

RESEARCH ARTICLE

Open Access

# Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods

Jarrett D Morrow<sup>1</sup>, Weiliang Qiu<sup>1</sup>, Divya Chhabra<sup>1,2</sup>, Stephen I Rennard<sup>3</sup>, Paula Belloni<sup>4</sup>, Anton Belousov<sup>5</sup>, Sreekumar G Pillai<sup>6,7</sup> and Craig P Hersh<sup>1\*</sup>

## Abstract

**Background:** Exacerbations of chronic obstructive pulmonary disease (COPD), characterized by acute deterioration in symptoms, may be due to bacterial or viral infections, environmental exposures, or unknown factors. Exacerbation frequency may be a stable trait in COPD patients, which could imply genetic susceptibility. Observing the genes, networks, and pathways that are up- and down-regulated in COPD patients with differing susceptibility to exacerbations will help to elucidate the molecular signature and pathogenesis of COPD exacerbations.

**Methods:** Gene expression array and plasma biomarker data were obtained using whole-blood samples from subjects enrolled in the Treatment of Emphysema With a Gamma-Selective Retinoid Agonist (TESRA) study. Linear regression, weighted gene co-expression network analysis (WGCNA), and pathway analysis were used to identify signatures and network sub-modules associated with the number of exacerbations within the previous year; other COPD-related phenotypes were also investigated.

**Results:** Individual genes were not found to be significantly associated with the number of exacerbations. However using network methods, a statistically significant gene module was identified, along with other modules showing moderate association. A diverse signature was observed across these modules using pathway analysis, marked by differences in B cell and NK cell activity, as well as cellular markers of viral infection. Within two modules, gene set enrichment analysis recapitulated the molecular signatures of two gene expression experiments; one involving sputum from asthma exacerbations and another involving viral lung infections. The plasma biomarker myeloperoxidase (MPO) was associated with the number of recent exacerbations.

**Conclusion:** A distinct signature of COPD exacerbations may be observed in peripheral blood months following the acute illness. While not predictive in this cross-sectional analysis, these results will be useful in uncovering the molecular pathogenesis of COPD exacerbations.

**Keywords:** Network analysis, Chronic obstructive pulmonary disease, Gene expression profiling, Biomarker

## Background

Chronic obstructive pulmonary disease (COPD) is characterized by progressive airflow obstruction accompanied by chronic inflammation. It is one of the leading causes of morbidity and mortality worldwide and is often caused by environmental exposure such as cigarette smoke [1]. COPD exacerbations, periods of acute deterioration, are a

major reason for COPD mortality and a major source of the high healthcare expenditure in patients with COPD. Acute exacerbations of COPD are characterized by symptoms of shortness of breath, cough, and sputum production. Although these exacerbations are often caused by bacterial or viral infections [2] or inhaled particles, the variability in occurrence within COPD patients and the familial aggregation of exacerbations indicate that other factors such as genetics are important in determining the onset, severity and frequency [3]. Also, the frequency of acute exacerbations appears to be a stable trait [4],

\* Correspondence: [craig.hersh@channing.harvard.edu](mailto:craig.hersh@channing.harvard.edu)

<sup>1</sup>Channing Division of Network Medicine, Brigham and Women's Hospital, 181 Longwood Avenue, Boston, MA 02115, USA

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

supporting genetic susceptibility, and loci associated with COPD exacerbations have been identified [5,6].

Despite the fact that gene expression data from lung tissues should provide greater sensitivity to detect the molecular signature of COPD exacerbations, COPD is a systemic disease, and blood is more accessible for genomics and biomarkers studies in large scale clinical trials and potentially in clinical practice than is lung tissue samples. Previous attempts to study lung disease via whole blood experiments have been successful while studying asthma [7] and idiopathic pulmonary fibrosis [8,9]. Further supporting the use of blood expression profiling, previous COPD studies have documented differential expression in overlapping genes from both blood and lung samples [10,11]. Gene expression in peripheral blood has been associated with COPD and related phenotypes [12].

Network medicine approaches provide a roadmap towards the understanding of complex diseases by studying interacting gene sets and pathways, instead of individual genetic determinants [13,14]. Network medicine techniques have been applied to the study of COPD [15].

We hypothesized that we could identify a signature of frequent COPD exacerbations using gene expression data and protein biomarker data, both collected from peripheral blood samples. The goal is to use network methods to understand the molecular pathogenesis of COPD exacerbations, and perhaps predict onset through minimally invasive means. This study built upon the prior publications involving the use of peripheral blood to examine the molecular pathogenesis of COPD and other complex diseases [8,11,12], and leveraged the power of network analysis methods to uncover gene expression signatures.

## Methods

### Study population

This analysis used expression data from 248 Caucasian COPD subjects from the Treatment of Emphysema with a Selective Retinoid Agonist (TESRA), a randomized controlled trial of palovarotene for treatment of COPD (clinicaltrials.gov identifier NCT00413205) [16,17]. TESRA subjects were former smokers with COPD who experienced two or fewer exacerbations requiring outpatient treatment with antibiotics or oral steroids or one exacerbation requiring hospitalization within the prior year. The total number of subjects in the study was 410. In our analysis, the number of exacerbations in the year prior to enrollment was considered both as a linear variable (0,1,2) and as a binary variable (0 vs. 1 or more). The baseline blood samples, from which our analysis data were obtained, were collected prior to randomization and any treatment associated with the clinical trial. Quantitative emphysema measurements were performed on chest computed tomography (CT) scans, including percent of lung with attenuation (%LAA)  $\leq -910$  HU and mean lung

attenuation at the 15th percentile on lung attenuation curve (Perc15). The baseline phenotype data in our analyses were obtained prior to randomization and any treatment. All subjects provided written informed consent. TESRA was approved by the Institutional Review Boards at all participating centers (see Additional file 1).

### Gene expression data

RNA was isolated from whole blood samples collected in PaxGene RNA tubes (PreAnalytiX, Franklin Lakes, NJ). Gene expression data were obtained using the Affymetrix Human Genome U133 Plus 2.0 Array. Data were available for 309 samples, including 39 replicates, as outlined in Additional file 2: Table S1. Of the 270 remaining unique arrays, three were found to be outliers during the quality control process and were discarded. All 54,675 probes passing quality control were retained regardless of level of variance across samples. These data were background corrected and quantile normalized using the robust multi-array average (RMA) method via the R Bioconductor package *affy* [18]. The 249 subjects that self-identified as Caucasian were included. Using information from a GWAS, one additional subject was omitted based on a genetic ancestry that did not cluster with the other Caucasian subjects, leaving 248 subjects for analysis.

### Protein biomarkers

A total of 140 plasma biomarkers were measured in duplicate at Rules Based Medicine (Austin, TX) and Quest Diagnostics (Valencia, CA), as previously described [17]. Biomarker data are available for 245 of the 248 subjects with associated gene expression data. Any biomarker with greater than 10% and less than 95% of its values below the lower limit of quantitation (LLOQ) were converted to a binary present vs. absent variable ( $n = 25$ ). Two biomarkers with greater than 95% values below the LLOQ were excluded from the analysis. An empirical normal quantile transformation was performed on the 113 remaining continuous biomarker values. Model #1 from Additional file 2: Table S2 was used to assess the association of biomarker levels with the number of recent exacerbations. Logistic regression was used for the binary biomarker values, and linear regression was used for the continuous biomarker values.

### Gene expression analysis

Microarray batch effects visible from MDS plots were addressed by adjusting for the microarray lot number as a covariate. For each expression probe, biomarker or module eigengene, we fitted a linear regression model to detect association with variables of interest using R statistical software (v 3.0.2). In the microarray data analysis, an empirical Bayes shrinkage method was used to obtain a moderated t-test statistic and its p-value in the R

Bioconductor package *limma* [19]. Adjustment for multiple testing controlled for false discovery rate (FDR) [20]. The regression models for each COPD phenotype are provided in Additional file 2: Table S2.

#### Weighted Gene Co-expression Analysis (WGCNA)

The WGCNA method was used to identify groups of probes that have similar expression in the sample population, using the R package WGCNA [21]. In order to identify possible sample outliers, a sample network based on squared Euclidean distances was built. An iterative approach was taken, where samples having a standardized connectivity below  $-5$  were removed [21]. One sample outlier was excluded using this process, leaving 247 subjects. Signed networks were built using biweight midcorrelation as the correlation function, and a soft thresholding power of 12. WGCNA produces a set of modules (labeled by color), each containing a set of unique genes. The module eigengene is defined as the first principal component of the expression matrix of the probes within the module. For association tests, the module eigengenes are included in the regression models from Additional file 2: Table S2, similar to the expression probes, and adjustment for multiple testing was performed using Bonferroni correction, where the burden is reduced in comparison with a probe-based analysis. Driver or hub genes within each module are the probes with both high gene significance (GS) and high module membership (MM) metrics, represented by the  $p$ -value for the test of gene expression association with phenotype, and the correlation of gene expression with the module eigengene, respectively.

#### Pathway analysis

We examined enrichment of curated gene sets from the gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases in the gene groups from association tests or the co-expression modules using a hypergeometric test in the R Bioconductor package *GeneAnswers* [22]. In the GO analysis, only the biological process (BP) category was considered. Further examination of the gene sets was performed using the gene set enrichment analysis (GSEA) to examine up- and down-regulated genes common to previous experiments and curated gene sets within all available MSigDB collections [23]. The query gene lists (top association test results or module genes) were preranked according to their log fold change values from the gene expression association tests. GSEA was performed on both a set of all probes and subsets of probes from WGCNA modules, using a similar concept to the GO and KEGG analyses. The Cytoscape plug-in *EnrichmentMap* [24] was used to plot the relationships between the individual GSEA results, as well as between the sets of results from different analyses.

## Results

#### Gene expression association with phenotype variables

Characteristics of study subjects are shown in Table 1. The top five expression probesets ranked by  $p$ -value for association with number of exacerbations in the past 12 months are provided in Additional file 2: Table S3. After controlling for FDR, no probes were statistically significant. The results for all other models are also provided in Additional file 2: Table S3, where only one probe across the COPD-related phenotypes was found to be significant at  $FDR < 0.05$ , for the six minute walk test. This probe did not have a gene annotation and maps to a region with no obvious biological significance. Enrichment tests were performed using the set of genes within the regression results having a  $p < 0.05$  for the exacerbation phenotype (Table 2). Although the KEGG results demonstrate an enrichment of autoimmune-related genes, the overall results indicate a lack of specificity, as demonstrated by the presumably-unrelated diseases represented. In addition, the GO results are very broad, with processes related to protein ubiquitination and regulation of calcium concentration being dominant. To improve specificity, we used a network approach, whereby genes with similar characteristics may be grouped prior to pathway analysis.

#### Network analysis

A weighted gene co-expression network was constructed as outlined in Methods. The final network consisted of 32 modules, ranging in size from 30 to 19,748 probes. Two dendrograms of the module structure are presented in Additional file 2: Figures S1-S2, as the number of probes (54,675) dictated that the network be built in two blocks. There was high correlation within the modules as expected, with the mean module membership (MM), or mean correlation of gene expression with the module eigengene, varying from 0.54 to 0.81 across all modules of interest. The grey module is a grouping of probes with outlying gene expression profiles and was not considered further.

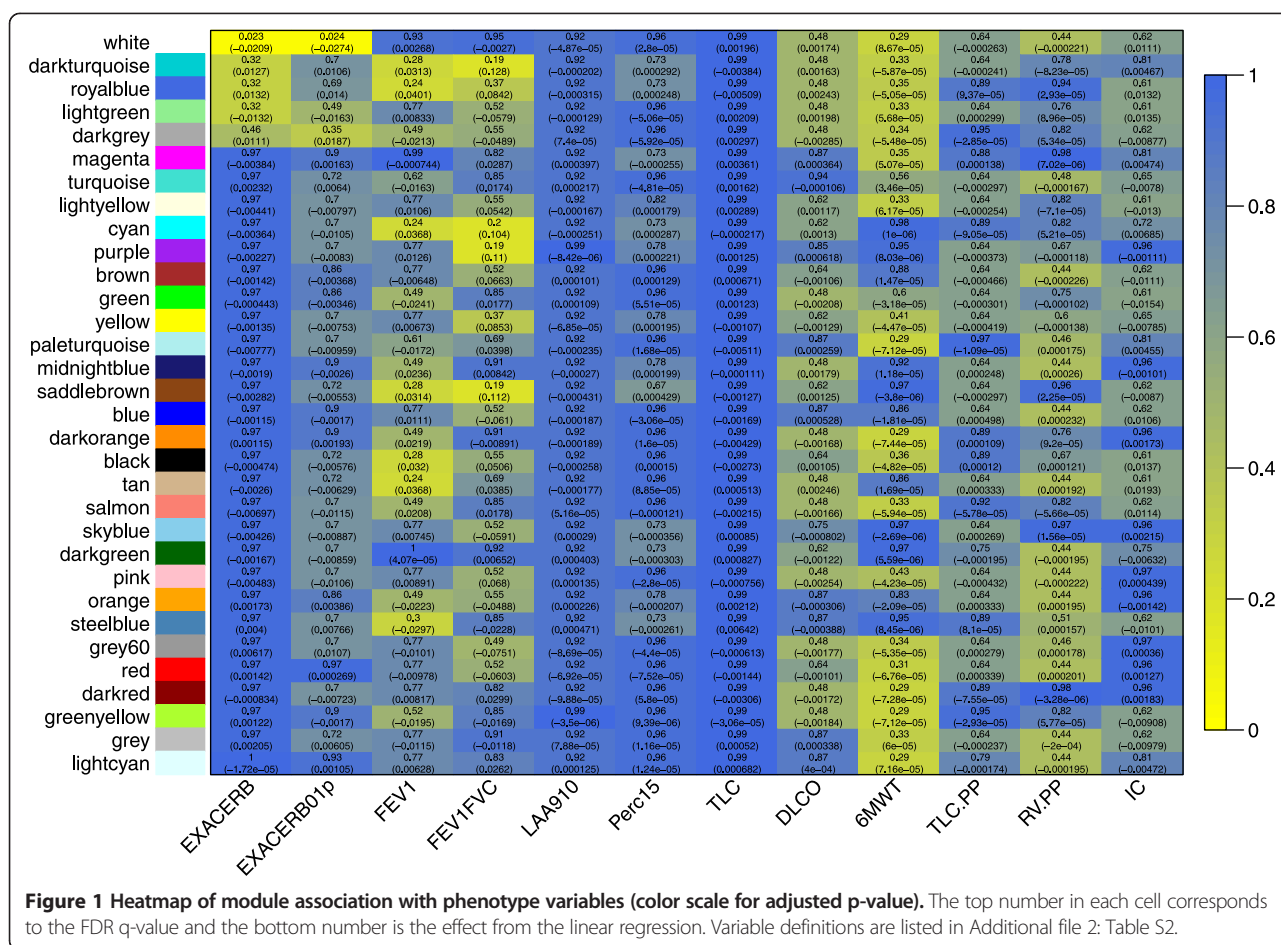
Tests of association between COPD phenotypes and the module eigengenes were performed for each model in Additional file 2: Table S2, and summarized in Figure 1. Across all phenotype variables and modules, only the white module (number of probes,  $n = 57$ ) was significant after controlling for FDR ( $q$ -value = 0.02) for the association with the number of exacerbations within the year prior to enrollment. In addition, four other modules were nominally significant ( $p < 0.05$ ) for the association with exacerbations (royalblue  $n = 118$ , lightgreen  $n = 147$ , darkturquoise  $n = 73$ , darkgrey  $n = 65$ ) (Additional file 2: Table S4). All probesets in these modules had variance above the lower 5%ile for all probesets. The top modules associated with both exacerbations phenotypes were similar (Additional file 2: Table S5).

**Table 1 TESRA study subjects**

Demographics	n = 248 expression (n = 245/248 biomarker)	Association with exacerbations
Age (years)	66.4 +/- 7.9	linear regression p-value = 0.11
Gender	Female 78 Male 170	Fisher's Exact Test p-value = 0.86
Race	Caucasian 100%	NA
Former smokers	100%	NA
Smoking history (pack-years)	47 +/- 24.3	linear regression p-value = 0.02
BMI	26 +/- 4.7	linear regression p-value = 0.39
<b>COPD outcomes</b>		
Exacerbations in past 12 months	0 n = 123 1 n = 102 2 n = 23	
FEV1 (L)	1.34 +/- 0.35	
FEV1 % predicted	48.8 +/- 9.3	
FEV1/FVC	0.43 +/- 0.09	
% LAA ≤ -910 HU	41.65 +/- 16.2	
Perc15 (HU)	-946 +/- 23.6	
TLC (L)	6.9 +/- 1.4	
TLC % predicted	101.6 +/- 15.4	
DLCO (mL/min/mmHg)	12.3 +/- 3.9	
6 minute walk test (m)	317.5 +/- 111.2	
Residual volume % predicted	120.1 +/- 31	
Inspiratory capacity (L)	2.4 +/- 0.67	

**Table 2 Top categories enriched in association results for exacerbation phenotype (gene n = 999)**

GO category	ID	Genes in category	p-value	FDR q-value
Protein ubiquitination	GO:0016567	23	2.78E-008	4.42E-006
Purine ribonucleoside triphosphate catabolic process	GO:0009207	17	2.40E-006	4.78E-005
Guanosine-containing compound catabolic process	GO:1901069	16	1.04E-006	2.69E-005
Calcium ion homeostasis	GO:0055074	15	3.16E-007	1.68E-005
Epidermal growth factor receptor signaling pathway	GO:0007173	14	1.18E-007	9.37E-006
Cellular calcium ion homeostasis	GO:0006874	14	1.12E-006	2.69E-005
Activation of protein kinase activity	GO:0032147	10	2.68E-004	2.66E-003
Negative regulation of protein serine/threonine kinase activity	GO:0071901	9	9.97E-007	2.69E-005
KEGG category	ID	Genes in category	p-value	FDR q-value
Cell adhesion molecules (CAMs)	04514	13	3.86E-003	1.44E-001
B cell receptor signaling pathway	04662	11	2.72E-004	4.75E-002
T cell receptor signaling pathway	04660	11	5.62E-003	1.54E-001
Antigen processing and presentation	04612	10	1.23E-003	5.73E-002
Chagas disease (American trypanosomiasis)	05142	10	1.20E-002	2.23E-001
Primary immunodeficiency	05340	7	5.37E-004	4.75E-002
Allograft rejection	05330	7	7.66E-004	4.75E-002
Autoimmune thyroid disease	05320	7	5.81E-003	1.54E-001



Lacking similar support across the results for the other phenotypes, we chose to focus exclusively on the exacerbations phenotype for the remainder of the study. The putative hub genes for the five modules of interest are highlighted in the lists of module genes in Additional file 2: Tables S6-S10. These hub genes with low gene significance (GS) p-values and high module membership (MM) correlations for the five modules include IGHM (immunoglobulin heavy constant mu; GS = 0.002, MM = 0.9), KLRD1 (killer cell lectin-like receptor subfamily D, member 1; GS = 0.006, MM = 0.89 and GS = 0.007, MM = 0.88), AFF3 (AF4/FMR2 family, member 3; GS = 0.0014, MM = 0.8), GPR56 (G protein-coupled receptor 56; GS = 0.00024, MM = 0.8), and GBP5 (guanylate binding protein 5; GS = 0.009, MM = 0.77 and GS = 0.016, MM = 0.77) and GBP1 (guanylate binding protein 1, interferon-inducible; GS = 0.028, MM = 0.87).

GO and KEGG pathway analyses were performed on the five modules above. The results are provided in Table 3 for the white module and Additional file 2: Tables S11-S14 for the other four modules. In contrast to the results for the overall gene expression association, the enrichment tests for each module provide more specific functional

signatures, with B cell receptor signaling and autoimmune disease being enriched in the results for the white module; B cell pathway is also seen in the lightgreen module results. The darkgrey and darkturquoise module results are highlighted by cytokine activity, with the darkgrey module showing a strong signal for interferon activity and viral response pathways.

GSEA was performed on the five modules, and the results meeting the FDR q-value < 0.25 are provided in Table 4. B cell, T-cell and viral response signatures dominate the results, with the white, royalblue and darkgrey modules demonstrating enrichments for viral response gene sets. EnrichmentMap networks were constructed from Table 4. These are shown in Figure 2 and Additional file 2: Figure S3, respectively. The darkgrey module network shows multiple highly-connected nodes related to viral response. An EnrichmentMap network was also produced using the GSEA results for the set of genes nominally associated (p < 0.05) with the exacerbations phenotype (model #1); the GSEA results are provided in Additional file 2: Table S15 and the EnrichmentMap network is shown in Additional file 3: Figure S4. Many similar

**Table 3 Top eight pathway analysis results for white module (unique gene symbols n = 32)**

GO category	ID	Genes in category	p-value	FDR q-value
B cell receptor signaling pathway	GO:0050853	2	0.00072	0.10
Antigen processing and presentation of exogenous peptide antigen via MHC class II	GO:0019886	2	0.0055	0.13
Antigen receptor-mediated signaling pathway	GO:0050851	2	0.011	0.13
Skeletal muscle tissue development	GO:0007519	2	0.029	0.18
Skeletal muscle organ development	GO:0060538	2	0.031	0.18
Clustering of voltage-gated potassium channels	GO:0045163	1	0.0036	0.13
Protein localization to juxtaparanode region of axon	GO:0071205	1	0.0048	0.13
Positive regulation of humoral immune response mediated by circulating immunoglobulin	GO:0002925	1	0.0073	0.13
KEGG category	ID	Genes in category	p-value	FDR q-value
B cell receptor signaling pathway	04662	4	7.6E-006	0.00018
Cell adhesion molecules (CAMs)	04514	4	7.4E-005	0.00088
Asthma	05310	2	1.4E-003	0.011
Allograft rejection	05330	2	2.1E-003	0.011
Graft-versus-host disease	05332	2	2.5E-003	0.011
Type I diabetes mellitus	04940	2	2.8E-003	0.011
Intestinal immune network for IgA production	04672	2	3.4E-003	0.012
Autoimmune thyroid disease	05320	2	4.0E-003	0.012

immune cell gene sets are seen, along with multiple non-specific results such as cancer gene sets.

#### Biomarker association with exacerbations phenotype

The association between the number of exacerbations and each of the plasma biomarkers was tested adjusting for covariates and the top ten results are provided in Table 5. Myeloperoxidase (MPO) was the top result (FDR q-value = 0.096), while club cell-16 (CC-16) showed nominal association ( $p < 0.05$ ).

#### Discussion

In order to identify genes associated with COPD-related phenotypes, we initially used standard gene expression models. Despite the sample size, we found a lack of significant results, and a lack of specificity with respect to pathways or diseases from the enrichment analyses. However, network methods were able to group similar probes into correlated modules with potential common functions, that were then tested for association with phenotypes of interest. Using these methods we were able to observe immunological functions across the modules associated with acute exacerbations of COPD. In addition, we were able to recapitulate the molecular signatures of previous studies involving viral lung infections and exacerbations of asthma, a related respiratory disease. The ability to uncover such signatures is limited given that only subjects having two or fewer exacerbations within the previous year were included in the study, as a wider range of exacerbation frequencies

would provide greater resolution to detect the molecular underpinnings of COPD exacerbations. Yet the network methods added value to address these limitations of the dataset.

WGCNA allowed for identification of the putative hub genes for each of the five modules associated with exacerbations. These hubs include the immunoglobulin heavy constant mu (IGHM) gene in the white module, that codes for the C region of the mu heavy chain (defines the IgM isotype). Lending to its role in lung disease, IGHM was found to be down-regulated in idiopathic pulmonary fibrosis (IPF) whole blood samples versus normal samples [9]. From the royalblue module, the putative hub was found to be KLRD1 (CD94) that codes for an antigen preferentially expressed on NK cells. A decrease in its expression has been observed in the blood of COPD patients versus controls [25]. The gene product for AFF3 from the lightgreen module is a tissue restricted nuclear transcriptional activator, formerly known as lymphoid nuclear protein related to AF4, and it was down-regulated with respect to controls in BEAS-2B cells (immortalized human bronchial epithelial cells) treated with low doses of cigarette smoke condensate [26]. From the darkturquoise module, GPR56 encodes G protein-coupled receptor 56; its involvement in COPD has not been previously established. Lastly, the hubs from the darkgrey module, GBP1 and GBP5, guanylate binding proteins whose expression is induced by interferon and are markers of M1 macrophage population, were found to be down-regulated by cigarette

**Table 4 GSEA results for modules associated with exacerbations phenotypes (gene sets with FDR q-value < 0.25 are shown and viral response gene sets are in bold)**

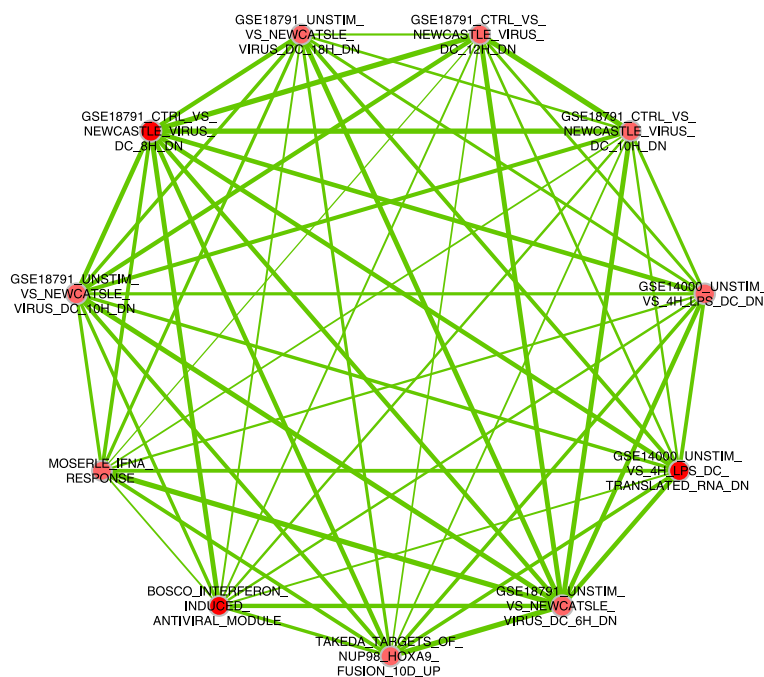
Module name	Gene set description	ID	Set overlap	P-value	FDR q-value	
White (negative)*	NAIVE_BCELL_VS_MONOCYTE_UP	GSE22886	15	0.084	0.207	
	CD4_TCELL_VS_BCELL_DN	GSE10325	18	0.122	0.230	
	<b>BCELL_VS_MONOCYTE_DAY7_FLU_VACCINE_P</b>	<b>GSE29618</b>	17	0.185	0.216	
Royalblue (positive)*	NAIVE_VS_PD1LOW_CD8_TCELL_DN	GSE26495	28	0.004	0.048	
	<b>HEALTHY_VS_RSV_INF_INFANT_PBMC_UP</b>	<b>GSE34205</b>	15	0.016	0.062	
	DEURIG_T_CELL_PROLYMPHOCYTIC_LEUKEMIA_DN		18	0.055	0.151	
Lightgreen (negative)*	BCELL_VS_MDC_DAY7_FLU_VACCINE_UP	GSE29618	17	0.006	0.095	
	CD4_TCELL_VS_BCELL_DN	GSE10325	19	0.012	0.05	
	BCELL_VS_MDC_UP	GSE29618	15	0.012	0.043	
	LUPUS_CD4_TCELL_VS_LUPUS_BCELL_DN	GSE10325	20	0.01	0.046	
	DODD_NASOPHARYNGEAL_CARINOMA_UP		15	0.023	0.068	
	BCELL_VS_MONOCYTE_DAY7_FLU_VACCINE_UP	GSE29618	21	0.025	0.064	
	BCELL_VS_MYELOID_UP	GSE10325	15	0.045	0.064	
	NAIVE_BCELL_VS_MONOCYTE_UP	GSE22886	17	0.06	0.078	
	BCELL_VS_MONOCYTE_UP	GSE29618	23	0.075	0.121	
	CAGGTG_V\$E12_Q6		16	0.116	0.149	
	MEMORY_CD4_TCELL_VS_BCELL_DN	GSE3982	17	0.137	0.14	
	Darkturquoise (positive)*	NAIVE_VS_PD1LOW_CD8_TCELL_DN	GSE26495	26	0.057	0.090
		NAIVE_VS_PD1HIGH_CD8_TCELL_DN	GSE26495	19	0.158	0.166
Darkgrey (positive)*	CTRL_VS_NEWCASTLE_VIRUS_DC_8H_DN	GSE18791	30	0.000	0.012	
	<b>BOSCO_INTERFERON_INDUCED_ANTIVIRAL_MODULE</b>		18	<b>0.003</b>	<b>0.038</b>	
	UNSTIM_VS_4H_LPS_DC_TRANSLATED_RNA_DN	GSE14000	23	0.002	0.030	
	MOSERLE_IFNA_RESPONSE		17	0.011	0.104	
	TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_10D_UP		20	0.016	0.094	
	UNSTIM_VS_NEWCATSLE_VIRUS_DC_18H_DN	GSE18791	19	0.013	0.093	
	UNSTIM_VS_NEWCATSLE_VIRUS_DC_6H_DN	GSE18791	31	0.031	0.105	
	CTRL_VS_NEWCASTLE_VIRUS_DC_10H_DN	GSE18791	21	0.013	0.095	
	CTRL_VS_NEWCASTLE_VIRUS_DC_12H_DN	GSE18791	19	0.026	0.134	
	UNSTIM_VS_4H_LPS_DC_DN	GSE14000	24	0.031	0.141	
	UNSTIM_VS_NEWCATSLE_VIRUS_DC_10H_DN	GSE18791	26	0.034	0.161	
	CTRL_VS_NEWCASTLE_VIRUS_DC_4H_DN	GSE18791	22	0.052	0.166	
	CTRL_VS_NEWCASTLE_VIRUS_DC_6H_DN	GSE18791	27	0.054	0.170	
	BCELL_VS_LUPUS_BCELL_DN	GSE10325	17	0.094	0.250	
	HECKER_IFNB1_TARGETS		21	0.102	0.246	

\*Positive or negative correlation between module and phenotype.

smoke in monocyte-derived macrophages [27]. Taken as a whole, these hubs lend additional evidence to the specific role of each module: B cell activity for the white and light-green modules, NK cell activity for the royalblue module, and macrophage involvement in the darkgrey module.

GO and KEGG pathway enrichment analysis provided common functions of the module genes, namely the immune system response to bacterial and viral infections.

Additionally, the significant enrichment for B cell and asthma pathways in the white module are of note, as are the cytokine and viral response related enrichments in the darkturquoise and darkgrey modules. However, the GSEA results were more informative, revealing signatures of gene expression previously identified in studies involving lung infections, which is consistent with the infectious etiology of most COPD exacerbations. One example was a



**Figure 2** EnrichmentMap network for GSEA results from darkgrey module. Nodes correspond to gene sets and edges correspond to an overlap of the genes within two sets.

top result for the darkgrey module (Table 4). The gene set, BOSCO\_INTERFERON\_INDUCED\_ANTIVIRAL\_MODULE, was a WGCNA module (n = 78 genes) composed of genes downstream of interferon signaling, that was associated with lung function in a study of sputum samples of children with acute asthma exacerbations [28]. This study found decreased activation of interferon pathways in asthmatic children with chronic airway obstruction compared with those without airway obstruction. The darkgrey module may harbor genes associated with differential responses during exacerbations that perhaps also play a role in onset, frequency and severity of COPD exacerbations.

**Table 5** Biomarker association with exacerbation phenotype

Biomarker	Beta	Standard error	P-Value	FDR q-value
MPO	0.33	0.096	0.00069	0.096
CC_16	0.20	0.098	0.041	0.98
ET1_Z	-0.44	0.25	0.074	0.98
HPT	0.16	0.094	0.086	0.98
ENA78	0.18	0.11	0.094	0.98
TNFB	-0.32	0.20	0.11	0.98
TP	-0.16	0.10	0.12	0.98
TSH	0.15	0.10	0.13	0.98
APOH	-0.15	0.10	0.14	0.98
TGFB3	0.30	0.20	0.14	0.98

In another example, the royalblue module enrichment results include the gene set GSE34205\_HEALTHY\_VS\_RSV\_INF\_INFANT\_PBMC\_UP, genes differentially expressed in blood samples from healthy infants and infants suffering from respiratory syncytial virus (RSV) infection [29]. RSV is a major respiratory pathogen that is more common in infants; however the elderly and patients with COPD or other lung diseases are also at risk of developing RSV infections [2]. Using network analysis methods and GSEA, we were able to identify modules that contain genes demonstrating a signature previously seen in the sputum of patients suffering from acute asthma exacerbations and the blood of patients with respiratory infections. However, we detected these signatures using expression data from whole blood of COPD patients, collected months following exacerbations.

The EnrichmentMap network of GSEA results for the darkgrey module (Figure 2) shows an overlap between the asthma exacerbation [28] and the viral-response gene sets, as would be expected given the virus-induced nature of many exacerbations. However the relatively narrow widths of many edges demonstrate that these gene sets are largely distinct. In the network for the lightgreen module in Additional file 2: Figure S3, relatively strong connections exist between the nodes, demonstrating an enrichment of B cell pathway activity within the gene sets. In contrast to the specificity of the networks for the darkgrey and lightgreen modules, the



EnrichmentMap for the probe-based analysis gene set shown in Additional file 2: Figure S4 is less tightly connected with many nodes having zero connections. The probe-based analysis is less specific than the module analyses. Although the red sub-network in Additional file 2: Figure S4 contains several viral response results, the blue sub-network contains various cancer gene sets, in addition to infectious diseases.

In the biomarker analysis, we found an association between MPO and the number of exacerbations, persisting months following illness. Serum MPO levels have been shown to be increased in COPD patients during acute exacerbations [30]. However in contrast to our finding, a previous study did not find a difference in sputum MPO levels between patients with stable COPD having frequent (>3/year) or infrequent (<2/year) exacerbations in the previous 12 months [31]. In addition, a study of bronchoalveolar lavage samples did not find significantly higher MPO levels in COPD patients with frequent ( $\geq 3$ /year) exacerbations as compared to infrequent (<3/year) exacerbations [32]. Perhaps the peripheral blood based signature of COPD exacerbations is able to persist due to systemic modulation of the immune system. Longitudinal studies have shown a reduction in serum MPO levels during treatment for acute exacerbations of chronic bronchitis, followed by the return to higher levels following treatment [33], however no association with the number of exacerbations or their severity was noted. Taken together, our finding and the previous studies provide evidence for MPO as part of the molecular signature to be considered when predicting the onset, severity and frequency of future COPD exacerbations, as it has been previously established as an inflammatory marker, and specifically in COPD [34].

The use of gene expression data from whole blood instead of lung tissue samples may limit the ability of our approach to detect signatures in the specific disease tissue. However, peripheral blood is more easily accessed than lung tissue and may provide specific insight into the infectious, immune and inflammatory mechanisms of COPD exacerbations. Like lung tissue, whole blood is a complex mixture of different cell types. Complete blood counts (CBC) with cell differentials are routinely measured in clinical practice, but were not available in the TESRA study. However, our analysis was able to distinguish B and T cell signatures, cell types which would not be discriminated in a standard CBC. An additional limitation of our approach involves the use biomarker and gene expression data as markers of exacerbations in the past year, yet this may indicate a stable molecular signature differing from that of lung function and other markers of severity. Lastly, one potential confounder is the external trigger of the exacerbation. Regardless of an individual's genetic predisposition for exacerbations, a viral, bacterial or

other environmental cause may lead to different signatures. Despite these limitations, we were able to effectively uncover expression signatures of frequent COPD exacerbations in peripheral whole blood samples.

## Conclusions

In this study a distinct signature of COPD exacerbations was observed in peripheral whole blood months following the illness. This signature was correlated with previous studies involving exacerbations in asthma and viral lung infections, and therefore, it may be useful in uncovering the molecular pathogenesis of COPD exacerbations. The moderate association of MPO and number of exacerbations adds to the body of information supporting the use of MPO as a biomarker. Future studies with longitudinal expression data would allow an investigation of the temporal nature of the network modules, separating causal modules from the downstream modules demonstrating expression changes as a result of the disease state. Similarly, the methods presented here would benefit from gene expression data obtained during acute exacerbations, in addition to samples obtained months following the illness, particularly with respect to prediction. Overall, these approaches will help to elucidate the genetic component determining the heterogeneity of onset, severity, and frequency of exacerbations seen in COPD patients.

## Additional files

**Additional file 1:** Institutional Review Boards for TESRA trial sites.

**Additional file 2:** **Table S1.** Expression data cleaning process. **Table S2.**

Regression models for association with outcomes of interest. **Table S3.**

Top five results from test of gene expression association with phenotype

variables. **Figure S1.** Clustering dendrograms based on topological

overlap in WGCNA (block #1). **Figure S2.** Clustering dendrograms based

on topological overlap in WGCNA (block #2). **Table S4.** WGCNA modules.

**Table S5.** Module eigengene association with exacerbations phenotypes.

**Table S6.** White module contents and information. **Table S7.** Royalblue

module contents and information. **Table S8.** Lightgreen module

contents and information. **Table S9.** Darkturquoise module contents and

information. **Table S10.** Darkgrey module contents and information. **Table**

**S11.** Top pathway analysis results for royalblue module. **Table S12.** Top

pathway analysis results for lightgreen module. **Table S13.** Top pathway

analysis results for darkturquoise module. **Table S14.** Top pathway analysis

results for darkgrey module. **Figure S3.** EnrichmentMap network for GSEA

results from lightgreen module. Nodes correspond to gene sets and edges

correspond to an overlap of the genes within two sets. **Table S15.** GSEA

results for probe association with exacerbations phenotype.

**Additional file 3:** **Figure S4.** EnrichmentMap for GSEA results for probe association with exacerbation phenotype.

## Competing interests

Drs. Morrow, Qiu, Chhabra, and Hersh report no competing interests related to this manuscript.

Dr. Rennard received a research grant from Roche for the TESRA study.

Dr. Belloni is an employee of Genentech.

Dr. Belousov is an employee of Roche.

Dr. Pillai was formerly an employee of Hoffman La Roche and is currently an employee of Eli Lilly and Company.

### Authors' contributions

JM: analysis and interpretation of data, manuscript preparation, and approval of the final version. WQ: analysis and interpretation of data, approval of final version of the manuscript. DC: analysis and interpretation of data, approval of final version of the manuscript. SR: acquisition of data, review of manuscript and approval of the final version of manuscript. PB: acquisition of data, review of manuscript and approval of the final version of manuscript. AB: acquisition of data, review of manuscript and approval of the final version of manuscript. SP: acquisition of data, review of manuscript and approval of the final version of manuscript. CH: analysis and interpretation of data, manuscript preparation, and approval of the final version.

### Acknowledgements

Supported by National Institutes of Health grants R01HL094635, R01NR013377, P01HL105339, R01HL111759. The TESRA study was supported by Roche.

Roche designed the TESRA study and collected the data. The funders had no role in the data analysis in this manuscript, the writing of the manuscript or the decision to submit the manuscript for publication.

TESRA Investigators: Ognian Georgiev, Sofia, Bulgaria; Dimitar Popov, Sofia, Bulgaria; Hristo Metev, Ruse, Bulgaria; Vasil Dimitrov, Sofia, Bulgaria; Yavor Ivanov, St. Plevan, Bulgaria; Libor Fila, Praha, Czech Republic; Vladimir Zindr Vitezna, Karlovy Vary, Czech Republic; Kamil Klenha, Tabor, Czech Republic; Jiri Votruba, Praha, Czech Republic; Jaromir Roubec, Ostrava, Czech Republic; Barna Szima, Szombathely, Hungary; Zsuzsanna Mark, Torokbalint, Hungary; Zoltan Baliko, Pecs Hungary; Zoltan Bartfai, Budapest, Hungary; Katalin Gomori, Balassagyarmat, Hungary; Andres Sigvaldason, Reykjavik, Iceland; Mordechai Kramer, Petach Tikva, Israel; Gershon Ya Fink, Rehovot, Israel; Zeev Weiler, Ashkelon, Israel; Joel Greif, Tel Aviv, Israel; Issahar Ben-Dov, Ramat Gan, Israel; Mordechai Yigla, Haifa, Israel; Leonardo Fabbri, Modena, Italy; Pierluigi Paggiaro, Pisa, Italy; Giorgio Canonica, Genova, Italy; Isa Cerveri, Pavia, Italy; Antra Bekere, Riga, Latvia; Aurika Babjoniseva, Riga, Latvia; Alvil Krams, Stopinu Pagasts, Rigas Rajons, Latvia; Wladyslaw Pierzchala, Katowice, Poland; Dariusz Nowak, Lodz, Poland; Robert Mroz, Bialystok, Poland; Hanna Szelerska-Twardosz, Poznan, Poland; Malgorzata Rzymkowska, Poznan, Poland; Ismail Abdullah, Durban, South Africa; Christo Van Dyk, Western Cape, Worcester, South Africa; Nyda Fourie, Bloemfontein, South Africa; John O'Brien, Cape Town, South Africa; J. Joubert, Bellville-Cape Province, South Africa; Abdool Gafar, Kwa-Zulu Natal, Amanzimtoti, South Africa; Mary Bateman, Cape Town, South Africa; Hannes Van Rensburg, Centurion, South Africa; Lyudmila Yashina, Kiev, Ukraine; Nadezda Monogorova, Donetsk, Ukraine; Oleksandr Dzyubliuk, Kiev, Ukraine; Volodymyr Gavrysyuk, Kiev, Ukraine; Yuriy Feshchenko, Kiev, Ukraine; David Parr, Coventry, United Kingdom, Stephen Rennard, Omaha, NE; Richard Casaburi, Torrance, CA; Gerard Criner, Philadelphia, PA; Mark Dransfield, Birmingham, AL; Charles Fogarty, Spartanburg, SC; Nicola Hanaia, Houston, TX; Carl Griffin, Oklahoma City, OK; Kathi Mcdavid, Oklahoma City, OK; Paul Kvale, Detroit, MI; Barry Make, Denver, CO; Joe Ramsdell, San Diego, CA; Michael D. Roth, Los Angeles, CA; Amir Harafkhaneh, Houston, TX; Peter Sporn, Chicago, IL. Steering Committee: Alvar Agustí (Spain), Peter Calverley (UK), Leonardo Fabbri (Italy), Klaus F. Rabe (Netherlands), Nicolas Roche (France), Michael Roth (US), Jorgen Vestbo (Denmark), Stephen Rennard (US).

### Author details

<sup>1</sup>Channing Division of Network Medicine, Brigham and Women's Hospital, 181 Longwood Avenue, Boston, MA 02115, USA. <sup>2</sup>Division of Biomedical Informatics, University of California, San Diego, CA, USA. <sup>3</sup>University of Nebraska Medical Center, Omaha, NE, USA. <sup>4</sup>Genentech, Member of the Roche Group, South San Francisco, CA, USA. <sup>5</sup>Roche Innovation Center, Penzberg, Germany. <sup>6</sup>Hoffman La Roche, Nutley, NJ, USA. <sup>7</sup>Current address: Eli Lilly and Company, Indianapolis, IN, USA.

Received: 3 July 2014 Accepted: 12 December 2014

Published online: 13 January 2015

### References

- Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187:347–65.
- Alfredo P, Gaetano C, Paolo C, Marco C, Johnston SL, Alberto P. Pathophysiology of viral-induced exacerbations of COPD. *Int J Chron Obstruct Pulmon Dis*. 2007;2:477–83.
- Foreman MG, DeMeo DL, Hersh CP, Reilly JJ, Silverman EK. Clinical determinants of exacerbations in severe, early-onset COPD. *Eur Respir J*. 2007;30:1124–30.
- Hurst JR, Vestbo J, Anzueto A, Locantore N, Müllerova H, Tal-Singer R, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med*. 2010;363:1128–38.
- Foreman MG, DeMeo DL, Hersh CP, Carey VJ, Fan VS, Reilly JJ, et al. Polymorphic variation in surfactant protein B is associated with COPD exacerbations. *Eur Respir J*. 2008;32:938–44.
- Takabatake N, Shibata Y, Abe S, Wada T, Machiya J, Igarashi A, et al. A single nucleotide polymorphism in the CCL1 gene predicts acute exacerbations in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2006;174:875–85.
- Hansel NN, Hilmer SC, Georas SN, Cope LM, Guo J, Irizarry RA, et al. Oligonucleotide-microarray analysis of peripheral-blood lymphocytes in severe asthma. *J Lab Clin Med*. 2005;145:263–74.
- Herazo-Maya JD, Noth I, Duncan SR, Kim S, Ma S-F, Tseng GC, et al. Peripheral blood mononuclear cell gene expression profiles predict poor outcome in idiopathic pulmonary fibrosis. *Sci Transl Med*. 2013;5:205ra136.
- Yang IV, Luna LG, Cotter J, Talbert J, Leach SM, Kidd R, et al. The peripheral blood transcriptome identifies the presence and extent of disease in idiopathic pulmonary fibrosis. *PLoS One*. 2012;7:e37708.
- Poliska S, Csanky E, Szanto A, Sztatmari I, Mesko B, Szeles L, et al. Chronic obstructive pulmonary disease-specific gene expression signatures of alveolar macrophages as well as peripheral blood monocytes overlap and correlate with lung function. *Respiration*. 2011;81:499–510.
- Bhattacharya S, Tyagi S, Srisuma S, Demeo DL, Shapiro SD, Bueno R, et al. Peripheral blood gene expression profiles in COPD subjects. *J Clin Bioinforma*. 2011;1:12.
- Bahr TM, Hughes GJ, Armstrong M, Reisdorph R, Coldren CD, Edwards MG, et al. Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol*. 2013;49:316–23.
- Silverman E, Loscalzo J. Network medicine approaches to the genetics of complex diseases. *Discov Med*. 2012;14:143–52.
- Haas BE, Horvath S, Pietiläinen KH, Cantor RM, Nikkila E, Weissglas-Volkov D, et al. Adipose co-expression networks across Finns and Mexicans identify novel triglyceride-associated genes. *BMC Med Genomics*. 2012;5:61.
- Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, et al. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity*. 2014;40:274–88.
- Jones P, Rames A. TESRA (Treatment of Emphysema with a Selective Retinoid Agonist) study results. *Am J Respir Crit Care Med*. 2011;A6418.
- Cheng DT, Kim DK, Cockayne DA, Belousov A, Bitter H, Cho MH, et al. Systemic soluble receptor for advanced glycation endproducts is a biomarker of emphysema and associated with AGER genetic variants in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2013;188:948–57.
- Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy-analysis of affymetrix GeneChip data at the probe level. *Bioinforma Oxf Engl*. 2004;20:307–15.
- Smyth GK. Limma: Linear Models for Microarray Data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. *Bioinforma Comput Biol Solut Using R Bioconductor*. New York: Springer; 2005. p. 397–420.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57:289–300.
- Horvath S. *Weighted Network Analysis: Applications in Genomics and Systems Biology*. 2011th ed. New York: Springer; 2011.
- Feng G, Du P, Krett NL, Tessel M, Rosen S, Kibbe WA, et al. A collection of bioconductor methods to visualize gene-list annotations. *BMC Res Notes*. 2010;3:10.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL. 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.
- Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*. 2010;5:e13984.
- Hodge G, Mukaro V, Holmes M, Reynolds PN, Hodge S. Enhanced cytotoxic function of natural killer and natural killer T-like cells associated with

- decreased CD94 (Kp43) in the chronic obstructive pulmonary disease airway. *Respirology*. 2013;18:369–76.
26. Veljkovic E, Jiricny J, Menigatti M, Rehrauer H, Han W. Chronic exposure to cigarette smoke condensate in vitro induces epithelial to mesenchymal transition-like changes in human bronchial epithelial cells, BEAS-2B. *Toxicol In Vitro*. 2011;25:446–53.
  27. Doyle I, Ratcliffe M, Walding A, Vanden Bon E, Dymond M, Tomlinson W, et al. Differential gene expression analysis in human monocyte-derived macrophages: impact of cigarette smoke on host defence. *Mol Immunol*. 2010;47:1058–65.
  28. Bosco A, Ehteshami S, Stern DA, Martinez FD. Decreased activation of inflammatory networks during acute asthma exacerbations is associated with chronic airflow obstruction. *Mucosal Immunol*. 2010;3:399–409.
  29. Ioannidis I, McNally B, Willette M, Peebles ME, Chaussabel D, Durbin JE, et al. Plasticity and virus specificity of the airway epithelial cell immune response during respiratory virus infection. *J Virol*. 2012;86:5422–36.
  30. Fiorini G, Crespi S, Rinaldi M. Serum ECP and MPO are increased during exacerbations of chronic bronchitis with airway obstruction. *Biomed Pharmacother*. 2000;54:274–8.
  31. Gompertz S, Bayley DL, Hill SL, Stockley RA. Relationship between airway inflammation and the frequency of exacerbations in patients with smoking related COPD. *Thorax*. 2001;56:36–41.
  32. Tunkaya M, Atis S, Ozge C, Delialioglu N, Polat G, Kanik A. Relationship between airway colonization, inflammation and exacerbation frequency in COPD. *Respir Med*. 2007;101:729–37.
  33. Gompertz S, O'Brien C, Bayley DL, Hill SL, Stockley RA. Changes in bronchial inflammation during acute exacerbations of chronic bronchitis. *Eur Respir J*. 2001;17:1112–9.
  34. Cockayne DA, Cheng DT, Waschki B, Sridhar S, Ravindran P, Hilton H, et al. Systemic biomarkers of neutrophilic inflammation, tissue injury and repair in COPD patients with differing levels of disease severity. *PLoS One*. 2012;7:e38629.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

