



Research Paper

Nasopharyngeal metabolomics and machine learning approach for the diagnosis of influenza

Catherine A. Hogan^{a,b,1,*}, Pranav Rajpurkar^{c,1}, Hari Sowrirajan^{c,1}, Nicholas A. Phillips^{c,1}, Anthony T. Le^a, Manhong Wu^d, Natasha Garamani^a, Malaya K. Sahoo^a, Mona L. Wood^{a,b}, ChunHong Huang^a, Andrew Y. Ng^c, Justin Mak^e, Tina M. Cowan^{a,e}, Benjamin A. Pinsky^{a,b,f}

^a Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA

^b Clinical Virology Laboratory, Stanford Health Care, Palo Alto, CA 94304, USA

^c Stanford Computer Science Department, Stanford University, Stanford, CA 94305, USA

^d Stanford Department of Anesthesiology, Stanford University, Stanford, CA 94305, USA

^e Stanford Biochemical Genetics Laboratory, Stanford Health Care, Palo Alto, CA 94304, USA

^f Division of Infectious Diseases and Geographic Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA



ARTICLE INFO

Article History:

Received 26 March 2021

Revised 2 August 2021

Accepted 6 August 2021

Available online xxx

Keywords:

Metabolomics

Mass spectrometry

Influenza

Diagnosis

Host response

ABSTRACT

Background: Respiratory virus infections are significant causes of morbidity and mortality, and may induce host metabolite alterations by infecting respiratory epithelial cells. We investigated the use of liquid chromatography quadrupole time-of-flight mass spectrometry (LC/Q-TOF) combined with machine learning for the diagnosis of influenza infection.

Methods: We analyzed nasopharyngeal swab samples by LC/Q-TOF to identify distinct metabolic signatures for diagnosis of acute illness. Machine learning models were performed for classification, followed by Shapley additive explanation (SHAP) analysis to analyze feature importance and for biomarker discovery.

Findings: A total of 236 samples were tested in the discovery phase by LC/Q-TOF, including 118 positive samples (40 influenza A 2009 H1N1, 39 influenza H3 and 39 influenza B) as well as 118 age and sex-matched negative controls with acute respiratory illness. Analysis showed an area under the receiver operating characteristic curve (AUC) of 1.00 (95% confidence interval [95% CI] 0.99, 1.00), sensitivity of 1.00 (95% CI 0.86, 1.00) and specificity of 0.96 (95% CI 0.81, 0.99). The metabolite most strongly associated with differential classification was pyroglutamic acid. Independent validation of a biomarker signature based on the top 20 differentiating ion features was performed in a prospective cohort of 96 symptomatic individuals including 48 positive samples (24 influenza A 2009 H1N1, 5 influenza H3 and 19 influenza B) and 48 negative samples. Testing performed using a clinically-applicable targeted approach, liquid chromatography triple quadrupole mass spectrometry, showed an AUC of 1.00 (95% CI 0.998, 1.00), sensitivity of 0.94 (95% CI 0.83, 0.98), and specificity of 1.00 (95% CI 0.93, 1.00). Limitations include lack of sample suitability assessment, and need to validate these findings in additional patient populations.

Interpretation: This metabolomic approach has potential for diagnostic applications in infectious diseases testing, including other respiratory viruses, and may eventually be adapted for point-of-care testing.

© 2021 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

Over the last decade, the diagnosis and monitoring of infectious diseases has been revolutionized by molecular testing, including the widespread use of Polymerase Chain Reaction (PCR) in Clinical Microbiology and Virology Laboratories. These methods are rapid and

highly accurate; however, important limitations remain unaddressed, including high cost, high complexity, inability to differentiate active infection from latency or colonization, and lack of sensitivity in direct patient specimens [1–4]. Moreover, although molecular testing availability has improved overall, it remains largely restricted to centralized laboratories especially in the context of kit shortages for point-of-care testing [5]. In addition, traditional rapid influenza diagnostic tests and digital immunoassays have been limited by suboptimal sensitivity, and alternatives are needed [6]. Accurate testing is particularly important for respiratory viruses including influenza, which are estimated to have caused over 35 million symptomatic illnesses

* Corresponding author at: 655 W 12th Ave, Room 2054, Vancouver, BC, Canada, V6R 2M7

E-mail address: catherine.hogan@bccdc.ca (C.A. Hogan).

¹ These authors contributed equally to this study.

Research in context

Evidence before this study

We performed a literature search of 3 major databases (PubMed, Embase, and Cochrane), and medRxiv for preprints. The search identified English and French studies published from January 1 2010 to May 31 2021 using the keywords “influenza”, and “metabolomics”, “mass spectrometry”, “quadrupole time of flight”, “triple quadrupole”, and similar terms, and “nasopharyngeal”, “samples”, “specimens” and similar terms. Overall, several studies have described the use of untargeted metabolomics for the diagnosis and characterization of influenza infection, mainly using viral cell culture lines and/or animal models. We found a single study that employed proteomics for characterization of influenza from nasopharyngeal swab specimens. Furthermore, comprehensive quantitative analysis and diagnostic test performance indicators were infrequently reported.

Added value of this study

We demonstrated the feasibility and high accuracy of an untargeted metabolomics approach from nasopharyngeal samples combined with machine learning for the identification of distinct metabolic signatures for the diagnosis of influenza infection. This study draws on a larger dataset than previously employed, and this approach maintained high performance after adaptation to clinically-adaptable LC/MS-MS instruments on an independent validation cohort.

Implications of all the available evidence

Our study reinforced the potential of metabolomics as a diagnostic approach for clinical virology application, and demonstrated successful adaptation for clinically-adaptable testing. Further work assessing performance for detection of other pathogenic targets and patient populations will be required to further characterize potential for clinical use.

signatures directly from patient specimens by liquid chromatography quadrupole time-of-flight mass spectrometry (LC/Q-TOF) using a simplified experimental workflow (**Fig. S1**) [20]. The objective of this study was to use this LC/Q-TOF method to generate data to develop and validate machine learning (ML) algorithms for classification of influenza infection status, and an interpretation method for biomarker discovery (**Fig. 1**). The developed top-20 ion feature signature was then successfully adapted to testing on a clinically-applicable, targeted triple quadrupole mass spectrometry instrument (LC/MS-MS; referred to as tandem mass spectrometry) for validation on upper respiratory tract specimens.

2. Methods

2.1. Resource availability

2.1.1. Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Catherine Hogan (hoganca@stanford.edu).

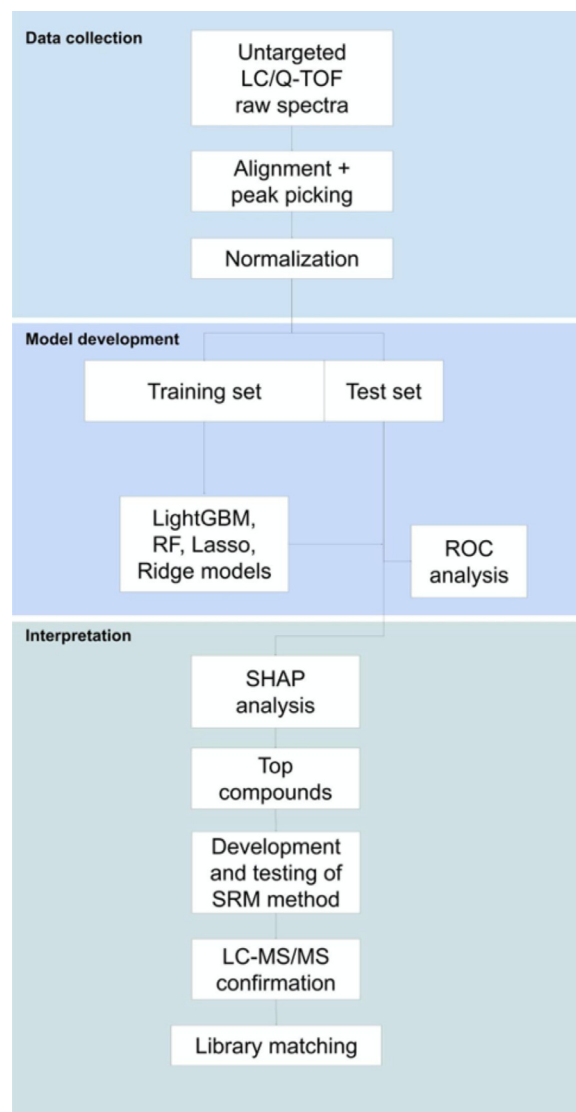


Fig 1. Conceptual diagram of the study from data collection to interpretation. The phases of data collection, model development, and interpretation are illustrated. LC/Q-TOF: liquid chromatography quadrupole time-of-flight; LC-MS/MS: liquid chromatography-mass spectrometry; RF: random forests; ROC: receiver operating characteristic curve; SHAP: Shapley Additive explanation; SRM: selected reaction monitoring.

during the 2018–2019 season alone in the United States [7]. These viruses infect respiratory epithelial cells, where they may induce metabolite alterations in the host [8,9]. The ‘-omics’ field, including genomics, proteomics, and metabolomics, has been studied for the diagnosis of influenza with variable success [10–16]. Of the three fields, genomics has been the most utilized for clinical virology purposes; however, there is significant interest in leveraging alternative approaches such as metabolomics and proteomics to address remaining gaps. Metabolomics, or the large-scale study of small molecules, represents a change in paradigm from routine clinical virology diagnostics as it detects host metabolic response rather than directly detecting the pathogen [17]. Metabolomics theoretically holds promise for infectious diseases applications as it can be performed directly from patient specimens from minimal sample volume, is inexpensive to run, provides a real-time assessment of host response and may accurately differentiate active infection from colonization [18,19]. However, approaches until now using viral culture have been hampered by low sensitivity and prolonged turnaround time.

Nasopharyngeal swab sampling followed by swab immersion in viral transport medium (VTM) is the most common collection technique for the diagnosis of respiratory viruses and enables the non-invasive collection of respiratory cells. We hypothesized that analysis of VTM after nasopharyngeal sampling using a recently reported and sensitive in-line two-column metabolomics method would reveal distinct signatures for the diagnosis of infectious diseases [20]. This method is well suited for the characterization of host metabolite

2.1.2. Materials availability

This study did not generate new unique reagents.

2.1.3. Data and code availability

The data and code generated during this study will be made available at <https://github.com/stanfordmlgroup/influenza-qtof>.

3. Experimental model and subject details

For the discovery cohort, we selected stored specimens collected from April 23 2015 to October 13 2019 to achieve a 1:1 ratio of positive to age and sex-matched negative controls. Age-matching was performed to the identical age, or within 2 years if not available. We included specimens from 96 children (2–17 years-old) and 140 adults (≥ 18 years-old). These corresponded to 123 males and 113 females. Mixed infections and samples from other sites (e.g., oropharyngeal swab, bronchoalveolar lavage and lung tissue) were excluded. Individual retrospective chart review was performed for all subjects in the untargeted phase of the study to identify age, sex, immunocompromised status, comorbidities, disease severity, antiviral treatment and clinical outcomes. LC/Q-TOF testing was performed to generate raw data on mass-to-charge ratio and retention time for each sample tested.

For the validation cohort, we prospectively selected negative and positive nasopharyngeal and nasal swab specimens from December 21 2019 to February 18 2020 in a 1:1 ratio without exclusion. Samples were subsequently stored at -80 °C until testing. Testing was performed at the Stanford Biochemical Genetics Laboratory using a validation sample set of 96 samples tested by LC/MS-MS. Of the individuals with available demographic data, there were 14 children and 80 adults, corresponding to 39 females and 55 males. There were three individuals with documented viral coinfection (seasonal coronavirus, RSV or CMV) in the validation cohort.

4. Method details

4.1. Study design and patient population

The research objective was to assess the diagnostic test performance of the LC/Q-TOF (biomarker discovery cohort) and targeted analysis (validation cohort) for the diagnosis of influenza-infected vs uninfected individuals, and to identify key metabolites for classification of these two groups. In both the discovery and validation cohorts, target sample size calculation was based on the DeLong method. The test cohort, which represented 20% of the discovery cohort corresponding to 48 patients, provides over 90% statistical power to detect an AUC of 0.925 between influenza-infected and uninfected individuals using an AUC hypothesis test of 0.50, with a significance level of 0.05.

A secondary endpoint of influenza A vs influenza B was established in the study design phase, and used as an exploratory endpoint. The target sample size was not changed during the study. Nasopharyngeal samples collected from adult patients from Stanford Health Care (SHC) and children from the Lucille Packard Children's Hospital (LPCH) were processed per routine clinical procedures. Briefly, a flocked swab is inserted in the nasal passage, rotated for collection of cells for 10–15 s and placed in viral transport medium (MicroTest M4RT, Remel Inc., San Diego, CA). Respiratory viral testing was performed on the ePlex Respiratory Pathogen (RP) panel (GenMark Diagnostics, Carlsbad, CA) at the Stanford Clinical Virology Laboratory, as per routine clinical test procedures. This automated qualitative (detected/not detected) nucleic acid amplification test (NAAT) identifies 15 viral targets, including influenza A, influenza H1N1 2009, influenza A H3 and influenza B. Results from this assay were used as the reference for this study. Specimens tested by the ePlex RP panel and resulted as indeterminate were not included for

the study. Specimens were aliquoted and stored at -80 °C without additional handling until subsequent LC/Q-TOF testing. Specimen processing was performed the same way for specimens from individuals with and without influenza infection. There were no adverse effects related to use of the reference (RT-PCR) and index tests (metabolomics).

For the discovery cohort, we selected stored specimens from individuals assessed at Stanford Health Care and Stanford Children's Health with and without influenza infection collected from April 23 2015 to October 13 2019 to achieve a 1:1 ratio of positive to age and sex-matched negative controls. A convenience set was selected. Specimens from pediatric individuals were included given the burden of respiratory viruses in these groups, with age-matching to account for potential metabolomic changes by age group. Infants aged less than 2 years-old were excluded due to the limited number of specimens available from this age group. Age-matching was performed to the identical age, or within 2 years if not available. We included specimens from 96 children (2–17 years-old) and 140 adults (≥ 18 years-old). These corresponded to 123 males and 113 females. Mixed infections and samples from other sites (e.g., oropharyngeal swab, bronchoalveolar lavage and lung tissue) were excluded. Individual retrospective chart review was performed for all subjects in the untargeted phase of the study to identify age, sex, immunocompromised status, comorbidities, disease severity, antiviral treatment and clinical outcomes. Data collection was performed after the reference test (RT-PCR) and before the index test (metabolomics). Testing was performed for symptomatic individuals, most commonly from an upper respiratory tract infection (URTI) viral syndrome. LC/Q-TOF testing was performed to generate raw data on mass-to-charge ratio and retention time for each sample tested. Single replicate testing was performed, and outlier data points were included for analysis. Clinical and reference testing data were not available to the individuals performing processing and performing set-up for metabolomics testing.

For the validation cohort, we prospectively selected negative and positive nasopharyngeal and nasal swab specimens from December 21 2019 to February 18 2020 in a 1:1 ratio without exclusion. Samples were subsequently stored at -80 °C until testing. Testing was performed at the Stanford Biochemical Genetics Laboratory. A validation sample set of 96 samples was tested. Of the individuals with available demographic data, there were 14 children and 80 adults, corresponding to 39 females and 55 males. There were three individuals with documented viral coinfection (seasonal coronavirus, RSV or CMV) in the validation cohort. LC/MS-MS testing was performed to generate raw data on mass-to-charge ratio and retention time for each sample tested. Single replicate testing was performed, and outlier data points were included for analysis. This method served to confirm the results from the LC/Q-TOF analysis in a separate participant cohort.

4.2. Ethics

This study was approved by the Stanford Institutional Review board (IRB protocol #48973). Per IRB assessment, informed consent was waived for this study.

4.3. LC/Q-TOF method

Liquid chromatography (LC) separation was performed on an Agilent 1290 Quaternary LC system (Agilent Technologies). In this unique chromatographic arrangement, two columns are used in-line: a reverse-phase (RP) column of 2.1×50 mm $1.8 \mu\text{m}$ HSS T3 (Waters Corporation, Milford, MA) is placed first followed by an ion exchange (IEX) column of 2.0×30 mm $3 \mu\text{m}$ Intrada (Imtakt USA, Portland, OR). Both columns are joined with EXP2 fittings (Optimize Technologies, OR). Mass spectrometry was performed on an Agilent 6545 Q-

TOF instrument with electrospray ionization. The mobile phases were (A) 150 mg of ammonium formate per liter water with 0.4% formic acid (v/v), (B) 1.2 g of ammonium formate per liter of methanol with 0.2% formic acid, and (C) water with 1% each formic acid and ammonium hydroxide, as previously described [20]. The flow rate was 0.5 mL/min, column temperature of 45 °C and injection volume of 5 μ L, for a total run time of 20 min (inject-to-inject). MS was performed on an Agilent 6545 Q-TOF with dual Agilent JetStream electrospray ionization, as previously described [20]. This LC/Q-TOF method has previously been shown to demonstrate high analytical data quality including peak area precision, with most QC samples showing <30% coefficient of variation [20]. The instrument was operated in sensitivity-mode with extended dynamic range and positive polarity, scanning from 50 to 1100 m/z. Two reference ions were used: purine (m/z 121.050873), and hexakis (1H,1H,3H-tetrafluoropropoxy) phosphazene (m/z 922.009798).

4.4. LC/Q-TOF metabolite extraction and analysis

A volume of 100 μ L of nasopharyngeal sample eluted in VTM was processed by ultrafiltration using Pall Omega 3kDa centrifugal devices (VWR, Radnor, PA) at 4 °C for 15 min at 17,000 x g. The filtrate was transferred to glass vials and analyzed, and each sample was run once. Two quality control (QC) samples one pooled QC sample and an independent normalization QC sample were used to assess for batch effect. The pooled QC was created by pooling an equal volume of aliquots from all the samples included in the run. Unsupervised principal component analysis was performed to visually assess appropriate performance of the pooled QC. The normalization QC was picked as a random nasopharyngeal swab clinical specimen not included in this study. In addition, blank VTM was run in triplicate to generate a mean background spectral distribution. Progenesis QI software (Waters Corporation) was used for run alignment, peak picking (automatic, level 4), adduct deconvolution, and feature identification. Positive polarity analysis was performed using the adducts [M+H], [M+NH₄] and [M+Na]. The fragment ions were obtained for the top 20 ion features by subjecting the samples to triple quadrupole system using a collision energy of 15V with the precursor ions identified from Progenesis analysis. Metabolite identification was first performed using a previously-developed authentic standard library [20]. If there was no identification match, preliminary annotation was performed in Progenesis QI software using the HMDB [21] and KEGG [22] plug-ins, and by manual review in the NIST 20 MSMS library and METLIN. A mass error setting of 30 ppm was used. Data were directly exported from Progenesis for machine learning analysis using peak area filter thresholds of 0; ≥ 5000 ; $\geq 10,000$ and $\geq 20,000$ relative abundance values. Outlier values were not excluded, and no additional instrument data processing was performed.

4.5. LC-MS/MS targeted method

The targeted analysis was performed using a clinically-validated method that detects pyroglutamic acid, as previously described [23,24]. Mass spectrometry was performed on an Agilent 6460 Triple Quadrupole mass spectrometer equipped with an Agilent JetStream electrospray ionization. Selected reaction monitoring (SRM) pairs based on the important ion features were added to the method (Table S4). The data were acquired using MassHunter WorkStation Acquisition version B.08.02 (Agilent) and exported for machine learning analysis.

4.6. LC-MS/MS metabolite extraction and analysis

A volume of 100 μ L of respiratory specimen eluted in VTM or phosphate buffered saline (PBS) and 10 μ L of pyroglutamic acid-D5 0.025 nm/L as internal standard (Cambridge Isotope Laboratories, Inc,

Tewksbury, MA) was processed by ultrafiltration using Pall Omega 3kDa centrifugal devices (VWR, Radnor, PA) at 4 °C for 15 min at 17,000 x g. The filtrate was transferred to glass vials and analyzed. The data were acquired using MassLynx version 4.2 (Waters Corp). The median percent coefficient of variation (CV) of the D5-pyroglutamic acid across all samples was <15%.

5. Quantification and statistical analysis

5.1. Statistical analysis

Statistical analysis was performed by Chi-squared test (categorical variables if 5 or more variables per cell) or Fisher's exact test (categorical variables if less than 5 variables per cell) and Mann-Whitney U test (continuous variables), using Stata v15.1 (Stata Corp, College Station, TX). Missing data are identified as unknown. A two-sided p value of <0.05 was considered significant. The sample size calculation was based on the DeLong method. The test cohort with 48 patients provides over 90% statistical power to detect an AUC of 0.925 between influenza-infected and uninfected individuals using a two-sided AUC hypothesis test of 0.50, with a significance level of 0.05.

5.2. Machine learning analysis

We developed machine learning methods for the task of determining whether a sample was positive or negative for influenza based on its metabolic profile. Machine learning is a class of techniques that uses data to learn a model that maps an input (the metabolic profile of a sample; includes mass-to-charge ratio (m/z), retention time and relative abundance for each sample) to its associated output (the influenza infection outcome of the sample based on the reference standard) and uses this learned model on new inputs (the metabolic profiles of new samples) to make predictions of new outputs (the influenza outcomes of new samples). We implemented two machine learning methods: gradient boosted decision trees and random forests (RFs). Each of these methods was used separately to build an independent model.

Gradient boosted decision trees and RFs are both ensemble learning methods that improve upon the performance of decision tree models. Decision tree learners construct a model by iteratively identifying which feature most effectively divides the data into groups with low within-group variation in the outcome and high between-group variation in outcome, and then repeat the process within each group. Gradient boosted decision trees (GBDT) construct several decision trees such that each tree learns from the errors of the prior tree [25]. Light gradient boosted model (LGBM) is a particular implementation of GBDT that is based on a unique algorithm to identify the split value of categorical variables. Random forests (RF) construct several decision trees such that each tree is constructed using different subsets of the data. The machine learning approaches of LGBM and RF were chosen over alternative machine learning methods because they can handle mixes of categorical and continuous covariates, capture non-linear relationships, and scale well to large amounts of data.

Ion features showing zero values through all samples tested were removed from the dataset. The remaining dataset was partitioned without normalization into a training set used to develop machine learning models, and a holdout test set used to evaluate the predictive performance of the machine learning models. The partitioning of the dataset was random such that 80% of the samples were included in the training set, and the other 20% in the test set. There was no overlap between the samples and patients between the two sets.

All models were developed on the training set, and their final performance reported on the holdout test set and/or the prospective cohort. The models were not retrained using SRM data to avoid overfitting and overestimating test performance. In addition, within the

training set, cross-validation was used to develop the models to avoid overfitting to the training set. In the cross-validation procedure, the training dataset was randomly partitioned into $k = 4$ equal sized subsamples consisting of an approximately equal percentage of each class. Of the k subsamples, a single subsample was retained as the validation data for the model, and the remaining $k - 1$ subsamples were used to train a model. The cross-validation process was then repeated k times, with each of the k subsamples used exactly once as the validation data. Grid search was used to find the best set of hyperparameters for model training; the same hyperparameter settings were used across all k folds. The resulting k models (one from each fold) were used to make k sets of predictions on the test set, which were then averaged using a simple mean to make the final prediction for each sample in the test set.

To determine the usefulness of capturing non-linear relationships with machine learning models, the modelling approaches using two machine learning methods, gradient boosted decision trees and random forests, were compared with two traditional linear models, Least absolute shrinkage and selection operator (Lasso) and Ridge. These models are variants of Logistic regression, a statistical model that uses the logistic function to model the outcome assuming a linear relationship between the features and the outcome. Lasso makes the same linear assumption but alters the model fitting process to select only a subset of the features for use in the final model rather than using all of them. Unlike Lasso, Ridge will not result in a sparse model, but rather addresses multicollinearity in the features by shrinking the weights assigned to correlated variables. The training and test sets, and the cross-validation strategy were identical across the machine learning models and traditional linear models.

The SHAP (SHapley Additive exPlanations) method was used to quantify the impact of each feature on the models. The method explains prediction by allocating credit among the input features; feature credit is calculated using Shapley Values [26,27] as the change in the expected value of the model's prediction of improvement for a symptom when a feature is observed versus unknown. To uncover clinically important ion features that were globally predictive of the outcome, the Shapley values for the top 20 ion features on individual predictions were aggregated and reported along with their averaged absolute Shapley contributions as a percent of the contributions of all the features. Our goal was to select the top 20 differentiating metabolites capable of performing classification in a prospective cohort of samples. While Lasso produces a sparse model, its top features are by no means guaranteed to perform well on the prospective cohort. Hence, we chose to use SHAP to identify the top metabolites used by the LGBM model, which recorded the highest performance on the retrospective dataset, as they would be the most likely to perform strongly in prospective analysis. Thus, the SHAP method assessment was based on the highest performing of the four classification models (RF, LGBM, Lasso, or Ridge) for the discovery and validation datasets. Limiting the number of features to the top 20 was also consistent with an approach to reduce the risk of overfitting.

An exploratory subgroup analysis was used to evaluate variation in model performance across patient subpopulations. We trained an LGBM model using the previously described cross-validation strategy on the discovery training set and generated predictions with this model on the discovery test set. We then split the test samples into disjoint subpopulations and reported the AUC and confidence interval using DeLong's method for each subgroup. We investigated the following subgroups: adult vs pediatric individuals, immunocompromised vs not, ICU-admitted vs not, antibiotic-treated vs not, bacterial coinfection vs not, and by time since symptom onset at the time of respiratory viral testing (<7 days vs ≥ 7 days).

A multivariable analysis was used to investigate the significance of potential confounders in the analysis. We first trained our model and generated predictions on the discovery test set using the previously described methods. We then performed an additional logistic

regression on the true label with predictors comprising predicted score, age, sex, number of days since symptom onset, Charlson comorbidity index score, and hospitalization status. The significance of each predictor was determined using the p -value from this regression.

The primary measure of model performance was the area under the receiver operating characteristic curve (AUC), which illustrates the diagnostic discriminative performance of the models. Performance measures for the models also included sensitivity, specificity, and accuracy at a high-sensitivity operating point used to binarize the model predictions. The high-sensitivity operating point was selected by selecting a high-sensitivity operating point on each of the k validation folds and averaging them: on each validation fold, an operating point that maximized the Youden's J statistic and produced a sensitivity of at least 0.9 was selected. To assess the variability in estimates, we provide 95% Wilson score confidence intervals for sensitivity, specificity, and accuracy and 95% DeLong confidence intervals for AUC [28].

Analyses were performed in Python version 3.6.8, using the LightGBM v2.2.3 implementation for gradient boosted decision trees, scikit-learn v0.20.2 for RF, stratified k-fold cross-validation and grid search [29], SHAP (SHapley Additive exPlanations) v0.29.1 for computing feature importance, and R version 3.5.0 for statistical analysis.

5.3. Role of the funding source

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

6. Results

6.1. Biomarker discovery phase

A total of 248 samples were analyzed by LC/Q-TOF for metabolite discovery. Of these, 6 were excluded prior to analysis due to technical errors and their 6 corresponding controls were excluded (Fig. S2). The final analysis included a total of 236 samples, with 118 positive influenza samples (40 influenza A 2009 H1N1, 39 influenza A H3 and 39 influenza B) and 118 negative age and sex-matched controls (Table 1). The discovery cohort training set consisted of 186 samples (94 positive, 92 negative), and the test set consisted of 50 samples (24 positive, 26 negative). Compared to individuals without influenza, those with a positive influenza result were more likely to have been tested at an outpatient clinic (63.6% vs 26.3%; $p < 0.001$ [Chi-squared]), less likely to be immunocompromised (22.9% vs 45.8%; $p = 0.001$ [Chi-squared]), less likely to have been hospitalized (24.6% vs 69.5%; $p < 0.001$ [Chi-squared]) and less likely to have been admitted to the intensive care unit (ICU) (5.1% vs 22.0%; $p < 0.001$ [Chi-squared]). Furthermore, individuals with a positive influenza result were more likely to have received antibiotic therapy (32.2% vs 13.6%; $p < 0.001$ [Chi-squared]), and to have confirmed bacterial coinfection or colonization (32.2% vs 6.7%; $p < 0.001$ [Chi-squared]). Symptoms most commonly involved an upper respiratory tract infection (URTI) syndrome, and patient characteristics were otherwise similar. All-cause 30-day mortality was identical in each group at 3/118 (2.5%).

6.2. LC/Q-TOF metabolomics combined with machine learning demonstrates high classification performance

Untargeted metabolomics by LC/Q-TOF identified a total of 3366 ion features. Of these, 48 ion features were removed given they showed zero values for all clinical samples tested, but were detected in a QC sample, leaving 3318 ion features for analysis. Principal component analysis representation of these data is presented separately (Fig. S3). Application of machine learning models to these features, specifically the LightGBM (LGBM) and random forest (RF) models, based on the 20% of data reserved for testing, achieved an area under

Table 1
Baseline demographic characteristics of all patients in the biomarker discovery (LC/Q-TOF) phase of the study.

		Non-influenza acute respiratory disease (n = 118)	Influenza acute respiratory disease (n = 118)	p-value*
Age (No. [%])	≥2yo–17yo	48 (40.7%)	48 (40.7%)	1.0
	≥18yo	70 (59.3%)	70 (59.3%)	
Sex (No. [%])	Male	62 (52.5)	61 (51.7)	0.9
	Female	56 (47.5)	57 (48.3)	
Immunocompromised (No. [%])	Yes	54 (45.8%)	27 (22.9%)	0.001
	No	63 (53.4%)	87 (73.7%)	
	Unknown	1 (0.8%)	4 (3.4%)	
Comorbidities (No. [%])	Leukemia/lymphoma	27 (22.9%)	10 (8.5%)	0.005
	Active malignancy	10 (8.5%)	2 (1.7%)	0.02
	Asthma	6 (5.1%)	7 (5.9%)	0.5
Median Charlson comorbidity index score (IQR)	1 (0–3)	0 (0–2)	0.002	
Days of symptoms at the time of testing (mean; SD)	3 (1–7)	3 (2–9)	0.4	
Patient location (No. [%])	ED	41 (34.8%)	36 (30.5%)	<0.001
	ICU	16 (13.6%)	4 (3.4%)	
	Inpatient ward	30 (25.4%)	3 (2.5%)	
	Outpatient clinic	31 (26.3%)	75 (63.6%)	
Antiviral treatment at time of testing (No. [%])	Yes	0	3 (2.5%)	0.1
	No	114 (96.6%)	96 (81.4%)	
	Unknown	4 (3.4%)	19 (16.1%)	
Antibiotic treatment at time of testing (No. [%])	Yes	16 (13.6)	38 (32.2)	<0.001
	No	82 (69.5%)	76 (64.4)	
	Unknown	20 (17.0)	4 (3.4)	
Confirmed bacterial coinfection or colonization at time of testing (No. [%])	Yes	8 (6.7)	38 (32.2)	<0.001
	No	110 (92.4)	76 (64.4)	
	Unknown	1 (0.8)	4 (3.4)	
Hospitalization (No. [%])	Yes	82 (69.5%)	29 (24.6%)	<0.001
	No	36 (30.5%)	89 (75.4%)	
ICU admission (No. [%])	Yes	26 (22.0%)	6 (5.1%)	<0.001
	No	92 (78.0%)	112 (94.9%)	
30-day all-cause mortality (No. [%])	Yes	3 (2.5%)	3 (2.5%)	1.0
	No	115 (97.5%)	116 (97.5%)	

The *p* values were calculated by Chi-squared if categorical variables, by Fisher's exact test if categorical variables with less than 5 datapoints per cell, and by Mann Whitney *U* test for continuous variables.

ED: emergency department; ICU: intensive care unit; IQR: inter-quartile range; SD: standard deviation; yo: years-old.

the receiver operating characteristic curve (AUC) of 1.00 (95% CI 0.99, 1.00) and 0.93 (95% CI 0.86, 1.00), respectively, on the test set (Fig. 2). Statistical models, specifically the Lasso and Ridge regression models, obtained AUCs of 0.94 (95% CI 0.88, 1.00) and 0.92 (95% CI 0.85, 1.00),

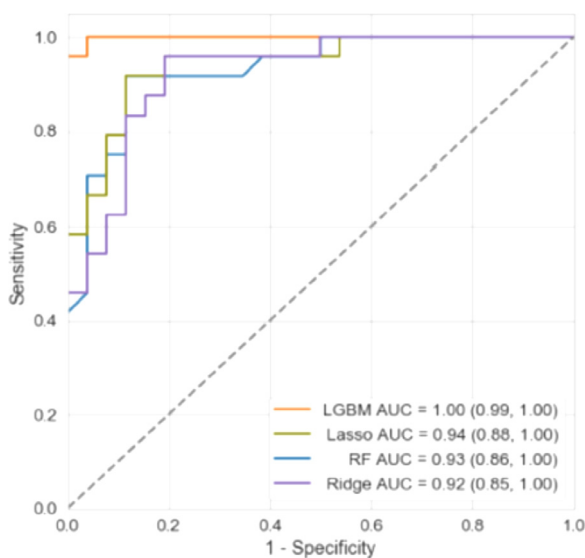


Fig. 2. Area under the receiver operating characteristic curve test performance of the biomarker discovery set. ROC curves comparing the performance of the machine learning models (RF, LGBM) with the traditional linear models (Lasso, Ridge) on the test set; bracketed values are 95% AUC confidence intervals calculated from a normal fit of the curves. AUC: area under the receiver operating characteristic curve; RF: random forests; ROC: receiver operating characteristic curve.

respectively. Subtraction of the background spectral data from the blank VTM replicates did not impact test performance of the model (Fig. S4). At an operating point optimized for sensitivity and Youden's J statistic, LGBM achieved a sensitivity of 1.00 (95% CI 0.86, 1.00) and a specificity of 0.96 (95% CI 0.81, 0.99), superior to other models (Table S1). Subgroup analysis of the performance of the LGBM model on adults and children showed an AUC of 0.99 (95% CI 0.97, 1.00) for adults and an AUC of 1.00 (95% CI 0, 1.00) for children (Table S2). The same model demonstrated an AUC of 1.00 (95% CI 0, 1.00) in immunocompromised hosts, and an AUC of 0.99 (95% CI 0.97, 1.00) in non-immunocompromised hosts (Table S2). Data from the other models including for individuals admitted to the ICU, with bacterial coinfection, antibiotic treatment, and by time since symptom onset are presented in Table S2. Furthermore, a separate multivariable model was performed including the variables age, sex, days since symptom onset and Charlson comorbidity index, and demonstrated evidence that only model outcome was significantly associated with influenza status classification (Table S3).

6.3. Top 20 signature validation by LC/MS-MS maintains high classification performance

After ranking the overall LC/Q-TOF features by importance, we identified the top 20 ion features associated with classification, of which only 13 contributed more than 1% to model predictions (Fig. 3). The top 20 ion feature signature identified by LC/Q-TOF was validated in a cohort of samples from 96 symptomatic individuals with nasopharyngeal swabs including 48 positives (24 influenza A H1N1, 5 influenza A H3 and 19 influenza B) and 48 negatives. Testing was performed by LC/MS-MS using the same sample set. The top 20 ion feature signature revealed an overall AUC of 1.00 (95% CI 0.998,

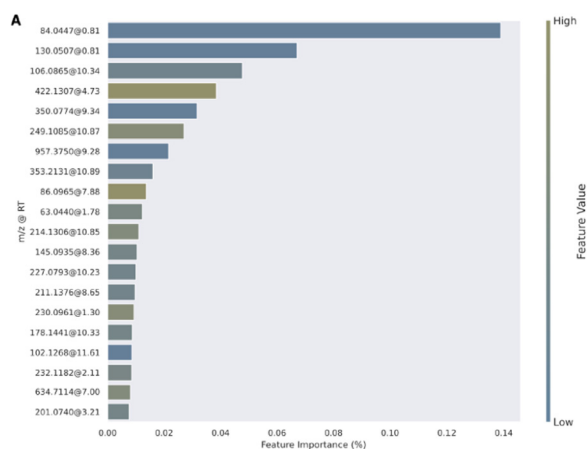


Fig 3. Feature importance analysis by SHapley Additive exPlanation (SHAP) values. Top 20 ion features by percentage importance using the SHAP method. Ion features are identified by mass-to-charge ratio @ retention time, and colors indicate the association between feature value and positive influenza classification. For example, low values of 84.0447@0.81 are indicative of positive classification, while the relative value of 106.0865@10.34 does not have a clear interpretation, despite being an important feature.

1.00), sensitivity of 0.94 (95% CI 0.83, 0.98) and specificity of 1.00 (95% CI 0.93, 1.00) (Fig. 4). High classification performance was maintained with parsimonious signatures including only the top 1, 3, 5 and 7 ion features (Fig. S5). Area under the curve for the top 20 features showed significantly lower pyroglutamic levels in influenza-infected individuals (Fig. S6). Heatmap analysis showed the top 20 ion feature signature varied slightly by influenza subtype compared to the negative subgroup (Fig. S7).

6.4. Pyroglutamic acid identified as top differentiating metabolite

Metabolite identification through in-house library matching revealed a Tier 1 match for compound 130.0507@0.81 as pyroglutamic acid [30], and compound 84.0447@0.81 as an in-source fragment ion of pyroglutamic acid (Table S4). Furthermore, compound 350.0774@9.34 was identified to be consistent with formylmethyl glutathione [31], though this identification will require further confirmation. Further metabolite annotation work will be required for the other metabolites listed as these did not definitively match the

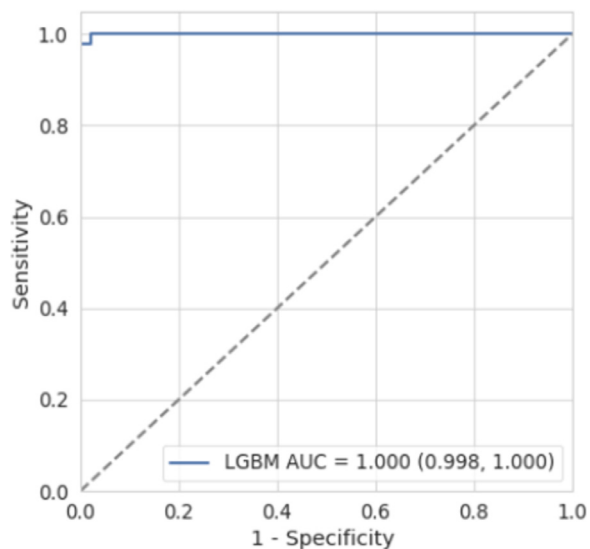


Fig 4. Area under the receiver operating characteristic curve test performance of the validation set. The ROC curve demonstrates the LGBM model performance on the 96-sample validation test set.

in-house library or large database screening (Table S5), and as larger mass error was noted for larger m/z ion features (Fig. S8).

7. Discussion

In this study of 236 nasopharyngeal swab samples from symptomatic individuals, we showed that the described LC/Q-TOF method combined with machine learning could differentiate between influenza-positive (including influenza A 2009 H1N1, H3 and influenza B) and influenza-negative samples with high test performance including AUC, sensitivity and specificity over 0.90. Given this untargeted approach presents significant upfront instrument expense and complexity in data reproducibility and processing, this was followed by a clinically-applicable targeted approach using tandem mass spectrometry (LC/MS-MS). The top 20 ion feature signature identified by LC/Q-TOF was adapted to LC/MS-MS testing on a 96-sample set, and demonstrated sustained high performance. Given LC/MS-MS is already employed in multiple laboratories for routine clinical testing, this work provides a model for feasibility of adaptation and roll-out to other centralized laboratory facilities [32,33]. This is particularly important given the significant burden of respiratory viruses in the U.S. and internationally [34,35]. Indeed, although molecular testing has revolutionized the diagnosis of respiratory viral infections in clinical laboratories, limitations to this technique remain, including high cost, target-specific approach and inability to differentiate active infection from persistent nucleic acid detection [1,36]. Furthermore, the high complexity of many molecular assays limits their use at the point of care where the patient need for a rapid and actionable diagnosis is highest. Similarly, an important gap remains for point-of-care influenza testing due to the lack of sensitivity rapid influenza diagnostic tests and digital immunoassays and resulting inability to confidently rule-out influenza [6]. Metabolomics, or the large-scale study of small molecules, represents the ‘-omics’ technology closest to phenotype and thus holds promise to address current gaps in molecular testing of infectious diseases [37–39]. This metabolomics approach allows for real-time monitoring of host response, uses very little sample volume, may be cost-effective, and allows for hypothesis-free untargeted exploration of novel biomarkers. Furthermore, our finding of a 20-ion feature signature demonstrating reproducible high test performance suggests that these biomarkers may be developed into an assay that could be performed at the point-of-care provided adaptation to a simple diagnostic such as a dipstick lateral flow test is performed.

Given the unique in-line two-column method of the approach presented in this study, comprehensive comparative test performance datapoints for metabolomics applications are lacking. Nonetheless, several published studies have described similar potential applications, mostly from cell culture or animal models [13–16]. However, this approach compared favorably to a previous study using unbiased proteomics from nasopharyngeal lavage sampling with normal saline from 15 previously healthy hosts experimentally infected with influenza A H3N2 or human rhinovirus [12]. The 10-peptide signature from that study was validated in a cohort of 80 subjects, achieving overall AUC of 0.86, sensitivity of 75% and specificity of 97.5% including paired samples. The metabolomics sample processing presented here is simpler and faster than the proteomic workflow (approximately 30 min for ultrafiltration compared to >20 h for proteomics), thus conferring a relative advantage even at similar performance. Previous studies using an untargeted metabolomics approach for the diagnosis of respiratory viruses present important heterogeneity in analytical methods (including MS (LC/Q-TOF, GC-MS) and nuclear magnetic resonance (NMR)), specimen type (including nasopharyngeal aspirate, serum, urine, cell culture), hosts (animal and human), viruses (including influenza, RSV, human rhinovirus) and metabolic signatures (ranging from 10 to 285 metabolites) [8,9,14,40]. These studies profiled metabolites and metabolic pathways but did not include quantitative analytical results of

classification model performance thus limiting assessment of potential clinical utility as diagnostic assays.

The top 20 ion features retained in our biomarker signature likely represent a heterogeneous group of compounds from a variety of biological pathways. The top two ion features were successfully identified through in-house library matching as pyroglutamic acid (130.0507@0.81) and an in-source fragment ion of pyroglutamic acid (compound 84.0447@0.81), which are decreased in specimens from influenza-infected individuals. Pyroglutamic acid (synonyms: pidolic acid, 5-oxoproline) is a cyclized derivative of L-glutamic acid which can form in one of three ways in the living cell: from the degradation of glutathione, from incomplete reactions following glutamate activation, or from the degradation of proteins containing pyroglutamic acid at the N-terminus [41]. Several recent studies have highlighted the complex interaction between glutathione metabolism important in reactive oxygen species (ROS) regulation and infection with influenza, which is known to increase the formation of ROS [42–45]. In a study using ultra-high-pressure LC/Q-TOF to detect early metabolic disturbances following infection with influenza H1N1 in A549 human lung epithelial cells, significant differences were found in 50 metabolites which were mainly mapped to purine, glutathione and lipid metabolism pathways [14]. In the reference study, the infected A549 cells were washed and lysed prior to metabolite analysis, and showed upregulation of glutathione metabolism with an increase in the intracellular concentration of pyroglutamic acid. Our results show a decrease in pyroglutamic acid in NP swabs from influenza-infected individuals. Given our specimens are not washed or lysed, the observed decrease in pyroglutamic acid in NP swabs from infected individuals may be due to decreased extracellular concentrations from increased use of glutathione in the intracellular space. Alternatively, a more complex mechanism involving oxidative stress and upstream metabolic effects may be at play. Though the mechanism giving rise to differential concentrations of pyroglutamic acid in our specimens is not yet known, our results conform to the findings in the current literature which highlight glutathione metabolism as a key pathway altered during influenza infection. In addition, the detected pyroglutamic acid was not identified to be an in-source fragment of glutamate, further supporting its independent role. In addition, although annotation was only preliminary for the additional features, an important characteristic noted was that several features demonstrated a closest match to terpenoids. The presence of non-routine compounds in human samples will require further investigation.

In this study, both statistical models and machine learning models were explored comprehensively to assess for best test performance for these untargeted metabolomics data. These results were reproducible across datasets and across models, adding confidence to our findings. Furthermore, the machine learning models were observed to consistently outperform the statistical models, consistent with findings in previous studies [46,47].

This study presents several strengths. First, it demonstrated high test performance in the discovery cohort, which was independently validated in a distinct cohort of consecutive clinical specimens, supporting the reproducibility, robustness and lack of overfitting of this approach. Furthermore, high performance on the clinically-applicable tandem mass spectrometry testing may facilitate uptake by a large number of laboratories, alleviating the need for complex testing by LC/Q-TOF, and enabling cost-effective testing. Second, it demonstrated a large effect size from a limited number of compounds in the SHAP feature importance analysis. This increases the feasibility of adapting this diagnostic approach to a point-of-care device such as portable mass spectrometry, though further work will be required to determine the optimal number of biomarkers required for this purpose. Third, this study was based on a real-world, diverse patient population of individuals who were naturally infected with influenza, which may better approximate metabolic changes compared to experimentally-infected, previously healthy volunteers. Furthermore,

cases and controls in the discovery cohort were tightly age- and sex-matched, thus reducing potential confounders in metabolomic analysis due to up- or downregulation of certain metabolic pathways based on these host factors [48]. Fourth, this cohort included a large number of samples, conferring over 90% power to detect a difference between influenza-infected and uninfected individuals. Finally, we proceeded with a systematic and comprehensive bioinformatics pipeline analysis strategy to identify the best model for untargeted and targeted metabolomics data.

This study also presents limitations. First, this study was performed at a single institution only and it is unclear at present if results are generalizable to other patient populations. However, our finding of consistent results across diverse patient groups lends support to the potential generalizability of this diagnostic approach. Second, only influenza positive and negative samples were compared in the untargeted approach such that we could not extrapolate to other respiratory viruses, and bacterial or viral co-infections. However, limited coinfection data in the validation cohort supported maintained performance. Further study will be important to better understand changes that occur across the spectrum of nasopharyngeal microbiome including bacterial colonization or coinfection, and to incorporate comparisons with other important respiratory viruses such as RSV and parainfluenza to better rival with current molecular diagnostic methods. Third, this study did not assess the respiratory metabolic profiles of healthy individuals as negative controls, which may help further isolate the metabolites that change in response to acute viral infection. Furthermore, sample adequacy was not assessed due to the proprietary nature of the internal control included in the commercial respiratory pathogen panel used for clinical testing. Fourth, we did not perform repeat longitudinal samples in the same individuals, and did not include paired plasma or urine samples, which would have strengthened findings of identified metabolites if reproducible. Fifth, due to a large proportion of features that could not be matched for identification, quantitation work to validate the full 20-ion feature signature could not be performed. Significant work remains to fully annotate these features, and for clinical adaptation using quantitative thresholds; this will require complementary mass spectrometry techniques, and fresh influenza samples which have not been available over the 2020–2021 season owing to the absence of influenza cases in our setting in the context of coronavirus disease (COVID-19). Finally, VTM contains small molecules that may have confounded the analysis. However, subtraction of background spectral data from the blank VTM sample replicates did not impact test performance of our model, suggesting these data did not significantly contribute to model classification.

In summary, we demonstrated the feasibility and high accuracy of an untargeted metabolomics approach from nasopharyngeal samples for the identification of distinct metabolic signatures for the diagnosis of influenza infection. This approach maintained high performance after adaptation to clinically-adaptable LC/MS-MS instruments. Significant work remains to be done to leverage the full potential of this method including expansion to other patient settings and in larger cohorts, additional pathogens and sample types, and to prospectively assess its potential as a prognostic tool. In addition, this method could be used to explore metabolic pathways that could eventually be harnessed for therapeutic potential.

Contributors

Conceptualisation, C.A.H., P.R., A.T.L. and B.A.P.; Data curation: P.S.R., H.S., N.P., C.A.H., A.T.L., M.W., Methodology, A.T.L., J.M.; Software, P.S.R., H.S., N.P.; Validation, J.M., C.A.H., A.T.L., M.W.; Formal analysis, P.S.R., H.S., N.P., C.A.H., M.K.S.; Investigation, C.A.H., A.T.L., M.U., N.G., J.M.; Resources, B.A.P., T.M.C.; Writing – original draft, C.A.H., P.R., H.S., N.P., A.T.L., M.W.; Writing – review and editing, C.A.H., P.R., H.S., N.P., A.T.L., J.M., B.A.P., T.M.C., M.W.; Supervision: B.A.P., J.M., T.M.C., A.

N.; Funding acquisition: B.A.P., M.W., T.M.C. All authors have read and approved the final version of the manuscript.

Data sharing statement

Data collected for this study will be made available through github. This will include de-identified participant data, and metabolomics results. Starting at publication, data and code will be available on github at the following addresses: <https://github.com/stanfordmlgroup/influenza-qtof>. Furthermore, raw metabolomics data will be available on MetaboLights. Signed data access agreement will be required.

Declaration of Competing Interest

A provisional patent covering the metabolomics approach combined with machine learning to recognize a medical condition has been filed (C.A.H., P.R., A.T.L., T.M.C., B.P.). The authors declare no other competing interests.

Acknowledgments

This project was partially supported a Precision Health Award from the Stanford Department of Pathology (M.W.) (<https://med.stanford.edu/pathology.html>), by anonymous gift funding to the Stanford Clinical Virology Laboratory, and by a R01HD102537 grant from the National Institute of Child Health and Human Development (T.M.C and A.T.L.).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ebiom.2021.103546](https://doi.org/10.1016/j.ebiom.2021.103546).

References

- Somerville LK, Ratnamohan VM, Dwyer DE, Kok J. Molecular diagnosis of respiratory viruses. *Pathology* 2015;47(3):243–9.
- Schreckenberger PC, McAdam AJ. Point-counterpoint: large multiplex PCR panels should be first-line tests for detection of respiratory and intestinal pathogens. *J Clin Microbiol* 2015;53(10):3110–5.
- Vergara A, Cilloniz C, Luque N, et al. Detection of human cytomegalovirus in bronchoalveolar lavage of intensive care unit patients. *Eur Respir J* 2018;51(2):1701332. doi: 10.1183/13993003.01332-2017.
- Tan SK, Burgener EB, Waggoner JJ, et al. Molecular and culture-based bronchoalveolar lavage fluid testing for the diagnosis of cytomegalovirus pneumonitis. *Open Forum Infect Dis* 2016;3(1):ofv212. doi: 10.1093/ofid/ofv212.
- Buchan BW, Ledeboer NA. Emerging technologies for the clinical microbiology laboratory. *Clin Microbiol Rev* 2014;27(4):783–822.
- Merckx J, Wali R, Schiller I, et al. Diagnostic accuracy of novel and traditional rapid tests for influenza infection compared with reverse transcriptase polymerase chain reaction: a systematic review and meta-analysis. *Ann Intern Med* 2017;167(6):394–409.
- Estimated influenza illnesses, medical visits, hospitalizations, and deaths in the United States – 2018–2019 influenza season. 2019.
- Ferrarini A, Righetti L, Martinez MP, et al. Discriminant biomarkers of acute respiratory distress syndrome associated to H1N1 influenza identified by metabolomics HPLC-QTOF-MS/MS platform. *Electrophoresis* 2017;38(18):2341–8.
- Stewart CJ, Hasegawa K, Wong MC, et al. Respiratory syncytial virus and rhinovirus bronchiolitis are associated with distinct metabolic pathways. *J Infect Dis* 2018;217(7):1160–9.
- Antonelli G. Emerging new technologies in clinical virology. *Clin Microbiol Infect* 2013;19(1):8–9.
- Mancone C, Ciccocanti F, Montaldo C, et al. Applying proteomic technology to clinical virology. *Clin Microbiol Infect* 2013;19(1):23–8.
- Burke TW, Henao R, Soderblom E, et al. Nasopharyngeal protein biomarkers of acute respiratory virus infection. *EBioMedicine* 2017;17:172–81.
- Purcaro G, Rees CA, Wieland-Alter WF, et al. Volatile fingerprinting of human respiratory viruses from cell culture. *J Breath Res* 2018;12(2):026015.
- Tian X, Zhang K, Min J, et al. Metabolomic analysis of influenza A virus A/WSN/1933 (H1N1) infected A549 cells during first cycle of viral replication. *Viruses* 2019;11(11):1007. doi: 10.3390/v11111007.
- Beale DJ, Oh DY, Karpe AV, et al. Untargeted metabolomics analysis of the upper respiratory tract of ferrets following influenza A virus infection and oseltamivir treatment. *Metabolomics* 2019;15(3):33.
- Lin YT, Lin CF, Yeh TH. Influenza A virus infection induces indoleamine 2,3-dioxygenase (IDO) expression and modulates subsequent inflammatory mediators in nasal epithelial cells. *Acta Otolaryngol* 2020;140(2):149–56.
- Nalbantoglu S. *Metabolomics: basic principles and strategies*. Molecular Medicine. IntechOpen; 2019.
- Pacchiarotta T, Deelder AM, Mayboroda OA. Metabolomic investigations of human infections. *Bioanalysis* 2012;4(8):919–25.
- Zurfluh S, Baumgartner T, Meier MA, et al. The role of metabolomic markers for patients with infectious diseases: implications for risk stratification and therapeutic modulation. *Expert Rev Anti Infect Ther* 2018;16(2):133–42.
- Le A, Mak J, Cowan TM. Metabolic profiling by reversed-phase/ion-exchange mass spectrometry. *J Chromatogr B Anal Technol Biomed Life Sci* 2020;1143:122072.
- Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res* 2018;46(D1):D608–17. doi: 10.1093/nar/gkx1089.
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28(1):27–30.
- Mak J, Cowan TM, Le A. Quantitative analysis of underivatized amino acids by liquid chromatography-tandem mass spectrometry. *Methods Mol Biol* 2019;2030:85–109.
- Le A, Ng A, Kwan T, Cusmano-Ozog K, Cowan TM. A rapid, sensitive method for quantitative analysis of underivatized amino acids by liquid chromatography-tandem mass spectrometry (LC-MS/MS). *J Chromatogr B Anal Technol Biomed Life Sci* 2014;944:166–74.
- Ke GMQ, Finely T, Wang T, Chen W, Ma W, Ye Q, Liu T. LightGBM: a highly efficient gradient boosting decision tree. *Neural Inf Process Syst Found* 2017;30:3146–54.
- Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* 2018;2(10):749–60.
- Lundberg SM, Lee S-I (2017). A Unified Approach to Interpreting Model Predictions, in *I Guyon; UV Luxburg; S Bengio; H Wallach; R Fergus; S Vishwanathan & R Garnett, ed., 'Advances in Neural Information Processing Systems 30'*, Curran Associates, Inc., pp. 4765–4774.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44(3):837–45.
- Pedregosa FVG, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay ES. Machine learning in python. *J Mach Learn Res* 2011;12:2825–30.
- ChemSpider. L-Pyroglytamic acid. (Accessed June 5 2021).
- ChemSpider. S-(Formylmethyl)glutathione. (Accessed June 5 2021).
- Seeger C, Salzmann L. After another decade: LC-MS/MS became routine in clinical diagnostics. *Clin Biochem* 2020;82:2–11.
- Garg U, Zhang YV. Mass spectrometry in clinical laboratory: applications in therapeutic drug monitoring and toxicology. *Methods Mol Biol* 2016;1383:1–10.
- Wang X, Li Y, O'Brien KL, et al. Global burden of respiratory infections associated with seasonal influenza in children under 5 years in 2018: a systematic review and modelling study. *Lancet Glob Health* 2020;8(4):e497–510.
- GBD 2016 Lower Respiratory Infections Collaborators. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the global burden of disease study 2016. *Lancet Infect Dis* 2018;18(11):1191–210.
- Whitley DM, Bialasiewicz S, Bletchly C, et al. Detection of novel influenza A(H1N1) virus by real-time RT-PCR. *J Clin Virol* 2009;45(3):203–4.
- Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* 2016;17(7):451–9.
- Patti GJ, Yanes O, Siuzdak G. Innovation: metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* 2012;13(4):263–9.
- Fiehn O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol* 2002;48(1–2):155–71.
- Turi KN, Romick-Rosendale L, Gebretsadik T, et al. Using urine metabolomics to understand the pathogenesis of infant respiratory syncytial virus (RSV) infection and its role in childhood wheezing. *Metabolomics* 2018;14(10):135.
- Kumar AB, Bachhawat AK. Pyroglutamic acid: throwing light on a lightly studied metabolite. *Curr Sci* 2012;102(2):288–97.
- Keshavarz M, Solaymani-Mohammadi F, Namdari H, Arjeini Y, Mousavi MJ, Rezaei F. Metabolic host response and therapeutic approaches to influenza infection. *Cell Mol Biol Lett* 2020;25:15.
- Amatore D, Celestino I, Brundu S, et al. Glutathione increase by the n-butanoyl glutathione derivative (GSH-C4) inhibits viral replication and induces a predominant Th1 immune profile in old mice infected with influenza virus. *FASEB Bioadv* 2019;1(5):296–305.
- Nencioni L, Iuvara A, Aquilano K, et al. Influenza A virus replication is dependent on an antioxidant pathway that involves GSH and Bcl-2. *FASEB J* 2003;17(6):758–60.
- Cai J, Chen Y, Seth S, Furukawa S, Compans RW, Jones DP. Inhibition of influenza infection by glutathione. *Free Radic Biol Med* 2003;34(7):928–36.
- Li Y, Wu FX, Ngom A. A review on machine learning principles for multi-view biological data integration. *Brief Bioinform* 2018;19(2):325–40.
- Trainor PJ, DeFilippis AP, Rai SN. Evaluation of classifier performance for multi-class phenotype discrimination in untargeted metabolomics. *Metabolites* 2017;7(2):30. doi: 10.3390/metabo7020030.
- Srivastava S. Emerging insights into the metabolic alterations in aging using metabolomics. *Metabolites* 2019;9(12):301. doi: 10.3390/metabo9120301.