily and has an important role in innate immunity through modulation of the classical pathway of complement activation (Gaboriaud et al., 2004). As the complement system has the potential to be damaging to host tissues, complement control proteins must tightly regulate activation. IC1 binds to complement protein C1 to inhibit activation of the classical complement pathway, and thus its discovery fits well with our understanding of the biology of these diseases.

Notably absent from this NPL protein analysis are proinflammatory cytokine and chemokine gene products which have previously been shown to be strong contributors to the host response both in peripheral blood and near the site of ARV infection (Kimura et al., 2013). Oshansky and colleagues assayed nasal lavage samples from a cohort of healthy and naturally-infected influenza patients using a multiplex cytokine and chemokine assay panel, reporting correlation of inflammatory cytokines MCP-3 and IFN- $\alpha$ 2 with disease progression (Oshansky et al., 2014). An aptamer-based detection method was subsequently used to screen the same cohort and generate quantitative measures of over 1000 protein analytes from nasal lavage, showing differential expression of 162 proteins including cytokines associated with immune response to infection (Marion et al., 2016). We did not identify



inflammatory cytokines to be differentially expressed in our pooled NPL analysis. It is possible that cytokine proteins in NPL samples are expressed at levels below the detection limits of LC/MS-based methods, and that coupling targeted methods capable of detecting and quantifying cytokines directly may provide enhanced datasets for biomarker discovery.

The inclusion of “sham” infected participants in the HRV #2 cohort provided an important opportunity to assess the accuracy of standard clinical symptom assessments in the diagnosis of ARV infection. The clinical definition of symptomatic, even when applying standardized symptom scoring and algorithms, includes elements of subjectivity and thus may be imperfect in describing such infections. The finding that, despite the apparent presence of symptoms of infection, MRM analysis classified all seven individuals as uninfected suggests that this NPL biomarker assay has a lower false positive rate in diagnosing HRV infection than standard symptom status measures.

Experimental challenge studies provide an excellent model for studying ARV infection and illness, with pre-screened volunteers, known time of exposure, standardized pathogen exposure, and extensive sampling and data collection through the course of illness. However, experimental challenges do not perfectly replicate natural ARV infection and illness in humans. Volunteers in these studies are young and healthy, and represent a relatively homogenous population. In contrast, patients presenting to clinical care are demographically heterogeneous, have a variety of comorbidities, present at various times in the course of their illness, and contain a far greater breadth of potential pathogens beyond the two studied here. As such, additional validation should be performed in a more diverse population, such as patients presenting to clinical care with community-onset disease. While the generalizability of this study's findings will require additional validation, the high TNR of 97.46% suggests this assay may provide value in ruling out ARV infection.

Categorizing infection based upon host response represents an emerging strategy with great potential for complementing current pathogen-based diagnostics, as well as providing additional insights into the pathobiology of infection. The results presented in this study provide evidence that a protein-based host response to ARV infection can be detected in the nasopharyngeal space, and that this response involves perturbation of pathways involved in acute inflammation and innate immune response. Further, this work demonstrates that targeted assays measuring peptides involved in this response allow classification of ARV infection with a high degree of accuracy. Validation of these findings across independent experimentally infected Influenza A and human rhinovirus cohorts suggests a robust and generalized response to viral infection. With further development as a clinical diagnostic, this signature may have utility in rapid screening for emerging infections, avoidance of inappropriate antimicrobial therapy, and more rapid implementation of appropriate therapeutic and public health strategies. Nonetheless, as with other validated biomarkers, *additional validation in community-based cohorts will be important to demonstrate the potential utility of such an assay in its clinical applications.* Furthermore, testing this approach across a larger series of upper respiratory viruses will be important to understand its full potential utility and limitations. An assay that combines host protein biomarkers with nasal viral antigen detection may be quite valuable in clinical care to optimize therapeutic decision making. And whilst a positive result from such an assay may avoid the use of inappropriate microbial therapy, it will likely require vigilance on the part of the clinician to exclude bacterial co-infection when clinically indicated.

Our previous demonstration of the potential for host response-based pre-symptomatic detection of H3N2 infection using blood RNA expression (McClain et al., 2016) raises the intriguing possibility that an NPL protein host response assay might be useful in early detection of ARV infection, and should be evaluated. The availability of a proteomic ‘signature’ that accurately classifies ARV infection and might be migrated to simple and inexpensive antibody-based tests that are routinely used

in both clinical laboratory and over-the-counter diagnostic applications represents an important advance, and may one day yield a ARV host response test that is safe, simple, rapid, inexpensive, and accurate.

## Funding

Support for this work was provided by the U.S. Defense Advanced Research Projects Agency (DARPA) through contract N66001-07-C-2024. E.L.T. and M.T.M. were supported by award numbers 1IK2CX000530 and 1IK2CX000611, respectively, from the Clinical Science Research and Development Service of the VA Office of Research and Development. The funders had no role in the preparation of this manuscript.

## Conflict of Interest

The authors declare that they have no competing interests. A provision patent application has been filed pertaining to the results presented in this article.

## Author Contributions

C.W.W., T.W.B. and G.S.G. conceptualized and directed the project; E.S., J. W. T. and M.A.M. developed the MS methods and generated proteomic data; R.H., J.E.L. contributed to the statistical analyses; R.B.T., R.L.-W., B.P.N. contributed to the conduct of experimental challenge studies and generation of samples, clinical data, and microbiological data; T.W.B., M.T.M., T.V., B.P.N., and C.W.W. contributed to the clinical and microbiological phenotyping of experimental challenge participants; M.N. provided experiment coordination and bioinformatic support; A.O.H. contributed to the host GEA-based phenotyping of research participants; all authors contributed to contributed to review and interpretation of data sets; T.W.B., R.H., E.L.T., and M.T.M. prepared the manuscript with input from all authors.

## Acknowledgements

We thank the staff of the Duke Clinical Research Unit, University of Virginia School of Medicine Research Unit, and hVIVO for their assistance in conducting the experimental challenge studies.

## Appendix A. Supplementary data

Supplementary material.

## References

- Addona, T.A., Abbatiello, S.E., Schilling, B., Skates, S.J., Mani, D.R., Bunk, D.M., Spiegelman, C.H., Zimmerman, L.J., Ham, A.J., Keshishian, H., Hall, S.C., Allen, S., Blackman, R.K., Borchers, C.H., Buck, C., Cardasis, H.L., Cusack, M.P., Dodder, N.G., Gibson, B.W., Held, J.M., Hiltke, T., Jackson, A., Johansen, E.B., Kinsinger, C.R., Li, J., Mesri, M., Neubert, T.A., Niles, R.K., Pulsipher, T.C., Ransohoff, D., Rodriguez, H., Rudnick, P.A., Smith, D., Tabb, D.L., Tegeler, T.J., Variyath, A.M., Vega-Montoto, L.J., Wahlander, A., Waldemarson, S., Wang, M., Whiteaker, J.R., Zhao, L., Anderson, N.L., Fisher, S.J., Liebler, D.C., Paulovich, A.G., Regnier, F.E., Tempst, P., Carr, S.A., 2009. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* 27, 633–641.
- Aebersold, R., Burlingame, A.L., Bradshaw, R.A., 2013. Western blots versus selected reaction monitoring assays: time to turn the tables? *Mol. Cell. Proteomics* 12, 2381–2382.
- Bhoj, V.G., Sun, Q., Bhoj, E.J., Somers, C., Chen, X., Torres, J.P., Mejias, A., Gomez, A.M., Jafri, H., Ramilo, O., Chen, Z.J., 2008. MAVS and MyD88 are essential for innate immunity but not cytotoxic T lymphocyte response against respiratory syncytial virus. *Proc. Natl. Acad. Sci. U. S. A.* 105, 14046–14051.
- Boja, E.S., Rodriguez, H., 2011. The path to clinical proteomics research: integration of proteomics, genomics, clinical laboratory and regulatory science. *Korean J. Lab. Med.* 31, 61–71.
- Byington, C.L., Ampofo, K., Stockmann, C., Adler, F.R., Herbener, A., Miller, T., Sheng, X., Blaschke, A.J., Crisp, R., Pavia, A.T., 2015. Community surveillance of respiratory viruses among families in the Utah better identification of groups-longitudinal viral epidemiology (BIG-LoVE) study. *Clin. Infect. Dis.* 61, 1217–1224.
- Cameron, C.M., Cameron, M.J., Bermejo-Martin, J.F., Ran, L., Xu, L., Turner, P.V., Ran, R., Danesh, A., Fang, Y., Chan, P.K., Mytle, N., Sullivan, T.J., Collins, T.L., Johnson, M.G.,

- Medina, J.C., Rowe, T., Kelvin, D.J., 2008. Gene expression analysis of host innate immune responses during Lethal H5N1 infection in ferrets. *J. Virol.* 82, 11308–11317.
- Cameron, M.J., Ran, L., Xu, L., Danesh, A., Bermejo-Martin, J.F., Cameron, C.M., Muller, M.P., Gold, W.L., Richardson, S.E., Poutanen, S.M., Willey, B.M., Devries, M.E., Fang, Y., Seneviratne, C., Bosinger, S.E., Persad, D., Wilkinson, P., Greller, L.D., Somogyi, R., Humar, A., Keshavjee, S., Louie, M., Loeb, M.B., Brunton, J., McGeer, A.J., Kelvin, D.J., 2007. Interferon-mediated immunopathological events are associated with atypical innate and adaptive immune responses in patients with severe acute respiratory syndrome. *J. Virol.* 81, 8692–8706.
- Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recogn. Lett.* 27, 861–874.
- Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22.
- Gaboriaud, C., Thielens, N.M., Gregory, L.A., Rossi, V., Fontecilla-Camps, J.C., Arlaud, G.J., 2004. Structure and activation of the C1 complex of complement: unraveling the puzzle. *Trends Immunol.* 25, 368–373.
- Gerszten, R.E., Carr, S.A., Sabatine, M., 2010. Integration of proteomic-based tools for improved biomarkers of myocardial injury. *Clin. Chem.* 56, 194–201.
- Hong, C.Y., Lin, R.T., Tan, E.S., Chong, P.N., Tan, Y.S., Lew, Y.J., Loo, L.H., 2004. Acute respiratory symptoms in adults in general practice. *Fam. Pract.* 21, 317–323.
- House, W., 2014. National Strategy for Combating Antibiotic-Resistant Bacteria (2014 Sep [accessed 2016 April 8]).
- Huang da, W., Sherman, B.T., Lempicki, R.A., 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13.
- Huang, Y., Zaas, A.K., Rao, A., Dobeigeon, N., Woolf, P.J., Veldman, T., Oien, N.C., McClain, M.T., Varkey, J.B., Nicholson, B., Carin, L., Kingsmore, S., Woods, C.W., Ginsburg, G.S., Hero 3rd, A.O., 2011. Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. *PLoS Genet.* 7, e1002234.
- Jackson, G.G., Dowling, H.F., Spiessman, I.G., Boand, A.V., 1958. Transmission of the common cold to volunteers under controlled conditions. I. The common cold as a clinical entity. *AMA Arch. Intern. Med.* 101, 267–278.
- Johnston, S.L., Sanderson, G., Pattermore, P.K., Smith, S., Bardin, P.G., Bruce, C.B., Lambden, P.R., Tyrrell, D.A., Holgate, S.T., 1993. Use of polymerase chain reaction for diagnosis of picornavirus infection in subjects with and without respiratory symptoms. *J. Clin. Microbiol.* 31, 111–117.
- Johnstone, J., Majumdar, S.R., Fox, J.D., Marrie, T.J., 2008. Viral infection in adults hospitalized with community-acquired pneumonia: prevalence, pathogens, and presentation. *Chest* 134, 1141–1148.
- Kimura, H., Yoshizumi, M., Ishii, H., Oishi, K., Ryo, A., 2013. Cytokine production and signaling pathways in respiratory virus infection. *Front. Microbiol.* 4, 276–284.
- Kiyonami, R., Schoen, A., Prakash, A., Peterman, S., Zabrouskov, V., Picotti, P., Aebersold, R., Huhmer, A., Domon, B., 2011. Increased selectivity, analytical precision, and throughput in targeted proteomics. *Mol. Cell. Proteomics* 10 (M110), 002931.
- Koyama, S., Ishii, K.J., Coban, C., Akira, S., 2008. Innate immune response to viral infection. *Cytokine* 43, 336–341.
- Liu, T.-Y., Burke, T., Park, L.P., Woods, C.W., Zaas, A.K., Ginsburg, G.S., Hero, A.O., 2016. An individualized predictor of health and disease using paired reference and target samples. *BMC Bioinforma.* 17, 1–15.
- Marion, T., Elbaheesh, H., Thomas, P.G., Devincenzo, J.P., Webby, R., Schughart, K., 2016. Respiratory mucosal proteome quantification in human influenza infections. *PLoS One* 11, e0153674.
- McClain, M.T., Nicholson, B.P., Park, L.P., Liu, T.-Y., Hero, A.O., Tsalik, E.L., Zaas, A.K., Veldman, T., Hudson, L.L., Lambkin-Williams, R., Gilbert, A., Burke, T., Nichols, M., Ginsburg, G.S., Woods, C.W., 2016. A genomic signature of influenza infection shows potential for presymptomatic detection, guiding early therapy, and monitoring clinical responses. *Open Forum Inf. Dis.* 3.
- O'Neill, J., 2015. Rapid Diagnostics: Stopping Unnecessary Use of Antibiotics (Review on Antimicrobial Resistance).
- Organization, W.H., 2015. Global Action Plan on Antimicrobial Resistance. World Health Organization.
- Oshansky, C.M., Gartland, A.J., Wong, S.S., Jeevan, T., Wang, D., Roddam, P.L., Caniza, M.A., Hertz, T., Devincenzo, J.P., Webby, R.J., Thomas, P.G., 2014. Mucosal immune responses predict clinical outcomes during influenza infection independently of age and viral load. *Am. J. Respir. Crit. Care Med.* 189, 449–462.
- Pirillo, A., Catapano, A.L., Norata, G.D., 2015. HDL in infectious diseases and sepsis. *Handb. Exp. Pharmacol.* 224, 483–508.
- Ramilo, O., Allman, W., Chung, W., Mejias, A., Ardura, M., Glaser, C., Wittkowski, K.M., Piqueras, B., Banchereau, J., Palucka, A.K., Chaussabel, D., 2007. Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* 109, 2066–2077.
- Ramilo, O., Mejias, A., 2009. Shifting the paradigm: host gene signatures for diagnosis of infectious diseases. *Cell Host Microbe* 6, 199–200.
- Ricklin, D., Hajishengallis, G., Yang, K., Lambris, J.D., 2010. Complement: a key system for immune surveillance and homeostasis. *Nat. Immunol.* 11, 785–797.
- Thompson, M., Shay, D., Zhou, H., Bridges, C., Cheng, P., Burns, E., Breesee, J., Cox, N., 2010. Estimates of deaths associated with seasonal influenza—United States, 1976–2007. *Morb. Mortal. Wkly Rep.* 59, 1057–1062.
- Tsalik, E.L., Henao, R., Nichols, M., Burke, T., Ko, E.R., McClain, M.T., Hudson, L.L., Mazur, A., Freeman, D.H., Veldman, T., Langley, R.J., Quackenbush, E.B., Glickman, S.W., Cairns, C.B., Jaehne, A.K., Rivers, E.P., Otero, R.M., Zaas, A.K., Kingsmore, S.F., Lucas, J., Fowler, V.G., Carin, L., Ginsburg, G.S., Woods, C.W., 2016. Host gene expression classifiers diagnose acute respiratory illness etiology. *Sci. Transl. Med.* 8, 322ra11.
- Woods, C.W., McClain, M.T., Chen, M., Zaas, A.K., Nicholson, B.P., Varkey, J., Veldman, T., Kingsmore, S.F., Huang, Y., Lambkin-Williams, R., Gilbert, A.G., Hero 3rd, A.O., Ramsburg, E., Glickman, S., Lucas, J.E., Carin, L., Ginsburg, G.S., 2013. A host transcriptional signature for presymptomatic detection of infection in humans exposed to influenza H1N1 or H3N2. *PLoS One* 8, e52198.
- Yoneyama, M., Fujita, T., 2010. Recognition of viral nucleic acids in innate immunity. *Rev. Med. Virol.* 20, 4–22.
- Zaas, A.K., Burke, T., Chen, M., McClain, M., Nicholson, B., Veldman, T., Tsalik, E.L., Fowler, V., Rivers, E.P., Otero, R., Kingsmore, S.F., Voora, D., Lucas, J., Hero, A.O., Carin, L., Woods, C.W., Ginsburg, G.S., 2013. A host-based RT-PCR gene expression signature to identify acute respiratory viral infection. *Sci. Transl. Med.* 5, 203ra126.
- Zaas, A.K., Chen, M., Varkey, J., Veldman, T., Hero 3rd, A.O., Lucas, J., Huang, Y., Turner, R., Gilbert, A., Lambkin-Williams, R., Oien, N.C., Nicholson, B., Kingsmore, S., Carin, L., Woods, C.W., Ginsburg, G.S., 2009. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host Microbe* 6, 207–217.
- Zaas, A.K., Garner, B.H., Tsalik, E.L., Burke, T., Woods, C.W., Ginsburg, G.S., 2014. The current epidemiology and clinical decisions surrounding acute respiratory infections. *Trends Mol. Med.* 20, 579–588.