

RESEARCH

Open Access



Dual RNA-seq reveals viral infections in asthmatic children without respiratory illness which are associated with changes in the airway transcriptome

Agata Wesolowska-Andersen¹, Jamie L. Everman¹, Rebecca Davidson¹, Cydney Rios¹, Rachelle Herrin¹, Celeste Eng², William J. Janssen³, Andrew H. Liu^{4,5}, Sam S. Oh², Rajesh Kumar⁶, Tasha E. Fingerlin^{1,7}, Jose Rodriguez-Santana⁸, Esteban G. Burchard^{2,9} and Max A. Seibold^{1,4,10*}

Abstract

Background: Respiratory illness caused by viral infection is associated with the development and exacerbation of childhood asthma. Little is known about the effects of respiratory viral infections in the absence of illness. Using quantitative PCR (qPCR) for common respiratory viruses and for two genes known to be highly upregulated in viral infections (*CCL8/CXCL11*), we screened 92 asthmatic and 69 healthy children without illness for respiratory virus infections.

Results: We found 21 viral qPCR-positive and 2 suspected virus-infected subjects with high expression of *CCL8/CXCL11*. We applied a dual RNA-seq workflow to these subjects, together with 25 viral qPCR-negative subjects, to compare qPCR with sequencing-based virus detection and to generate the airway transcriptome for analysis. RNA-seq virus detection achieved 86% sensitivity when compared to qPCR-based screening. We detected additional respiratory viruses in the two *CCL8/CXCL11*-high subjects and in two of the qPCR-negative subjects. Viral read counts varied widely and were used to stratify subjects into Virus-High and Virus-Low groups. Examination of the host airway transcriptome found that the Virus-High group was characterized by immune cell airway infiltration, downregulation of cilia genes, and dampening of type 2 inflammation. Even the Virus-Low group was differentiated from the No-Virus group by 100 genes, some involved in eIF2 signaling.

Conclusions: Respiratory virus infection without illness is not innocuous but may determine the airway function of these subjects by driving immune cell airway infiltration, cellular remodeling, and alteration of asthmogenic gene expression.

Keywords: Asymptomatic, Virus, Infection, Transcriptome, Airway epithelium, RNA-seq, Asthma, Children, Host response, Host-virus interactions

* Correspondence: seiboldm@njhealth.org

¹Center for Genes, Environment, and Health, National Jewish Health, Denver, CO, USA

⁴Department of Pediatrics, National Jewish Health, 1400 Jackson St, Denver, CO 80206, USA

Full list of author information is available at the end of the article



Background

The detection of viruses from human biological samples and the determination of their influence on the host cell transcriptome are paramount for understanding complex human diseases that are initiated or exacerbated by virus infections and their sequelae. Even virus detection alone has proven challenging for current methods. First, PCR-based viral amplicon detection strategies are susceptible to cross-reactivity, lack strain specificity, and are inefficient for broad screening due to the lack of a common gene shared across virus species [1–3]. Solution-based capture methods followed by fluorescent detection or sequencing have generated quality data but can be confounded by sequence differences between viral species/strains and can only detect viruses for which the capture panel is designed [4, 5]. A third method involves the physical isolation of virions from clinical samples followed by sequencing to determine viral species present [6–9]. However, physical separation methods (e.g., ultra-centrifugation and ultra-filtration) are cumbersome to apply to multiple samples, and the following sequence-based detection may be unsuccessful when the sample viral load is low.

Another strategy is to perform shotgun sequencing of nucleic acids isolated from the entire, unseparated clinical sample (i.e., metagenomics). This method has been previously applied to total nucleic acids (RNA and DNA) extracted from clinical samples for viral detection [10–13]. Although metagenomic sequencing of total nucleic acids can detect viruses, its feasibility is highly dependent on the level of virus present in the sample and the amount of host cell nucleic acid present, which can in turn demand high levels of sequence depth for virus detection. Moreover, since the sequencing library contains reads from host DNA, the host transcriptome cannot be analyzed. An alternative is to isolate and specifically sequence RNA isolated from a clinical sample. The sequencing reads generated from this RNA sample can then be mapped to the human genome and analyzed to establish the host transcriptome profile, and unmapped reads can be queried against viral databases to establish the presence of both RNA and DNA viruses present in the sample through their transcriptional activity. For viruses with RNA genomes (the majority of respiratory viruses), the viral reads detected with this method can also originate from the virus genome in addition to active viral transcription. Total viral read counts generated, as a reflection of both infection level and transcriptional activity, can be tested for association with the host transcriptome data from the same sample. For multi-transcript viruses the read counts for specific viral transcripts could also be related to host transcript levels. This dual purposing of a single sequencing library is not only efficient, but powerful in that a level of

internal control is achieved by use of a common sample, library, and sequencing reaction to generate both host and virus data. Such dual RNA-seq approaches have recently been applied to interrogate host-viral interactions both *in vitro* [14] and *in vivo* [15, 16], aiding greatly in the deconvolution of viral-driven disease mechanisms. In particular, the work of Perez-Losada et al., 2015 [16] established the first dual RNA-seq study in the context of asthma, focusing on differences in the host-bacterial microbiome interactions between asthmatic and healthy children.

Upper respiratory tract infections are among the most prevalent of all viral infections in humans and render a dramatic cost on the health care system in terms of lost economic productivity and general well-being. Beyond their acute effects, respiratory virus infections are strongly associated with the development, symptoms, and exacerbation of asthma [17, 18]. For example, an early-life infection with respiratory syncytial virus (RSV), which induces respiratory bronchiolitis, is associated with a sixfold increase in asthma risk by age 6 years [19]. Early-life rhinovirus-associated wheezing has an even stronger association with subsequent asthma by age 6 and 13 years [20, 21]. Rhinovirus infections are also associated with >80% of childhood asthma exacerbations [22], and recently rhinovirus detection was associated with increased day-to-day asthma symptoms throughout the year while, at times, dissociated from nasal symptoms and exacerbations [23, 24]. Despite the dramatic increase in risk associated with respiratory tract infections, it is also clear that most people affected by respiratory viruses are resilient to development of asthma or a significant exacerbation. These facts suggest a complex interaction between the host and virus, which is important for determining risk. This interaction first occurs at the level of the nasal airway epithelium, which is the primary site of infection and replication for respiratory viruses. The response of the airway epithelium to viral infection is complex and involves both the mucociliary epithelial and interdigitating immune cells that form the mucosal surface. This intricate cellular milieu is difficult to recreate for *in vitro* experimental infection studies; therefore, novel methods are needed to capture this biology in the *in vivo* setting.

In this study, we detail methods for the detection of common respiratory viruses in asthmatic and healthy children using polyA-selected whole transcriptome sequencing. Using these efficient methods, we found that a large proportion of children without respiratory illness were carrying pathogenic and transcriptionally active respiratory viruses. Virus detection using a dual RNA-seq workflow resulted in 86% sensitivity when compared to quantitative PCR (qPCR)-based screening. Additionally, we detected respiratory viruses not targeted by viral

qPCR primers in two subjects with high expression of viral response genes, demonstrating the versatility of this method. Despite absence of respiratory illness, viral carriers exhibited a dramatically altered nasal epithelial transcriptome, reflective of active viral infection. Virus levels determined by RNA-seq correlated strongly with host airway viral response gene expression levels. Even virus levels at the limits of detection by qPCR and RNA sequencing had a discernible effect on the host airway epithelial transcriptome. We detail modules of activated gene expression in viral carriers reflective of epithelial cell subtypes, as well as immune cell types. We also report modules of asthmogenic genes that are strongly associated with viral infection. We conclude that viral infection, even in the absence of observable illness, induces epithelial responses in children with and without asthma.

Results

Quantitative PCR screening for respiratory virus infections in 161 subjects without respiratory illness

We first used qPCR assays for six common respiratory virus species to identify which of the 161 Genes-environments and Admixture in Latino Americans II (GALA II) cohort subjects (92 asthmatics and 69 healthy controls) were infected with a respiratory virus, despite the absence of respiratory illness (Additional file 1: Table S1, Additional file 2: Figure S1). Specifically, the qPCR assays tested for human rhinovirus (HRV), respiratory syncytial virus (RSV), human metapneumovirus (HMPV), human parainfluenza virus 1 (HPIV1), human parainfluenza virus 2 (HPIV2), and human parainfluenza virus 3 (HPIV3) [25]. We identified 21 subjects harboring respiratory viruses (9 asthmatics, 12 controls): 14 HRV, 4 RSV, 3 HPIV1, and 1 HMPV (1 subject was qPCR-positive for 2 viruses) (Table 1). To screen for additional subjects carrying other viral species, we performed a qPCR assay on the subject RNA for expression of two putative viral biomarker genes. These genes, *CCL8* and *CXCL11*, were previously reported as among the most upregulated during the peak stage of acute respiratory illness [26]. Our qPCR screen found that most subjects exhibited low expression for both of these genes, and that subjects with high expression of these genes were mostly qPCR positive for viruses (Fig. 1, Table 1). Moreover, among the virus-positive subjects, we observed high correlation between both the *CCL8* and *CXCL11* gene expression and virus levels ($\rho_{CCL8} = 0.84$, $\rho_{CXCL11} = 0.80$). Two subjects that were ranked in the top fifth percentile of expression for both biomarker genes were negative for virus by qPCR, and were therefore suspected to be carriers of viral species not covered by the qPCR panel (Fig. 1). There was no significant difference in viral carriage rate observed between asthmatics and healthy controls.

Respiratory virus detection in a host-dominated transcriptome

We next generated polyA-selected whole transcriptome RNA-seq data on the 21 qPCR virus-positive subjects, 2 *CCL8/CXCL11*-high but viral qPCR-negative subjects, and 25 randomly selected viral qPCR-negative subjects. We applied an adapted Kapa mRNA-seq protocol to generate the transcriptome libraries from these 48 samples (Fig. 2a). The samples were sequenced to a mean read depth of $1.26 \times 10^7 \pm 2.71 \times 10^6$ using the Ion Torrent Proton sequencer. This sequencing depth was chosen based on what was previously determined to yield informative airway epithelial transcriptome data [27]. We analyzed these data using a pipeline of publicly available bioinformatic tools to detect and quantify respiratory virus reads in the samples, in addition to host transcriptome generation (Fig. 2b). Across samples we found that 3.5–7.0% of reads failed to map with high stringency to the human genome. We therefore attempted a second low-stringency alignment of these unmapped reads to the human genome using a different aligner, which resulted in an average of 87.5% ($\pm 6.4\%$) of these additional reads being mapped. To identify reads of viral origin, we performed a Basic Local Alignment Search Tool (BLAST) search of unmapped reads against the National Center for Biotechnology Information (NCBI) Nucleotide database. We found respiratory virus reads in 22 of 48 subjects sequenced. The viral reads were assembled into longer contigs to enable more accurate taxonomic assignment of the read source. All reads not mapping to the human genome for these samples were subsequently mapped to their respective contig-matching virus genome. Only a single virus species was detected in each sample, and the number of viral reads varied significantly, from as little as 4 reads ($2.29 \times 10^{-5}\%$ of the total reads) to 164,870 reads (1.54% of the total reads) generated per sample. Viral genome read depth plots were generated for all virus-infected samples with >20 viral reads (Fig. 3, Additional file 2: Figure S2). For all but one of these viral carriers, we observed coverage spread throughout the length of the detected viral genome (Fig. 3a, b), providing us with confidence that the virus-matching reads were not sequencing or mapping artifacts. In the sample with uneven viral read distribution (Fig. 3c), Control-7, nearly all HCoV reads mapped to the virus nucleocapsid gene. Interestingly, viral genome read depth plots for multi-transcript viruses, such as parainfluenza and RSV, revealed breaks in the read depth across the genome aligned with viral transcripts, indicating that the reads originated from viral transcription and not simply from viral RNA genomes (Fig. 3a).

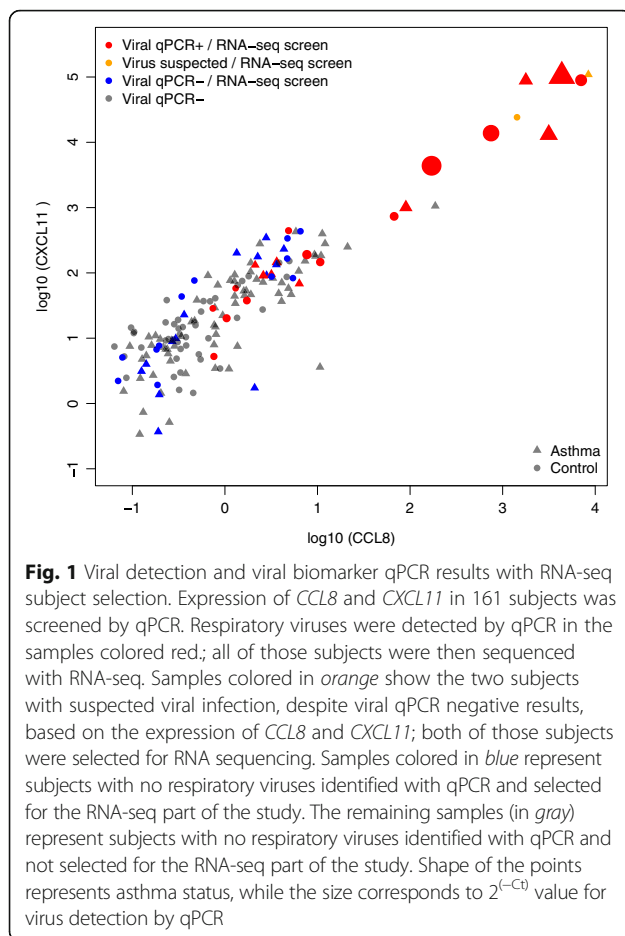
RNA-seq versus qPCR virus detection

In total, respiratory virus reads were detected in 18 of 21 qPCR virus-positive samples, resulting in a sensitivity of

Table 1 Amplification and mapping metrics for samples with respiratory viruses detected by RNA-seq and/or by qPCR. Detected virus coverage was scaled by the total number of raw reads divided by 10e6. If the closest reference differed between the multiple contigs for a sample, then the most frequent closest reference sequence was reported. In all cases the closest reference for all contigs of a sample were of the same viral species (i.e., no dual infections were detected)

Subject	qPCR		RNA-seq mapping metrics					Detected virus				
	Virus Species	Ct	Gene		Total reads	Resp. virus reads	% Viral reads	Species	Strain/serotype	Closest reference (NCBI acc)	Virus mapped to NCBI reference	Coverage (scaled)
			CCL8	CXCL11								
Asthma-1	None	NA	8426.9	109137.3	10,720,551	164,870	1.54	HPIV	4a	KF483663.1	98.9%	1271.6x
Control-1	HRV	29.36	7.7	190.8	9,818,450	74,935	0.763	HEV	C105	JX393302.1	92.8%	1259.7x
Asthma-2	RSV	20.05	4356.0	103,107.1	11,597,132	58,263	0.502	RSV	A	KJ672478.1	99.8%	473.5x
Control-2	HPIV1	24.37	749.7	13819.3	14,029,017	26,045	0.186	HPIV	1	KF687314.1	99.8%	172.2x
Control-3	HRV	24.19	170.7	4377.0	10,591,814	4081	0.039	HRV	A55	DQ473511.1	80.8%	66.9x
Asthma-3	HPIV1	25.29	3148.7	13,129.6	9,738,089	6648	0.068	HPIV	1	KF530212.1	99.8%	64.4x
Control-4	HRV	25.86	7042.3	89477.7	16,450,743	2878	0.017	HRV	A59	JN541266.1	99.3%	37.9x
Asthma-4	HRV	38.15	6.4	68.5	15,103,156	829	0.005	HRV	C	JF317015.1	99.9%	11.7x
Control-5	None	NA	1430.1	24,191.0	14,863,334	2671	0.018	Influenza	B	CY171998.1	5.7%	7.8x
Control-6	HRV	30.24	67.3	734.7	9,732,129	203	0.002	HRV	A59	JN541266.1	99.5%	4.5x
Asthma-5	HRV	29.17	90.1	1010.9	8,269,118	156	0.002	HRV	A100	FJ445175.1	91.0%	3.6x
Control-7	None	NA	4.7	166.0	10,579,380	780	0.007	HCoV	OC43	KF530099.1	95.0%	3.2x
Asthma-6	HRV	28.60	1780.1	89078.1	9,944,802	97	0.001	HRV	A55	DQ473511.1	85.6%	1.7x
Control-8	HMPV	30.78	10.7	146.7	9,978,447	107	0.001	HMPV	A2	GQ153651.1	99.1%	1.2x
Control-9	HRV	40.98	0.7	28.6	14,107,522	76	0.0005	HRV	C	JF317015.1	100%	1.2x
Control-10	HRV	44.08	1.3	58.1	14,005,041	39	0.0003	HEV	CY135	KF322116.1	15.4%	0.7x
Control-11	HRV	33.35	1.0	20.2	15,787,336	35	0.0002	RSV	A	EF155421.1	100%	0.5x
Control-12	None	NA	0.5	76.5	10,880,596	24	0.0002	RSV	A	KJ672484.1	100%	0.2x
Asthma-7	HRV	37.60	2.6	90.8	16,883,056	9	5.33e-05	HCoV	NA*	JX504050.1	66.7%	NA*
Asthma-8	RSV	35.95	3.6	145.9	15,782,330	7	4.44e-05	RSV	NA*	KJ672484.1	85.7%	NA*
Asthma-9	HRV	42.14	2.1	132.1	13,910,896	5	3.59e-05	HRV	NA*	KC342104.1	40%	NA*
Control-13	HRV	33.51	1.7	37.8	17,449,199	4	2.29e-05	HRV	NA*	AB904651.1	25%	NA*
Control-14	HRV	37.76	0.8	5.3	14,074,760	0	0	NA	NA	NA	NA	NA
Asthma-10	RSV/HPIV1	40.04/37.25	3.2	95.6	15,013,585	0	0	NA	NA	NA	NA	NA
Control-15	RSV	37.8	4.9	443.7	17,496,674	0	0	NA	NA	NA	NA	NA

*Too few reads to perform assembly into contigs

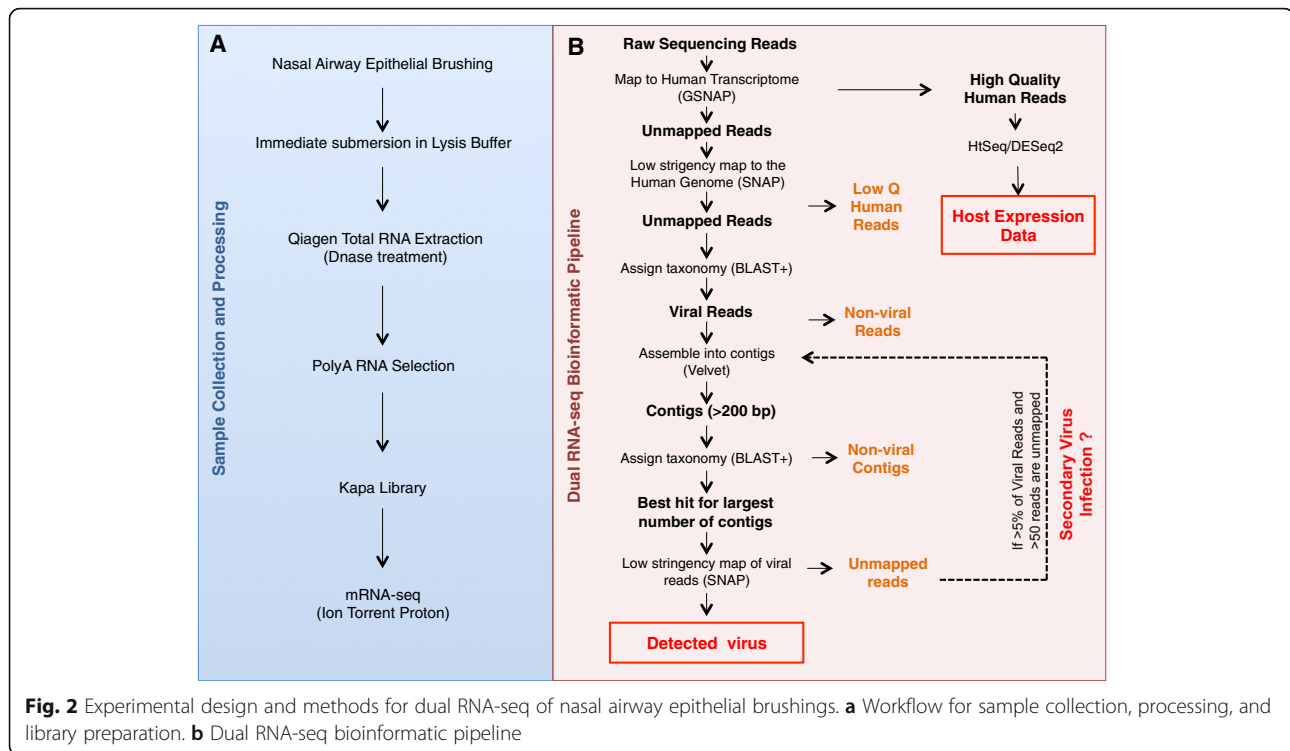


86% using RNA-seq at this depth (Table 1). The three samples with viruses not detected by RNA-seq all had viral qPCR Ct values at the edge of detection ($Ct > 37$, Table 1). The respiratory viruses detected by RNA-seq matched the viruses identified with qPCR in 15 out of 18 cases. In the three discordant cases, the viral species identified with qPCR was HRV; however, RNA-seq-based screening identified human enterovirus for two of the subjects and human coronavirus for the third. Enterovirus and HRV are part of the same viral genus and share high sequence similarity [28], which is likely why the HRV primers gave a positive signal in the enterovirus-infected samples. Careful examination of the third sample confirmed all reads mapped to coronavirus, confirming these results. Although HRV reads were not detected in this sample, the level of HRV infection detected in this sample was very low ($Ct > 37$). We found that 23 of the 25 qPCR virus-negative subjects were RNA-seq-negative for respiratory viruses. One of the RNA-seq virus-positive samples, Control-12, exhibited 24 RSV reads. The other sample, Control-7, exhibited 780 reads matching an unscreened human coronavirus (HCoV).

Examining the two qPCR-negative, but suspected virus carriers based on nasal airway epithelial viral biomarker expression, we detected human parainfluenza virus 4a (HPIV-4a) in one sample and influenza virus B in the other sample. Neither viral species was tested in our qPCR assays. Examining all samples with virus detected by both methods, there was a strong positive correlation between qPCR viral signal and the proportion of viral reads in RNA-seq data ($\rho = 0.75$, $P = 3 \times 10^{-3}$). Therefore, RNA-seq viral detection at this low read depth is highly quantitative, is nearly as sensitive as qPCR viral assay, and has likely unlimited breadth of detection. Examining all subjects either qPCR or RNA-seq positive for virus, we observed a skew in the frequency of virus-infected subjects by season ($P = 0.042$). This skew was driven by a significant decrease in viral infection frequency in summer (virus positive 4%, virus negative 23%, $P = 0.029$) and a trend toward increased infection in the fall (virus positive 24%, virus negative 10%, $P = 0.092$).

Respiratory virus infection without respiratory illness drives a dramatic shift in the host transcriptome

We then examined whether the host airway epithelial transcriptome generated from the same sequencing libraries was reflective of viral infection status. Due to the wide variation in the number of viral reads detected, we divided the viral carriers into two groups based on RNA-seq virus detection: (1) Virus-High subjects: with viral genome coverage of $> = 1x$ ($n = 14$) and (2) Virus-Low: remaining subjects harboring a respiratory virus ($n = 10$) (Table 1). Given the near monogenic read pileup for subject Control-7, we did not include this subject in any of the downstream analyses. The 23 subjects with no respiratory virus detected by either qPCR or RNA-seq comprised the No-Virus group. We first performed multidimensional scaling (MDS) of the airway epithelial transcriptome data (Fig. 4, Additional file 1: Table S2). The first dimension generated from this analysis accounted for 36.9% of total transcriptome variance. It completely separated the Virus-High subjects from No-Virus subjects. The Virus-Low subjects clustered in between the Virus-High and No-Virus subjects; however, they were not completely distinct from the No-Virus subjects. Dimension 1 values correlated with observed viral depth ($\rho = 0.47$). We therefore found the stratification of samples by high versus low viral genome coverage depth to be strongly associated with the intensity of the airway expression response. To identify specific genes associated with virus carriage without illness, we performed single gene differential expression analysis between the Virus-High and No-Virus samples. We found 8126 differentially expressed genes (false discovery rate, FDR 5%), including 4061 upregulated and 4065 downregulated genes (Additional file 1: Table S3).



Among the top genes we observed host virus response genes (*RAD2*, *OASL*, etc.) and many interferon-induced genes (*ISG15*, *IFIT2*, *IFIT1*, *IFITM3*, *IFIT3*, etc.). Moreover, Gene Ontology (GO) analysis of the top differentially expressed genes (FDR-corrected P value ≤ 0.01 and \log_2 fold change absolute value > 1 , genes = 2148) resulted in hits for defense response to virus ($P = 4.5 \times 10^{-23}$) and the virus responsive, interferon-gamma-mediated signaling pathway ($P = 2.1 \times 10^{-21}$) (Additional file 1: Table S4).

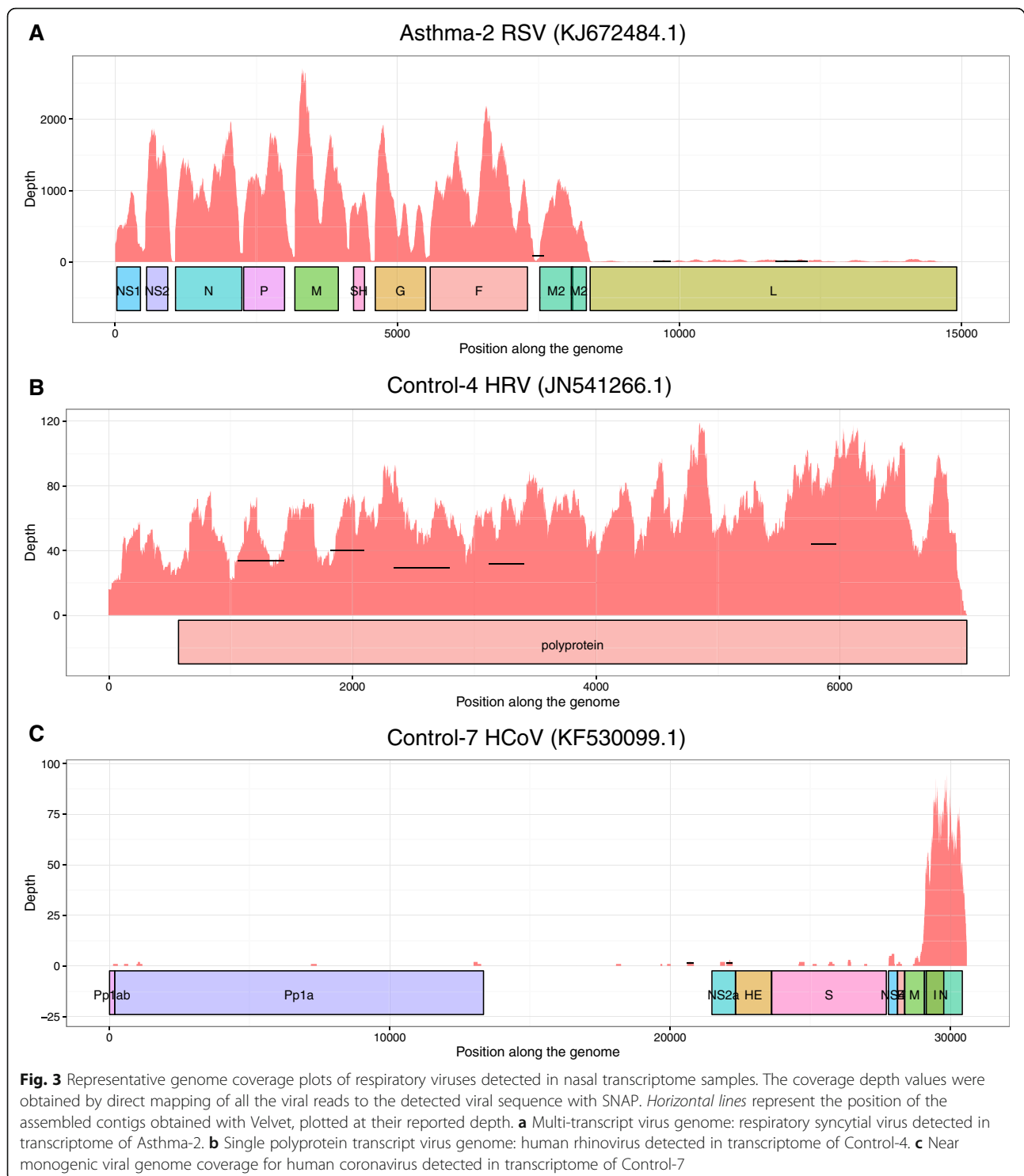
We next examined whether genes previously shown to be upregulated in acute viral infection of the nasal airway are discriminatory for infected subjects without illness. We found 55 of the 56 genes previously reported as upregulated during the peak stage of acute viral respiratory illness were upregulated in our Virus-High subjects (all 55 genes FDR corrected differential expression P values $< 1.0 \times 10^{-4}$ and \log_2 expression fold changes > 2) [26].

Secondly, we determined the in vitro primary airway epithelial transcriptional response to acute respiratory virus infection to draw comparisons to the transcriptional signature exhibited in vivo among the Virus-High subjects without illness. To accomplish this, we performed whole transcriptome paired differential expression analysis of HRV-A16-infected and mock-infected mucociliary differentiated airway epithelial cell cultures from three donors (Additional file 1: Table S5). We

identified 493 airway epithelial genes significantly differentially expressed between the HRV-A16- and mock-infected cultures (1% FDR and absolute \log_2 fold change in expression > 2 , Additional file 1: Table S6). We found 92.8% of these genes were differentially expressed in our Virus-High subjects compared to No-Virus subjects. Moreover, the fold changes in expression between our in vitro and in vivo datasets were highly correlated (Additional file 2: Figure S3, $\rho = 0.94$, $P < 2.2 \times 10^{-16}$). These results indicate that respiratory virus infection of the airway epithelium without clinical illness can mediate epithelial expression changes characteristic of active acute infection.

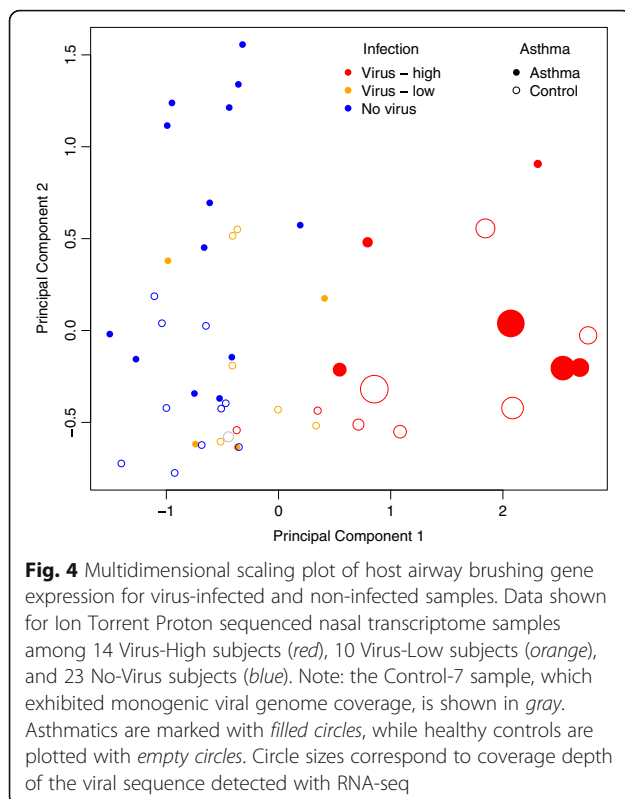
Low-level viral infection without illness is associated with airway epithelial viral responses

We next investigated whether any of the host viral responses seen in Virus-High subjects were also present in the Virus-Low individuals. Performing transcriptome-wide single gene differential expression, we identified 100 differentially expressed genes (FDR 5%) between the Virus-Low and No-Virus subjects (Additional file 1: Table S7). We found that 80 of those genes were also significantly differentially expressed between the Virus-High and No-Virus subjects. We found that 42 of the 100 genes were downregulated in virus-infected samples, and were significantly enriched in genes from the eIF2 signaling pathway ($P = 1.67 \times 10^{-30}$, Ingenuity Pathway



Analysis (IPA)). Closer investigation of those genes revealed downregulation of the ribosomal subunit genes, which predicted activation of *EIF2AK2*. The *EIF2AK2* gene encodes for a kinase which is activated by virus infection to inhibit expression of translational machinery,

as a host viral defense mechanism [29]. Importantly, *EIF2AK2* expression was among the 58 significantly up-regulated genes in the Virus-Low subjects. The up-regulated genes showed a significant enrichment in non-activated ($P = 3.89 \times 10^{-09}$) and activated neutrophil



($P = 1.82 \times 10^{-07}$) signatures, as well as an activated macrophage signature ($P = 3.20 \times 10^{-02}$). These 100 genes separated the Virus-Low subjects from No-Virus subjects in an MDS analysis (Fig. 5a). Hierarchical clustering of all 48 subjects from this study using these genes demonstrated even higher expression levels in the Virus-High group (Fig. 5b). In fact, the collapsed expression of these genes is highly correlated with the read depth of respiratory virus carried ($\rho = 0.73$).

Network analysis of epithelial response to viral infection without illness

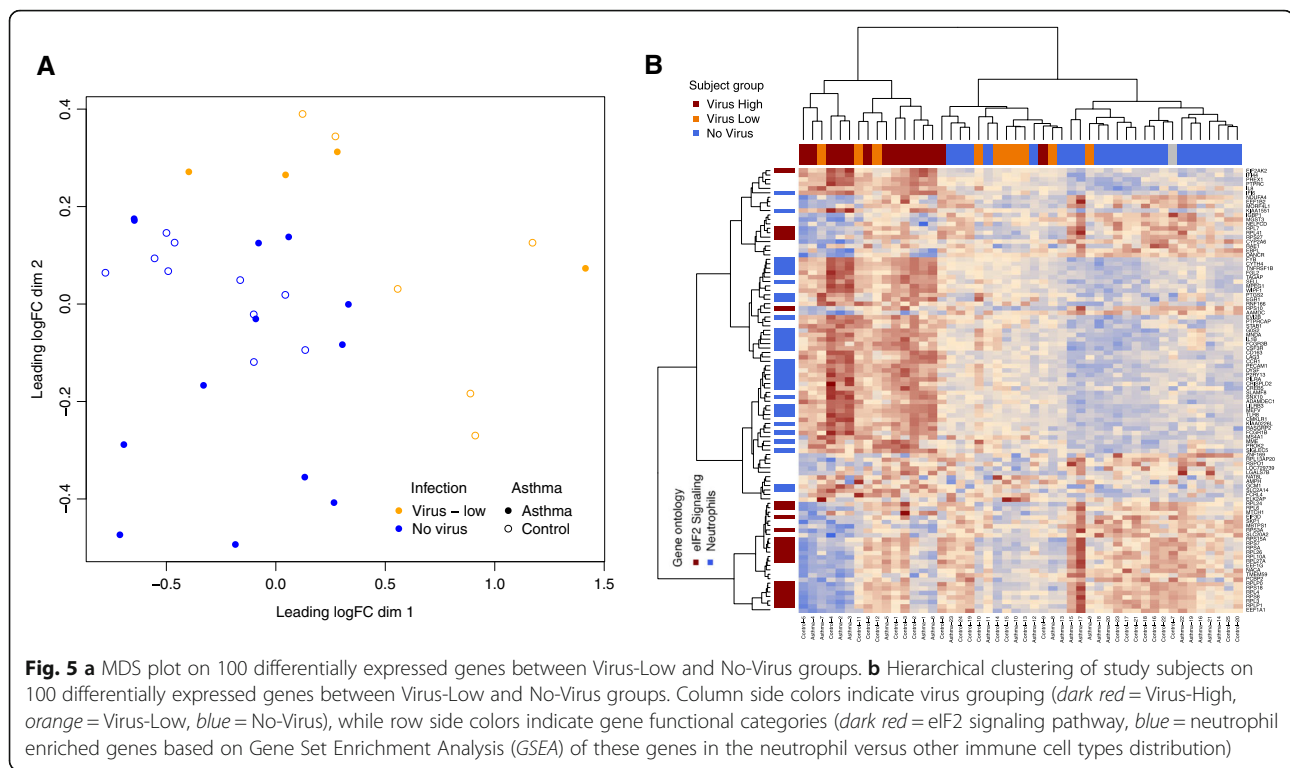
In an effort to better understand the mechanisms activated among virus-infected samples, we applied weighted gene co-expression network analysis (WGCNA) [30] to the 8126 differentially expressed genes from the Virus-High to No-Virus comparison. WGCNA utilizes a correlation network approach to identify modules of highly correlated genes, likely representing a common biological process or pathway. WGCNA analysis detected nine distinct network modules (Table 2, Additional file 2: Figure S4, Additional file 1: Table S8) comprising 3378 of the differentially expressed genes. Modules were assigned arbitrary color names by the WGCNA package, and expression values of genes within each module were summarized by a single module eigengene value for each subject/module.

Activated immune cell modules

Five of the detected modules (pink, blue, black, magenta, and yellow) had highly correlated expression patterns, which were characterized by high module eigengene expression values in Virus-High subjects and low values in the No-Virus samples, regardless of asthma status (Table 2). Enrichment analysis of the module member genes for Gene Ontology (GO) terms found that immune response, defense response, and leukocyte activation terms were enriched in pink and blue modules (Additional file 1: Table S9). We hypothesized that these modules represent a signature of immune cells infiltrating upper airways in response to a viral infection. To test this hypothesis, we compared the results of in vitro stimulation of epithelial cells with HRV with our in vivo data and identified a set of 199 genes highly upregulated in the in vivo samples but not in the in vitro experiment (i.e., representing genes that may be expressed by immune cell types that are not represented in the in vitro airway epithelial cell-restricted experiment; Additional file 2: Figure S3). Interestingly, we found *CCL8* among the 199 genes, suggesting that this gene is only expressed by the immune cells infiltrating the nasal airway. We found strong over-representation of these 199 in vivo infection-specific genes in the blue and pink modules by hypergeometric testing (blue: 121 genes, $P = 3.10 \times 10^{-65}$; pink: 28 genes, $P = 2.11 \times 10^{-20}$), with no over-representation of these genes found among the remaining expression modules. We then explored which activated leukocyte populations might be represented by these modules by examining enrichment of the module genes in the publicly available expression datasets of specific activated immune cell populations [31] (Additional file 1: Table S10). The blue module indicated strong activated macrophage and neutrophil signatures, as well as enrichment for a dendritic cell signature. The pink module was most strongly enriched for neutrophil genes, with a weaker but still significant enrichment for activated macrophage and eosinophil genes. Black, magenta, and yellow modules were all significantly enriched for the activated macrophage signatures, and the yellow module was also enriched for neutrophil and eosinophil signatures. Additionally, we examined cytopins from second brushings of several infected study subjects, which were collected at the same time as the first brushing used for RNA-seq. We observed both macrophages and neutrophils among the airway epithelial cells in these cytopins (Additional file 2: Figure S5).

Virus-Low group module

The green module was significantly enriched in ribosomal genes ($P = 1.52 \times 10^{-33}$) and genes involved in the eIF2 signaling pathway, and had a statistically significant overlap with the genes differentially expressed in the



Virus-Low group ($P = 3.23 \times 10^{-08}$, Table 2, Additional file 1: Table S7). Green module eigengenes were down-regulated in the virus carriers, including the Virus-Low group. Overall module expression was similarly down-regulated in the Virus-High subjects.

Modules of asthma-associated genes

The remaining three modules were associated to varying degrees with asthma among non-viral carrying samples (Table 2). In fact, 59% of the 3378 virus-associated genes incorporated into modules from our network analysis were within these three asthma-associated modules. In particular, the brown module's expression was much higher among non-infected asthma samples compared to non-infected healthy controls. Close examination of this module's genes revealed a signature of type 2 airway epithelial inflammation. In fact, the biomarkers of type 2 asthmatic airway inflammation (*POSTN*, *CLCA1*, *CST1*, and *ALOX15*) [32, 33] were all present in this module. Ingenuity Pathway Analysis (IPA®, QIAGEN, Redwood City, CA, USA, <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>) predicted IL-13, the primary driver of Th2 inflammation [34], to be the main upstream regulator of this module. These genes and the overall module eigengene expression were significantly lower in virus-infected versus non-infected subjects. In contrast, the red module eigengene values were higher in non-infected asthmatics versus non-

infected controls, but were upregulated with viral infection among both asthmatic and control subjects. The red module was enriched for genes that encode for the production of actin cytoskeleton and the remodeling of epithelial adherens junctions genes (Additional file 1: Table S9). Expression eigengene values for the turquoise module were lower among non-infected asthmatics versus non-infected controls, and were downregulated among virus carriers. The turquoise module was the largest module, comprising 1615 genes. Genes in this module were greatly enriched in both cilium and microtubule cytoskeleton GO categories, and decreased formation of cilia was among the enriched functions identified by IPA. This is supported by recent work observing significant downregulation of cilium genes in airway epithelial cultures in response to rhinovirus infection [35]. Lastly, we found that genes within this module were highly enriched for *RFX3* transcription factor binding sites. The RFX family of transcription factors is known to drive the expression of genes involved in motile cilia formation of metazoans [36].

Discussion

Dual RNA sequencing of host and microbe has been proposed as a groundbreaking method to simultaneously detect/quantify microbes and reveal their influence on the host transcriptome [37]. To date, however, there are still few examples of this method applied to human

Table 2 WGCNA module summary and functional enrichments

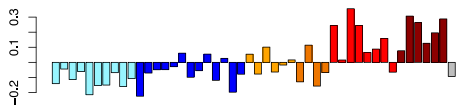



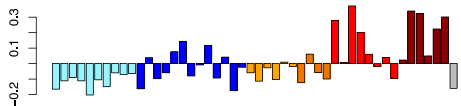
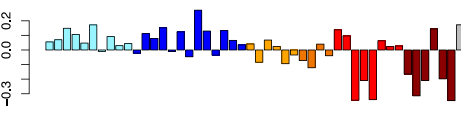
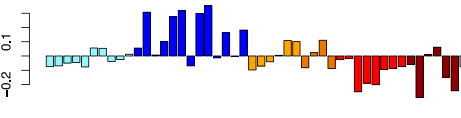
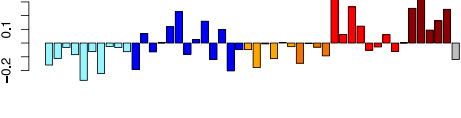
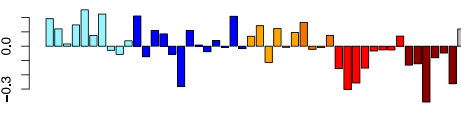
Module (gene number)	Hub genes	Viral Cor.*	Enrichment terms	Immune cell type enrichment
Pink (100) 	SPI1	0.83	GO:0006952 defense response ($P_{adj} = 6.17e-07$)	Neutrophils ($P = 7.46e-11$)
	PLEK		GO:0006955 immune response ($P_{adj} = 2.25e-06$)	Act. neutrophils ($P = 4.44e-09$)
	HCK		IPA path: Fcγ receptor-mediated phagocytosis in macrophages and monocytes ($P = 4.97e-10$)	Eosinophils ($P = 3.74e-04$)
	FGR		IPA upstream regulator: IFNG ($P = 4.77e-16$)	Act. macrophages ($P = 9.78e-03$)
	TLR4		IPA function: leukocyte migration (increased) ($P = 3.82e-22$)	Act. eosinophils ($P = 2.67e-02$)
Blue (726) 	SLAMF7	0.75	GO:0006955 immune response ($P_{adj} = 5.31e-76$)	Act. macrophages ($P = 2.49e-16$)
	SP100		GO:0006952 defense response ($P_{adj} = 2.15e-29$)	Dendritic cells ($P = 3.01e-11$)
	DTX3L		GO:0045321 leukocyte activation ($P_{adj} = 3.79e-26$)	Act. neutrophils ($P = 3.78e-05$)
	GIMAP4		IPA upstream regulator: IFNG ($P = 1.66e-100$)	Neutrophils ($P = 1.99e-04$)
	OAS3		IPA path: interferon signaling ($P = 3.35e-22$)	NK cells ($P = 2.64e-02$)
Black (117) 	DCAKD	0.28	GO:0031982 ~ vesicle ($P_{adj} = 1.38e-02$)	Act. macrophages ($P = 4.06e-03$)
	GNA13		GO:0012505 endomembrane system ($P_{adj} = 2.10e-02$)	
	PICALM		IPA upstream regulator: CHEK2 ($P = 6.49e-05$)	
	CREB1		IPA path: ephrin receptor signaling ($P = 3.26e-04$)	
	LARP4		IPA function: viral infection (increased) ($P = 1.02e-04$)	
Magenta (85) 	TNKS2	0.32	No significant DAVID enrichments	Act. macrophages ($P = 4.57e-02$)
	VCPIP1		IPA path: oxidative phosphorylation ($P = 8.05e-05$)	
	MIER1		IPA path: mitochondrial dysfunction ($P = 6.51e-04$)	
	CD2AP		IPA path: NF-κB signaling ($P = 6.14e-03$)	
	ARAP2			
Yellow (189) 	CHST9	0.70	No significant DAVID enrichments	Act. neutrophils ($P = 4.00e-04$)
	ITGA5		IPA function: inflammatory response ($P = 2.09e-11$)	Act. macrophages ($P = 9.78e-03$)
	C8orf47		IPA function: quantity of leukocytes ($P = 1.67e-10$)	Neutrophils ($P = 1.56e-02$)
	OXSRI		IPA function: differentiation of cells ($P = 1.74e-10$)	Act. eosinophils ($P = 1.83e-02$)
	ITPRIPL2			

Table 2 WGCNA module summary and functional enrichments (*Continued*)

 <p>Green (179)</p>	RPL10A	-0.46	GO:0006414 translational elongation ($P_{\text{adj}} = 3.65\text{e-}44$)	B cells ($P_{\text{adj}} = 1.07\text{e-}07$)
	RPL3		KEGG: hsa03010 ribosome ($P_{\text{adj}} = 3.34\text{e-}37$)	Th2 cells ($P_{\text{adj}} = 8.30\text{e-}05$)
	EEF2		GO:0006412 translation ($P_{\text{adj}} = 7.72\text{e-}36$)	T cells ($P_{\text{adj}} = 2.33\text{e-}04$)
	RPS14		IPA path: eIF2 signaling ($P = 4.07\text{e-}47$)	Mast cells ($P_{\text{adj}} = 1.33\text{e-}03$)
	RPL4			Th1 cells ($P_{\text{adj}} = 2.42\text{e-}02$)
 <p>Brown (203)</p>	ELOVL5	-0.74	Type 2 inflammation ($P = 3.30\text{e-}04$)	<i>No significant immune cell enrichments</i>
	CDH26		<i>No significant DAVID enrichments</i>	
	ALOX15		IPA upstream regulator: IL13 ($P = 2.77\text{e-}05$)	
	FETUB			
	VWF			
 <p>Red (164)</p>	CFL1	0.50	GO:0015629 actin cytoskeleton ($P_{\text{adj}} = 2.41\text{e-}06$)	Act. neutrophils ($P_{\text{adj}} = 2.35\text{e-}02$)
	CAP1		GO:0031252 cell leading edge ($P_{\text{adj}} = 6.68\text{e-}05$)	Act. macrophages ($P_{\text{adj}} = 3.04\text{e-}02$)
	VASP		IPA path: remodeling of epithelial adherens junctions ($P = 2.42\text{e-}09$)	
	DIAPH1		IPA upstream predictor: TGFB1 ($P = 6.88\text{e-}09$)	
	TMBIM1		IPA function: organization of cytoskeleton (increased) ($P = 3.65\text{e-}15$)	
 <p>Turquoise (1,615)</p>	C6orf165	-0.43	GO:0005929 cilium ($P_{\text{adj}} = 2.78\text{e-}30$)	<i>No significant immune cell enrichments</i>
	ARMC2		GO:0015630 microtubule cytoskeleton ($P_{\text{adj}} = 1.02\text{e-}23$)	
	CAPSL		GO:0005930 axoneme ($P = 3.01\text{e-}21$)	
	EFCAB6		IPA upstream regulator: RFX3 ($P = 1.74\text{e-}05$)	
	ANKRD66		IPA function: formation of cilia (decreased) ($P = 1.47\text{e-}43$)	

*Cor = Spearman correlation of module eigengenes with depth of detected viral genomes (for virus high samples only)

biological samples [16, 38]. This method is especially needed for viral detection, due to the high sequence variability of viruses and their lack of a universal marker gene. We detail here a dual RNA-seq workflow for both the detection/quantification of respiratory viruses and generation of host transcriptome data from a single nasal respiratory epithelial brushing library. Our analysis confirms the common nature of respiratory virus carriage without illness in children and finds that this carriage, even at low levels, has a dramatic impact on respiratory mucosa function.

We suspect that several features of our study design aided in successful virus detection in these samples. First, our immediate lysis of the complete sample and application of the total lysate onto the RNA column without pre-processing steps allowed us to avoid cumbersome, potentially biased, separation methods that could have resulted in loss of viral RNA. Secondly, we took advantage of the fact that most respiratory viruses have polyadenylated RNA genomes and that their expressed transcripts are also polyadenylated [39–41]. We performed standard polyA enrichment of total RNA

isolated from the nasal samples and were therefore able to enrich libraries for respiratory virus sequences in the context of a standard library preparation for human whole transcriptome sequencing. Importantly, we find that our modest transcriptome sequencing depth ($\sim 1.2 \times 10^7$ /sample) was sufficient to detect virus infection at species and even strain level among subjects without illness.

Detection of viruses in asymptomatic children has been previously reported by multiple PCR studies, which detected viruses in 33–52% of asymptomatic subjects using nasopharyngeal swabs and washings [23, 24, 42–45]. The observed prevalence in these studies is higher than the rate observed in our samples (15.5%, detected by RNA-seq or qPCR). Furthermore, unlike previous studies, we did not detect any dual viral infections. We believe that this may be explained by sample type (brushings versus washes). We hypothesize that washes may sample a larger surface area of the nasal passages, and represent a sort of “environmental sampling” of nasal airway. In contrast, the brushings result in a point sampling of airway cells and may reflect viral infection of airway epithelial cells, rather than viral material passively inhaled and deposited on the nasal epithelium surface. Although we observed close correlation in viral qPCR and RNA-seq detection results, other studies have found discrepancies in PCR-based and other detection methods. For example, one study found that PCR-based testing of samples resulted in asymptomatic carriage rates of 41.7% versus 4.4% using conventional methods (immunochromatography, direct fluorescent antibody techniques, etc.) in the same samples [43]. In contrast to PCR-based detection of short viral sequences, with RNA-seq-based detection we are able to cover large portions of the virus genomes, as well as determine the host epithelial responses, which lends high credibility to the metatranscriptomics method. In particular, the virus genome read pileups with breaks between viral transcripts, for the multi-transcript virus genomes, indicate that this method detects transcriptionally active viruses. The theoretically unlimited breadth of detection for metatranscriptomic sequencing was apparent with our detection of respiratory virus species not targeted with viral qPCR primers, including HPIV strain 4a and influenza virus B. Additionally, in three cases, the viral qPCR method indicated that the respiratory virus present in the sample was human rhinovirus; however, closer examination of all the viral sequencing reads revealed that the actual virus present was human enterovirus for two of the subjects and human coronavirus in the third case. We take the RNA-seq result to be correct since it is based on detection of a much larger portion of the viral genome sequence than the qPCR assay result. This shows the higher specificity of respiratory virus detection RNA-seq, which is also not subject to the limitations of primer cross-

reactivity. Lastly, we were able to strain-type viruses based on small numbers of reads, a distinction difficult to achieve using qPCR-based methods, and that could have important disease implications.

The dual RNA-seq method applied to host-dominated samples allowed us to determine that a clinically “silent” carriage of respiratory viruses results in dramatic alterations of the airway molecular phenotype with 8126 genes significantly affected. Comparison of acute *in vitro* experimental viral infection of airway epithelium and GO analyses of viral infection categories revealed high similarity between signatures of viral carriage without illness and active viral response. We did not observe significant differences in airway expression between asthmatic and healthy control subjects carrying a respiratory virus. We did note, however, that three of the nine modules associated with viral carriage were also associated with asthma status in uninfected subjects. Interestingly, we found that genes associated with type 2 airway inflammation, a common airway endotype in asthma, were also expressed much lower in virus-infected asthmatic and healthy children than in subjects without any viral reads detected. The association of viral infection with dampened type 2 inflammation may be due to anti-viral immune cells infiltrating the airways. Supporting this, we found viral gene response modules strongly characteristic of activated macrophage and neutrophil populations. This finding is significant, as it strongly suggests that not only are epithelial cells responding to the presence of a virus, but that the local immune system is being directed by the level of virus carriage we detect. Additionally, we identified a subset of the differentially expressed genes that were significantly altered even in subjects with very few viral reads present. These genes represented host suppression of ribosomal subunit genes by activated *EIF2AK2*, leading to inhibition of viral replication, and infiltration of airway epithelium by activated neutrophils.

We acknowledge multiple factors in our study that limit interpretation of our results and raise important questions beyond the scope of this work. Namely, mild respiratory illness may not have been reported or recognized in some subjects due to community recruitment. Additionally, symptoms and severity of rhinitis were not explicitly scored on study subjects, preventing tests of association here. Additionally, the cross-sectional nature of our study does not allow us to determine if the viral carriage we detect pre-dates, post-dates, or is unrelated to a clinically significant viral infection. It is unlikely that our detection post-dates infection, since a recent respiratory illness (6 weeks prior) was among the study exclusion criteria for the asthmatics. We acknowledge that the lack of recruitment of symptomatic patients to the

study prevented us from conducting a comprehensive comparison of the effects of the viral infection without illness to viral infection effects in patients with illness. Moreover, the collection of only a single sample from subjects did not allow us to determine the persistence of this viral carriage. Furthermore, the polyA-enrichment protocol limits this method to detection of viruses with polyadenylated genomes and transcripts and would miss any viruses without those characteristics potentially present in the samples. Finally, while dual RNA-seq was initially proposed for studying host-microbiome interactions in bacterial infections [37], the approach used here would not be applicable for studying bacterial pathogens, since most bacteria do not have polyadenylated transcripts. Our results indicate that virus carriage, despite absence of illness, dramatically modulates the airway endotype of asthmatic subjects, in particular for genes related to asthma disease and asthmatic airway inflammation. These questions merit further investigation, especially considering the strong association and contribution of rhinitis severity to asthma severity in children receiving guidelines-based care [46]. This method, applied to an appropriately large sample size, would also allow for investigation of whether particular virus species and strains have gene- and pathway-specific effects on host expression. Similarly, this method will help investigate how the expression levels of individual viral genes might affect specific host genes and pathways.

Conclusions

We applied a protocol and analytical workflow for dual RNA-seq of nasal airway epithelial brushings to concurrently detect respiratory viruses and generate the host airway transcriptome in one sample preparation and sequencing experiment. The viral detection with this method achieved 86% sensitivity when compared with qPCR-based respiratory virus detection. The sensitivity, specificity, and scope of this method resulted in the detection of multiple respiratory viruses and provided sufficient resolution to distinguish between virus species and strains in infected subjects with low total viral read percentages. Co-generation of the host transcriptome data allowed us to determine that molecular function of the airway is greatly altered by a “silent” virus carriage, which is characterized by immune cell infiltration of the airway, downregulation of ciliated cell gene expression, and modulation of type 2 inflammatory and other asthma-associated gene expression patterns.

Methods

GALA study subjects

All nasal brushing analysis subjects included in this study were recruited as a part of the Genes-environments and

Admixture in Latino Americans II (GALA II) childhood asthma cohort. The asthmatics and controls included in this study were randomly selected. Asthma was defined by physician diagnosis and the presence of at least two asthma symptoms (coughing, wheezing, or shortness of breath). Study recruitment excluded subjects with respiratory illness as judged by a recruiting nurse evaluating cold symptoms. Additional study exclusion for asthmatics included a reported respiratory illness in the past 6 weeks. Clinical characteristics including medication usage and asthma symptoms in the past 2 weeks are listed in Additional file 1: Table S1 for study subjects.

Processing of nasal brushings to extract RNA

Cytology brushes were used to collect airway epithelium from the posterior surface of the inferior turbinate from children with and without asthma (controls) as part of the GALA II study [47] (Fig. 2a, Additional file 1: Table S1). The anatomical location of our sampling resulted in the collection of airway epithelial cells. Collected brushes were immediately submerged in RLT Plus lysis buffer and β -mercaptoethanol and frozen at -80°C until extraction, so that nucleic acids could be released from all sampled cells and microorganisms (host epithelial cells, viral particles, etc.). Total RNA was extracted, followed by DNase treatment.

qPCR screening for respiratory virus infection

Respiratory virus infection was determined by measuring the following six viral species by qPCR assay: human rhinovirus (HRV), respiratory syncytial virus (RSV), human metapneumovirus (HMPV), human parainfluenza virus 1 (HPIV1), human parainfluenza virus 2 (HPIV2), and human parainfluenza virus 3 (HPIV3). Each assay was conducted using RNA extracted from the nasal airway brushings of 161 GALA II study subjects (92 asthmatics and 69 healthy controls). Specifically, RNA was reverse transcribed using oligo-dT primers and SuperScript III Reverse Transcriptase (Invitrogen). The cDNA synthesized for each sample was used for qPCR assays (5 ng per reaction) using specific primers and probes for each viral species as previously described [25]. The qPCR amplification cycling was conducted as follows: 3 minutes at 95°C , followed by 45 cycles of denaturation for 10 seconds at 95°C and annealing for 1 minute at 60°C . The Ct values for each virus were used to determine the presence (Ct <45) or absence (undetectable Ct) of that viral infection within each donor.

Viral biomarker qPCR screening for putative respiratory virus infections

Expression of the *CCL8* and *CXCL11* genes was measured by qPCR assays in RNA extracted from the nasal

airway brushings of the same cohort of 161 GALA II study subjects (92 asthmatics and 69 healthy controls) as used for the virus detection qPCR assays. The cDNA was synthesized as described above, and each qPCR reaction was run using 5 ng of cDNA template and custom PrimeTime qPCR assays for the biomarkers *CCL8* (Hs.PT.58.39289889.g), *CXCL11* (Hs.PT.58.26723814), and the housekeeping gene *GusB* (Hs.PT.51.2648420) (IDT Technologies). The qPCR amplification cycling was conducted as follows: 3 minutes at 95 °C, followed by 40 cycles of denaturing for 10 seconds at 95 °C, and annealing for 30 seconds at 60 °C. The Ct values for *CCL8* and *CXCL11* were normalized to *GusB* expression levels within each sample, and gene expression values were determined using the comparative Ct analysis method [48].

Viral pathogens detection in Ion Proton RNA-seq data

Kapa Biosystems mRNA-seq library kits (catalog number KK8421) were used to generate sequencing libraries for Ion Torrent whole transcriptome sequencing for the 21 samples with respiratory virus detected by qPCR, 2 samples with suspected viral infection based on high *CCL8* and *CXCL11* expression, and 25 randomly selected samples with no suspected asymptomatic virus carriage. Barcodes and adapters compatible with the Ion Torrent Proton sequencing instrument were used. Sequencing was conducted with P1 chips. The viral pathogen detection pipeline is schematically illustrated in Fig. 2b. Raw sequencing reads were mapped to the reference human genome hg19 (GrCHR37) with the Genomic Short-read Nucleotide Alignment Program (GSNAP) [49], and unmapped reads from this step were mapped again to the human genome with SNAP [50] using an edit distance of 14. Reads that did not map to the human genome with either algorithm and were at least 50 nucleotides long were then queried against the NCBI Nucleotide database with BLAST+ [51]. Reads of viral origin identified in this step were then assembled into longer contigs using Velvet [52]. At this step, the estimated coverage of pathogen was used as a parameter for Velvet, and was calculated as the number of viral reads, multiplied by the mean read length divided by the size of the viral genome with the highest number of BLAST hits. Assembled contigs were then queried again against the BLAST nucleotide database to verify species of viral origin. The sequence for the most common BLAST contig hit was downloaded from NCBI and indexed with SNAP. All the viral reads were mapped to this sequence with SNAP using an edit distance of 30, and final coverage depth was estimated with the BEDTools suite [53]. The final virus coverage plot was generated with input generated by BEDTools using the ggplot2 (version 1.0.0) [54] R package in the R statistical environment. If less than 95%

of the viral reads mapped to the identified closest NCBI reference sequence, and there were >50 such unmapped reads, another iteration of the virus detection pipeline was executed, starting from assembly of the unmapped reads from this step into contigs.

Host differential expression analysis

Reads uniquely mapping as human with GSNAP [49] were used for quantification of host expression using the human hg19 iGenomes GTF file with htseq-count script [55]. For subjects with more than 12 million raw sequencing reads mapped to human genes, we performed downsampling of the reads to a maximum of 12 million reads per sample (Additional file 1: Table S2). Differential gene expression analysis between Virus-High and No-Virus subjects, as well as between Virus-Low and No-Virus subjects, was performed in R using the DESeq2 package (version 1.8.1) [56] while adjusting for asthma status. All 8126 genes significantly differentially expressed between the Virus-High and No-Virus subjects were used to build a co-expression network with the WGCNA R package (version 1.51) [30]. Gene lists forming each co-expressed module were analyzed with the DAVID functional enrichment tool [57], DAVID R Web Service (version 1.6.0) [58] and QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN, Redwood City, <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>). Enrichment for type 2 inflammation genes was performed with hypergeometric testing using a list of 70 genes from Poole et al., 2014 [32]. Enrichment for activated and non-activated immune cell types was performed using the Human Immune Cell Transcriptome dataset (GSE3982) [31] obtained from the NCBI Gene Expression Omnibus (GEO) with the Generally Applicable Gene-set Enrichment (GAGE) R package (version 2.18.0) [59] and Gene Set Enrichment Analysis (GSEA) [60], in the same way as previously described [61]. All other statistical analyses and plots were generated in the R environment using the following packages: edgeR (version 3.10.2) [62] for generating MDS plots, and ggplot2 (version 2.1.0) [54] for generating coverage plots of viral genomes. Reported correlations with viral depth were calculated using Spearman correlations. Collapsed gene expression for the 100 Virus-Low differentially expressed genes was calculated with the collapseRows function [63] implemented in the WGCNA R package.

In vitro HRV stimulation

Human tracheal airway epithelial cells from three donors were cultured using a modified Schlegel method as previously described [64–67]. Basal cells were seeded and grown at an air-liquid interface as described previously [68]. Briefly, 6.5-mm inserts (Corning, catalog number

3470) were coated with 2.7 mg/ml collagen (PureCol Bovine Collagen #5005-B, Advanced Biomatrix, San Diego, CA, USA). Epithelial cells were seeded in airway media 1 [68] at 1.2×10^5 cells per insert. At 24 hours post-plating, airway media 1 was replaced with airway medium 2 containing 2% Ultrosor G [68]. Cultures were air-lifted upon the appearance of confluent cell monolayers 72 hours after cells were seeded, and maintained in airway media 2 for 21 days for cell differentiation. After 21 days of growth at the air-liquid interface, paired differentiated cells from each donor were mock-infected (control) or infected with human rhinovirus A16 (HRV-A16). Both apical and basolateral chambers of the inserts were washed with phosphate-buffered saline (PBS), and fresh airway media 2 was added to the basolateral chamber. Apical chambers were infected with HRV-A16 at a multiplicity of infection (MOI) of 5 in X-Vivo 10 serum-free media (Lonza) for 24 hours in a 37 °C humidified incubator containing 5% CO₂; control inserts were incubated with X-Vivo media alone in the apical chamber. Both the apical and basolateral chambers of the inserts were washed three times with warmed PBS, fresh airway media 2 containing USG was added to the basolateral chamber, and the apical chamber was left at the air-liquid interface for an additional 24 hours in a 37 °C incubator. Inserts were washed with PBS, cell monolayers lysed in RLT buffer, and RNA was extracted using an RNA Miniprep Kit (QIAGEN).

RNA-seq libraries were prepared using the standard protocol for the Ion AmpliSeq Transcriptome Human Gene Expression Kit. The prepared libraries were multiplexed and sequenced on the Ion Torrent Proton sequencer. Raw sequencing reads were mapped to human transcriptome reference (hg19) and gene count tables were generated with *ampliseqRNA* plugin v.5.0.0.0 and *Torrent Suite* v5.0.4 (Additional file 1: Table S5). Differential expression analysis was performed with the DESeq2 R package (version 1.4.5) [56], including information on sample pairing in the applied model. We used 483 of 493 genes (ones with matching gene IDs between the two studies) with adjusted *P* values <0.01, and absolute log₂ expression fold changes ≥ 1 were then used to assess whether viral infections without illness in vivo demonstrate similar expression changes as in vitro airway epithelial HRV infection.

Hematoxylin and eosin (H&E) staining of nasal airway brushing cytopins

A second nasal brushing collected at the same time as the brushing used for RNA-seq was solubilized, and the cells were cytopun, H&E stained, and imaged by standard methods.

Additional files

Additional file 1: Table S1. Clinical characteristics of subjects. **Table S2.** Gene count table for 48 RNA-seq nasal brushing subjects from GALA cohort. **Table S3.** Virus-High DESeq differential expression analysis. **Table S4.** DAVID enrichment for Virus High DEGs. **Table S5.** Gene count table for HRV in vitro stimulation dataset. **Table S6.** HRV in vitro DEGs. **Table S7.** Virus Low Differential Expression Analysis. **Table S8.** WGCNA gene2-module. **Table S9.** DAVID for modules. **Table S10.** immune cell enrichments for modules. (XLS 18739 kb)

Additional file 2: Figure S1. Study design. **Figure S2.** Genome coverage for viruses detected in Ion Torrent Proton sequenced samples. **Figure S3.** Comparison of in vivo and in vitro log₂ gene expression fold changes by infection status. **Figure S4.** WGCNA gene correlation dendrogram showing nine detected gene co-expression modules. **Figure S5.** H&E staining of nasal airway epithelial brushing cell cytopin from a viral-infected subject. (PDF 30208 kb)

Acknowledgements

Not applicable.

Funding

Funding was provided through NIH 5R01HL128439-02.

Availability of data and materials

The datasets supporting the conclusions of this article, (1) count tables for host expression data and (2) in vitro HRV stimulation RNA-seq count tables, are available in Additional file 1: Tables S2 and S5, and (3) BAM files of non-human reads are available in the NCBI SRA repository with BioProject ID: PRJNA348728 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA348728>) or with SRA identifier SRP091658 (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?cmd=search_obj&m=&s=&term=SRP091658&go=Search). The source code for the virus detection pipeline, as well as all the analyses carried out in this paper, is available at Github (https://github.com/agaves/Virus_manuscript_Genome_Biology) and Zenodo (<http://doi.org/10.5281/zenodo.204929>).

Authors' contributions

AWA designed the study, analyzed and interpreted data, and drafted the manuscript. JLE, CR, and RH performed the experiments, interpreted data, and helped draft the manuscript. RD, AHL, SSO, RK, and TEF suggested strategies, interpreted data, advised on analyses, and helped draft the manuscript. WJJ performed immune cell staining and helped draft the manuscript, CE, JRS, and EGB participated in the study design, interpreted data, and helped draft the manuscript, and MAS conceived, designed and led the study, and helped draft the manuscript. All authors read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The GALA study (nasal brushings) was approved by local institutional review boards (UCSF, IRB number 10-00889, Reference number 153543, NJH HS-2627), and written assent/consent was received from all subjects and their parents. The tracheal airway epithelial cells were obtained in a de-identified fashion from the National Jewish Health (NJH) live cell core. The NJH live cell core is an institutional review board-approved study (HS#2240) for the collection of tissues from consented patients for researchers at NJH. Experimental methods in this study comply with the Helsinki Declaration.

Author details

¹Center for Genes, Environment, and Health, National Jewish Health, Denver, CO, USA. ²Department of Medicine, University of California, San Francisco, CA, USA. ³Department of Medicine, National Jewish Health, Denver, CO, USA. ⁴Department of Pediatrics, National Jewish Health, 1400 Jackson St, Denver, CO 80206, USA. ⁵Children's Hospital Colorado and University of Colorado

School of Medicine, Aurora, CO, USA. ⁶Department of Pediatrics, The Ann and Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. ⁷Department of Biomedical Research, National Jewish Health, Denver, CO, USA. ⁸Centro de Neumología Pediátrica, San Juan, Puerto Rico. ⁹Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA. ¹⁰Division of Pulmonary Sciences and Critical Care Medicine, University of Colorado School of Medicine, Aurora, CO, USA.

Received: 27 October 2016 Accepted: 16 December 2016

Published online: 19 January 2017

References

- Ellis C, Misir A, Hui C, Jabbour M, Barrowman N, Langill J, Bowes J, Slinger R. Detection of respiratory viruses and bacteria in children using a twenty-two target reverse-transcription real-time PCR (RT-qPCR) panel. *World J Pediatr.* 2016;12:183–9.
- Parker J, Fowler N, Walmsley ML, Schmidt T, Scharrer J, Kowaleski J, Grimes T, Hoyos S, Chen J. Analytical sensitivity comparison between singleplex real-time PCR and a multiplex PCR platform for detecting respiratory viruses. *PLoS One.* 2015;10, e0143164.
- García-Arroyo L, Prim N, Marti N, Roig MC, Navarro F, Rabella N. Benefits and drawbacks of molecular techniques for diagnosis of viral respiratory infections. Experience with two multiplex PCR assays. *J Med Virol.* 2016;88:45–50.
- Wylie TN, Wylie KM, Herter BN, Storch GA. Enhanced virome sequencing using targeted sequence capture. *Genome Res.* 2015;25:1910–20.
- Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, Lipkin WI. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio.* 2015;6:e01491–01415.
- Sakudo A, Chou H, Ikuta K, Nagatsu M. Integration of antibody by surface functionalization of graphite-encapsulated magnetic beads using ammonia gas plasma technology for capturing influenza A virus. *Bioorg Med Chem Lett.* 2015;25:1876–9.
- Bagga K, Brougham DF, Keyes TE, Brabazon D. Magnetic and noble metal nanocomposites for separation and optical detection of biological species. *Phys Chem Chem Phys.* 2015;17:27968–80.
- Jue E, Yamanishi CD, Chiu RY, Wu BM, Kamei DT. Using an aqueous two-phase polymer-salt system to rapidly concentrate viruses for improving the detection limit of the lateral-flow immunoassay. *Biotechnol Bioeng.* 2014;111:2499–507.
- Kang W, Cannon JL. A membrane-based electro-separation method (MBES) for sample clean-up and norovirus concentration. *PLoS One.* 2015;10, e0141484.
- Fischer N, Indenbirken D, Meyer T, Lutgehetmann M, Lellek H, Spohn M, Aepfelbacher M, Alawi M, Grundhoff A. Evaluation of unbiased next-generation sequencing of RNA (RNA-seq) as a diagnostic method in influenza virus-positive respiratory samples. *J Clin Microbiol.* 2015;53: 2238–50.
- Wang Y, Zhu N, Li Y, Lu R, Wang H, Liu G, Zou X, Xie Z, Tan W. Metagenomic analysis of viral genetic diversity in respiratory samples from children with severe acute respiratory infection in China. *Clin Microbiol Infect.* 2016;22:458.e1–9.
- Zoll J, Rahamat-Langendoen J, Ahout I, de Jonge MI, Jans J, Huijnen MA, Ferwerda G, Warris A, Melchers WJ. Direct multiplexed whole genome sequencing of respiratory tract samples reveals full viral genomic information. *J Clin Virol.* 2015;66:6–11.
- Graf EH, Simmon KE, Tardif KD, Hymas W, Flygare S, Eilbeck K, Yandell M, Schlaberg R. Unbiased detection of respiratory viruses using RNA-seq-based metagenomics: a systematic comparison to a commercial PCR panel. *J Clin Microbiol.* 2016;54:1000–7.
- Woodhouse SD, Narayan R, Latham S, Lee S, Antrobus R, Gangadharan B, Luo S, Schroth GP, Klenerman P, Zitzmann N. Transcriptome sequencing, microarray, and proteomic analyses reveal cellular and metabolic impact of hepatitis C virus infection in vitro. *Hepatology.* 2010;52:443–53.
- Strong MJ, Xu G, Coco J, Baribault C, Vinay DS, Lacey MR, Strong AL, Lehman TA, Seddon MB, Lin Z, et al. Differences in gastric carcinoma microenvironment stratify according to EBV infection intensity: implications for possible immune adjuvant therapy. *PLoS Pathog.* 2013;9, e1003341.
- Perez-Losada M, Castro-Nallar E, Bendall ML, Freistat RJ, Crandall KA. Dual transcriptomic profiling of host and microbiota during health and disease in pediatric asthma. *PLoS One.* 2015;10, e0131819.
- Sigurs N, Bjarnason R, Sigurbergsson F, Kjellman B, Bjorksten B. Asthma and immunoglobulin E antibodies after respiratory syncytial virus bronchiolitis: a prospective cohort study with matched controls. *Pediatrics.* 1995;95:500–5.
- Sigurs N. Epidemiologic and clinical evidence of a respiratory syncytial virus-reactive airway disease link. *Am J Respir Crit Care Med.* 2001; 163:52–6.
- Sigurs N, Bjarnason R, Sigurbergsson F, Kjellman B. Respiratory syncytial virus bronchiolitis in infancy is an important risk factor for asthma and allergy at age 7. *Am J Respir Crit Care Med.* 2000;161:1501–7.
- Jackson DJ, Gangnon RE, Evans MD, Roberg KA, Anderson EL, Pappas TE, Printz MC, Lee WM, Shult PA, Reisdorf E, et al. Wheezing rhinovirus illnesses in early life predict asthma development in high-risk children. *Am J Respir Crit Care Med.* 2008;178:667–72.
- Rubner FJJD, Evans MD, Gangnon RE, Tisler CJ, Pappas TE, Gern JE, Lemanske RF. Early life rhinovirus wheezing, allergic sensitization and asthma risk at adolescence. *J Allergy Clin Immunol.* 2016. doi:10.1016/j.jaci.2016.03.049.
- Kloepfer KM, Lee WM, Pappas TE, Kang TJ, Vrtis RF, Evans MD, Gangnon RE, Bochkov YA, Jackson DJ, Lemanske Jr RF, Gern JE. Detection of pathogenic bacteria during rhinovirus infection is associated with increased respiratory symptoms and asthma exacerbations. *J Allergy Clin Immunol.* 2014;133:1301–7. 1307 e1301–1303.
- Tovey ER, Stelzer-Braid S, Toelle BG, Oliver BG, Reddel HK, Willenborg CM, Belessis Y, Garden FL, Jaffe A, Strachan R, et al. Rhinoviruses significantly affect day-to-day respiratory symptoms of children with asthma. *J Allergy Clin Immunol.* 2015;135:663–9. e612.
- Stelzer-Braid S, Tovey ER, Willenborg CM, Toelle BG, Ampon R, Garden FL, Oliver BG, Strachan R, Belessis Y, Jaffe A, et al. Absence of back to school peaks in human rhinovirus detections and respiratory symptoms in a cohort of children with asthma. *J Med Virol.* 2016;88:578–87.
- Ali SA, Gern JE, Hartert TV, Edwards KM, Griffin MR, Miller EK, Gebretsadik T, Pappas T, Lee WM, Williams JV. Real-world comparison of two molecular methods for detection of respiratory viruses. *Virol J.* 2011;8:332.
- McErlean P, Berdnikovs S, Favoreto Jr S, Shen J, Biyasheva A, Barbeau R, Easley C, Barczak A, Ward T, Schleimer RP, et al. Asthmatics with exacerbation during acute respiratory illness exhibit unique transcriptional signatures within the nasal mucosa. *Genome Med.* 2014;6:1.
- Gordon ED, Palandra J, Wesolowska-Andersen A, Ringel L, Rios CL, Lachowicz-Scroggins ME, Sharp LZ, Everman JL, MacLeod HJ, Lee JW, et al. IL1RL1 asthma risk variants regulate airway type 2 inflammation. *JCI Insight.* 2016;1, e87871.
- Tapparel C, Junier T, Gerlach D, Van-Belle S, Turin L, Cordey S, Muhlemann K, Regamey N, Aubert JD, Socal PM, et al. New respiratory enterovirus and recombinant rhinoviruses among circulating picornaviruses. *Emerg Infect Dis.* 2009;15:719–26.
- Farrell PJ, Sen GC, Dubois MF, Ratner L, Slattery E, Lengyel P. Interferon action: two distinct pathways for inhibition of protein synthesis by double-stranded RNA. *Proc Natl Acad Sci U S A.* 1978;75:5893–7.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
- Jeffrey KL, Brummer T, Rolph MS, Liu SM, Callejas NA, Grumont RJ, Gillieron C, Mackay F, Grey S, Camps M, et al. Positive regulation of immune cell function and inflammatory responses by phosphatase PAC-1. *Nat Immunol.* 2006;7:274–83.
- Poole A, Urbaneck C, Eng C, Schageman J, Jacobson S, O'Connor BP, Galanter JM, Gignoux CR, Roth LA, Kumar R, et al. Dissecting childhood asthma with nasal transcriptomics distinguishes subphenotypes of disease. *J Allergy Clin Immunol.* 2014;133:670–8. e612.
- Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, Koth LL, Arron JR, Fahy JV. T-helper type 2-driven inflammation defines major subphenotypes of asthma. *Am J Respir Crit Care Med.* 2009;180:388–95.
- Wills-Karp M, Luyimbazi J, Xu X, Schofield B, Neben TY, Karp CL, Donaldson DD. Interleukin-13: central mediator of allergic asthma. *Science.* 1998;282:2258–61.
- Bai J, Smock SL, Jackson Jr GR, MacIsaac KD, Huang Y, Mankus C, Oldach J, Roberts B, Ma YL, Klappenbach JA, et al. Phenotypic responses of differentiated asthmatic human airway epithelial cultures to rhinovirus. *PLoS One.* 2015;10, e0118286.

36. Thomas J, Morle L, Soulavie F, Laurencon A, Sagnol S, Durand B. Transcriptional control of genes involved in ciliogenesis: a first step in making cilia. *Biol Cell*. 2010;102:499–513.
37. Westermann AJ, Gorski SA, Vogel J. Dual RNA-seq of pathogen and host. *Nat Rev Microbiol*. 2012;10:618–30.
38. Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, Giannoukos G, Boylan MR, Ciulla D, Gevers D, et al. Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A*. 2014;111:E2329–2338.
39. Macnaughton MR, Dimmock NJ. Different patterns of polyadenylation of rhinovirus: specified RNA during multiplication in human aneuploid and diploid cells. *J Gen Virol*. 1976;30:263–7.
40. Pritlove DC, Poon LL, Fodor E, Sharps J, Brownlee GG. Polyadenylation of influenza virus mRNA transcribed in vitro from model virion RNA templates: requirement for 5' conserved sequences. *J Virol*. 1998;72:1280–6.
41. Cowton VM, McGivern DR, Fearns R. Unravelling the complexities of respiratory syncytial virus RNA synthesis. *J Gen Virol*. 2006;87:1805–21.
42. Jartti T, Jartti L, Peltola V, Waris M, Ruuskanen O. Identification of respiratory viruses in asymptomatic subjects: asymptomatic respiratory viral infections. *Pediatr Infect Dis J*. 2008;27:1103–7.
43. Advani S, Sengupta A, Forman M, Valsamakis A, Milstone AM. Detecting respiratory viruses in asymptomatic children. *Pediatr Infect Dis J*. 2012;31:1221–6.
44. DeMuri GP, Gern JE, Moyer SC, Lindstrom MJ, Lynch SV, Wald ER. Clinical features, virus identification, and sinusitis as a complication of upper respiratory tract illness in children ages 4–7 years. *J Pediatr*. 2016;171:133–9. e131.
45. Singleton RJ, Bulkow LR, Miernyk K, DeByle C, Pruitt L, Hummel KB, Bruden D, Englund JA, Anderson LJ, Lucher L, et al. Viral respiratory infections in hospitalized and community control children in Alaska. *J Med Virol*. 2010;82:1282–90.
46. Liu AH, Babineau DC, Krouse RZ, Zoratti EM, Pongracic JA, O'Connor GT, Wood RA, Khurana Hershey GK, Kerckmar CM, Gruchalla RS, et al. Pathways through which asthma risk factors contribute to asthma severity in inner-city children. *J Allergy Clin Immunol*. 2016;138:1042–50.
47. Kumar R, Nguyen EA, Roth LA, Oh SS, Gignoux CR, Huntsman S, Eng C, Moreno-Estrada A, Sandoval K, Penaloza-Espinosa RI, et al. Factors associated with degree of atopy in Latino children in a nationwide pediatric sample: the Genes-environments and Admixture in Latino Asthmatics (GALA II) study. *J Allergy Clin Immunol*. 2013;132:896–905. e891.
48. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc*. 2008;3:1101–8.
49. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*. 2010;26:873–81.
50. Zaharia M, Bolosky WJ, Curtis K, Fox A, Patterson D, Shenker S, Stoica I, Karp RM, Sittler T. Faster and more accurate sequence alignment with SNAP. In *arXiv*; 2011
51. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421. doi:10.1186/1471-2105-10-421.
52. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008;18:821–9.
53. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
54. Wickham H. *ggplot2: elegant graphics for data analysis*. New York: Springer; 2009.
55. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166–9.
56. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
57. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol*. 2007;8:R183.
58. Fresno C, Fernandez EA. RDAVIDWebService: a versatile R interface to DAVID. *Bioinformatics*. 2013;29:2810–1.
59. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: Generally Applicable Gene set enrichment for pathway analysis. *BMC Bioinformatics*. 2009;10:161.
60. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545–50.
61. Gordon ED, Simpson LJ, Rios CL, Ringel L, Lachowicz-Scroggins ME, Peters MC, Wesolowska-Andersen A, Gonzalez JR, MacLeod HJ, Christian LS, et al. Alternative splicing of interleukin-33 and type 2 inflammation in asthma. *Proc Natl Acad Sci U S A*. 2016;113:8765–70.
62. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.
63. Miller JA, Cai C, Langfelder P, Geschwind DH, Kurian SM, Salomon DR, Horvath S. Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics*. 2011;12:322.
64. Liu X, Ory V, Chapman S, Yuan H, Albanese C, Kallakury B, Timofeeva OA, Nealon C, Dakic A, Simic V, et al. ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. *Am J Pathol*. 2012;180:599–607.
65. Supryniewicz FA, Upadhyay G, Krawczyk E, Kramer SC, Hebert JD, Liu X, Yuan H, Chelvaraju C, Clapp PW, Boucher Jr RC, et al. Conditionally reprogrammed cells represent a stem-like state of adult epithelial cells. *Proc Natl Acad Sci U S A*. 2012;109:20035–40.
66. Chu HW, Rios C, Huang C, Wesolowska-Andersen A, Burchard EG, O'Connor BP, Fingerlin TE, Nichols D, Reynolds SD, Seibold MA. CRISPR-Cas9-mediated gene knockout in primary human airway epithelial cells reveals a proinflammatory role for MUC18. *Gene Ther*. 2015;22:822–9.
67. Reynolds SD, Rios C, Wesolowska-Andersen A, Zhuang Y, Pinter M, Happoldt C, Hill CL, Lallier SW, Cosgrove GP, Solomon GM, et al. Airway progenitor clone formation is enhanced by Y-27632-dependent changes in the transcriptome. *Am J Respir Cell Mol Biol*. 2016;55:323–36.
68. Karp PH, Moninger TO, Weber SP, Nesselhauf TS, Launspach JL, Zabner J, Welsh MJ. An in vitro model of differentiated human airway epithelia. Methods for establishing primary cultures. *Methods Mol Biol*. 2002;188:115–37.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

