Supplementary information

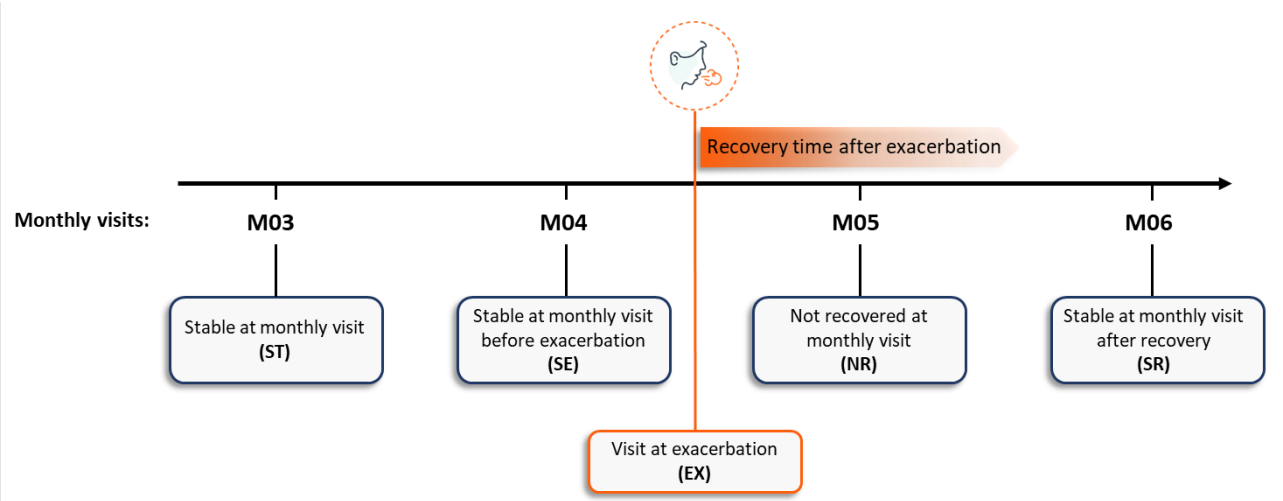
Supplementary section 1. Sample status definition

In this work, all described samples from stable-state or exacerbation visits correspond to samples classified as **ST**/**EX**, on the basis of the following information and definitions.

Sample status was defined as a function of the patient's chronic obstructive pulmonary disease (COPD) status at stable-state or exacerbation visits (see Supplementary Fig. 1):

- **ST**: Sample collected at monthly visits when the patient had stable COPD symptoms. The patient had not experienced an exacerbation within at least 1 week after and 4 weeks before sample collection, and one stable visit after recovery.
- **EX**: Sample collected when the patient visited the hospital due to the onset of an acute exacerbation.
- **SR**: Sample collected at the first monthly visit after recovery from acute exacerbation.
- **SE**: Sample collected at monthly visits when the patient was stable and an acute exacerbation occurred within a week after the monthly visit.
- **NR**: Sample collected at monthly visits but patient has not yet recovered from exacerbation.

Supplementary Fig. 1: Sample status as function of collection time and COPD status at visit.



COPD, chronic obstructive pulmonary disease.

Supplementary section 2. Microarrays Quality Control (QC) metrics

The following microarray quality assessment combined metrics were used during the QC step to accept microarrays:

- Scaling factor (SF) ≤ 11
- 49.12 ≤ Percent of presence (PP) ≤ 60.04
- Glyceraldehyde 3-phosphate dehydrogenase ratio ≤ 2.5

Using those criteria, 585 microarrays, corresponding to 112 patients, passed the QC thresholds, while microarrays from 111 samples did not pass and where excluded from further analysis. A summary of the final set of microarrays passing QC is given in Supplementary Table 1.

Supplementary Table 1. Number of patients and time points with whole blood samples and number of microarrays passing QC thresholds.

	Whole blood samples		Microarrays passing QC threshold	
Time point	**Samples**	**Patients**	**Microarrays**	**Patients**
Enrollment	30	30	27	27
6 months	66	66	58	58
12 months	84	84	75	75
18 months	80	80	62	62
24 months	84	84	68	68
Total routine	**344**	**113**	**290**	**109**
Exacerbation	379	95	295	90
Total samples	**723**	**113**	**585**	**112**

QC, quality control.

Supplementary section 3. Multi-network gene community detection (MNGCD) network layers modelling

We considered in each of the five networks only the 2657 common nodes among all layers and applied the Infomap algorithm for a multiplex, using the default parameters[1]. MNGCD was applied by

modelling different layers of gene interactions. The layers integrated in a multi-network stack of layers were:

- A gene co-expression network by a weighted correlation network analysis method (WGCNA)[2] between genes among the samples in the selected group/condition. The edge weight between genes i and j was defined as: $w_{ij} = |cor_{ij}|^t$ where cor_{ij} is the Pearson correlation between the expression vectors of the genes i and j; t is a scale parameter chosen to reach a power-law distribution of the node degree of the network layer;
- A Gaussian radial basis function (RBF) kernel transformation as similarity measure derived as decreasing function of the Euclidean distance ("RBF expression similarity network"), where genes with a close level of expression had a high weight w of their edge: $w_{ij} = \exp(-\gamma \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2)$, where x_i and x_j were expressions of the genes i and j among the samples in the selected group/condition; γ is a scaling factor identified when the node degree of the network reaches a power law distribution (similar to the WGCNA approach). The number of edges in each of the two expression layers was around 7.0 million;
- A transcription factor co-targeting network, defined from ENCODE experimentally validated interactions[3]. These data were assessed by chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq) for data sets corresponding to 119 different transcription factors in five main cell lines (K562, GM12878, HeLa-S3, H1-hESC and HepG2). In these experiments available from ENCODE database, the largest portion of the transcription factors were canonical sequence-specific factors, a smaller portion gathered POL2 and other basal transcription factors, while the remaining was made up of factors manipulating chromatin. According to the ENCODE database design, the interactions between transcription factors and genes happened in regions of the genome proximal (within ± 2.5 kilobases) to the annotated gene transcription start site. We built our transcription factor co-regulating network: nodes were the genes and those regulated by at least one transcription factor were connected. The weight of the link between gene i and gene j was the number of regulators in common. The number of edges in the transcription factor co-targeting network layer was 6.8 million.
- A microRNA co-targeting network, where the weight of each edge between genes was the number of common microRNAs that regulated them. This information was inferred from experimentally-validated interactions by different assays and collected in miRTarBase database release 7.0[4]. The number of edges in the microRNA co-targeting network layer was 2.1 million.

- A binary human protein-protein interaction network based only on experimentally validated interactions taken from the public database APID (Agile Proteins Interactomes DataServer)[5]. The number of edges in the protein-protein interaction network layer was around 23,000.

With respect to the usual approach[6], in this integrative analysis, we introduced an innovative method defining two layers of interactions based on the gene expression levels to model two very distinct behaviors of genes across the samples. The first layer tends to cluster genes with correlated signals, the second with similar level of expression. The same network layers, moreover, did not introduce any filter on the weight of the links.

The community detection therefore aimed to detect the communities in each multi-network, defined as groups of vertices among which a random walker, with transition probability depending on the weight of their edges, circulated easily and rapidly[7]. The Infomap algorithm explores the multi-network by a random walk through the multi-layers, allowing the detection of disjoint communities. Because of the stochasticity of the procedure, we introduced a robustness step assessment by repeating the algorithm for multiple seeds (100). We then computed a final consensus clustering[8], made up of communities of genes out of a set of stochastic partitions. Therefore, we assessed if genes in each community over-represented specific biological functions, embodied in annotated gene sets; this evaluation was done by the hypergeometric test. The universe was made up of the 2657 common nodes among all layers while the foreground genes were those ones in each community. The p-values of the hypergeometric enrichment were corrected by Benjamini-Hochberg method (significance $p<0.01$). For multiple communities enriched with the same gene set, we considered the minimum p-value. Moreover, for greater confidence in the detected communities, the only enrichments results kept were those detected in the consensus clustering with a rate higher than 80%, iterating the entire algorithm 20 times. Finally, p-values of all the iterations were then log averaged.

Scripts execution flow

The execution environment is a Unix HPC with *slurm* parallel execution queue manager, Infomap and R are required. Here we report a description of the executed code to run Infomap and the enrichment analysis of the identified gene communities (scripts available at https://github.com/muzzial1/MNGCD_for_COPD).

1. Run the following bash scripts in the following order after changing in each script the line *for j in (ls -1d output_*vero)* replacing the term *output_*vero* with the name of the file to analyse:

a) *./execallt.sh* generates 100 partitions using Infomap:

- it executes *perparcopy2.sh*, which in turn runs *stability_sb2.R* 20 times in parallel.

- *stability_sb2.R* calls *myscript_varifile_copy_cal.sh*, which executes Infomap 100 times with different seeds for each iteration.

- Seeds also vary across the 20 invocations of *perparcopy2.sh*.

b) *./execallt_cons.sh* calculates the consensus on the 100 partitions:

- It executes *perparconsensi_v2.sh*, which runs *stabilityperconsenso_v2.R* 20 times in parallel.

- *stabilityperconsenso_v2.R* calculates the consensus as follows:

  - A for loop is executed, starting with the 100 partitions generated by Infomap.

  - In each iteration, *generoperconsenso_cal.sh* is called.

  - The process continues until all generated partitions are identical, producing the consensus matrix.

c) *./execallt_arr.sh* (with argument h or c3 or c5 or BTM, depending on the gene-set collection to use for the enrichment of communities).

d) *./execallt_pval.sh* (with argument h or c3 or c5 or BTM, depending on the gene-set collection to use for the enrichment of communities).

e) *./alltabs.sh* (with argument h or c3 or c5 or BTM, depending on the gene-set collection to use for the enrichment of communities).

2. In the folder final_result in the folder prova_nmi/output_nameofthemultinetwork is the final result (functions enriched in each community with their rate) for the gene set collection selected.

Supplementary section 4. Differentially expressed genes (DEGs) analysis

Gene expression analysis models

DEGs were assessed using the Limma (Linear models for microarrays data) modelling approach[9] with False Discovery Rate (FDR) threshold of 0.05 and Fold Change (FC) threshold of 2 ($|\log_2 FC| > 1$).

Different types of Limma models were applied according to the biological question and sample status.

Model 1: Stability of gene expression profile with time in COPD stable conditions

$$Gene\ expression \sim Group + Age + Gender + COPD\ status + BMI + Run$$

For seasonality *Group*, three models were considered with regards to associations to:

- High/low season, with *Group* = (High season, Low season)
- Season, with *Group* = (Summer, Fall, Winter, Spring)
- Month, with *Group* = (January, February, March, April, May, June, July, August, September, October, November, December)

BMI represents the subject body mass index (available only for this comparison at stable condition visits).

Run represents the experimental analysis batch, a variable that we considered in all the analyses. In fact, the samples were grouped in 3 sets and processed in consecutive batches.

Model 2: Comparison of acute exacerbations of COPD (AECOPD) and stable COPD

$$Gene\ expression \sim Group + High/Low\ season + Age + Gender + COPD\ status + Run$$

With *Group* = (ST, EX)

Model 3: Comparison of infections, pooling samples at stable and exacerbation visits

$$Gene\ expression \sim Group + GRP + Season + Age + Gender + COPD\ status + Run$$

Where *Group* = (ST, EX) and

GRP indicates the type of infection in the studied group (e.g. if we compare the presence of a pathogen versus (vs) its absence, the variable GRP is 0 for absence and 1 for presence).

Model 4: Comparison of infections, samples at stable or exacerbation visits only

$$Gene\ expression \sim GRP + Season\ + Age\ + Gender + \ COPD\ status\ + \ Run$$

Where *GRP* indicates the type of infection in the studied group (e.g. if we compare the presence of a pathogen vs its absence, the variable GRP is 0 for absence and 1 for presence).

We also ran in parallel for each comparison the gene set enrichment analysis (GSEA) on all the transcripts overcoming the interquartile range (IQR) filtering (results not shown for contrasts without DEGs, the only GSEA results reported are in section 5, Supplementary information).

Stability of gene expression profile with time in COPD stable conditions

The aim of this analysis is to assess the evolution of gene expression over time on ST samples. A uniform profile implies a stable profile over time for all patients.

Regression modelling was used to find genes highly correlated with season/month. The global linear model used was Model 1.

From the contrast performed between high and low season as well as from the different associations to High/Low season, Season and Month, there is no gene for which the expression is significantly associated to the season/month in samples taken at stable visits. Therefore, one must consider gene expression profiles are uniform across seasons/month.

However, based on literature, we decided to include seasonality (high/low season) in the models used to compare samples taken at exacerbation vs stable visits.

Comparison of AECOPD and stable COPD

The aim of these analyses is to find a set of DEGs that differentiate exacerbation state vs stable state of patients. The global linear model used was Model 2. In addition, samples were filtered depending of the comparison performed:

- No filter (all samples, pooled)
- Paired samples within patients
- Filter on the type of infection (Bacterial, Viral, Eosinophilic in blood, Eosinophilic in sputum)
- Filter on the pathogen bacteria in sputum (*Haemophilus influenzae*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Moraxella catarrhalis,* and *Pseudomonas aeruginosa*).

All samples

Gene expression was compared between samples taken at exacerbation visits and samples of the same patients at the last available stable visit (paired comparison).

As gene expression in stable COPD samples is considered uniform (see section above), the comparison of AECOPD and stable COPD was also performed by contrasting all exacerbation time points (EX) to all stable time points (ST).

The comparison between AECOPD vs stable visits shows there is no evidence for genes differentially expressed between the two conditions in: pooled comparison (all ST vs all EX samples) and paired comparison (pairs ST-EX within the same patient).

According to type of infection

Gene expression was compared between samples taken at exacerbation visits and samples of the same patients by contrasting all exacerbation time points (EX) to all stable time points (ST). Samples were filtered by type of infection for the comparison.

The comparison of gene expression profiles in whole blood samples taken at exacerbation vs stable visits shows there is no evidence for genes differentially expressed between these two conditions in:

- Bacterial infections;
- Viral infections;
- Eosinophilic infections (blood and sputum).

We should note that in the contrast involving eosinophils in sputum, while the gene FAM19A1 has a significant FDR, it does not have a fold change high enough to be considered.

According to pathogen

Gene expression was compared between samples taken at exacerbation visits and samples of the same patients by contrasting all exacerbation time points (EX) to all stable time points (ST). Samples were filtered by bacterial pathogen for the comparison.

The comparison of gene expression profiles in whole blood samples taken at exacerbation vs stable visits shows there is no evidence for genes differentially expressed between these two conditions in bacterial infections where those pathogens are present:

- *H. influenzae;*
- *S. pneumoniae;*
- *S. aureus;*
- *M. catarrhalis;*
- *P. aeruginosa*.

Comparison of infection conditions

The aim of these analyses is to find a set of DEGs that differentiate distinct types of infections detected by PCR in the sputum samples. For instance, we compared the whole blood gene expression in presence/absence of bacteria or viruses in sputum. We performed different analyses by further selecting samples with the absence of other bacteria and/or viruses.

Moreover, we compared the whole blood gene expression for samples with only bacteria in the sputum with specimens with only viruses or only eosinophils in the sputum.

To answer questions involving infection, we applied Models 3 and 4.

Moreover, we ran several analyses for DEG using the variable HL_Season and separately the $Season\_$ variable as $Season$ variable. These comparisons involved presence vs absence of respectively bacteria, viruses and eosinophils. However, we are reporting in this document the outcomes using the variable HL_Season in the model.

The results of the search for DEGs associated with type of infection (bacterial vs viral vs eosinophilic) are summarized in Supplementary Table 2.

Supplementary Table 2: Number of DEGs associated with type of infection.

ANALYSIS	STABLE	AECOPD	ALL
Presence vs absence of bacteria	NS	NS	NS
Presence vs absence of viruses	NS	NS	NS
B vs V, with eosinophils counted in sputum	NED	NS	NS

B vs V, with eosinophils counted in blood	NED	NS	NS
(Bacterial^{+}, Virus^{-}) vs (Virus^{+}, Bacterial^{-})	NED	NS	NS
Eosinophilic in sputum vs not eosinophilic in sputum	NS	NS	NS
Eosinophilic in blood vs not eosinophilic in blood	NS	NS	NS
B vs E, with eosinophils counted in sputum	NS	NS	NS
B vs E, with eosinophils counted in blood	NS	NS	NS
V vs E, with eosinophils counted in sputum	NED	NS	NS
V vs E, with eosinophils counted in blood	NED	NS	NS

AECOPD, acute exacerbations of COPD; B, bacterial infection; COPD, chronic obstructive pulmonary disease; DEG, differentially-expressed genes; E, eosinophilic; NED, not enough data to perform the analysis; NS, not significant; V, viral infection; vs, versus.

As shown in supplementary Table 3, there is no significant DEG in association to co-infection occurence.

Supplementary Table 3: Number of DEGs associated with type of co-infection

ANALYSIS	STABLE	AECOPD	ALL
(V not B) vs (BV)	NED	NS	NS
(B not V) vs (BV)	NS	NS	NS
B vs BV, with eosinophils counted in sputum	NS	NS	NS
B vs BV, with eosinophils counted in blood	NS	NS	NS
V vs BV, with eosinophils counted in sputum	NED	NS	NS
V vs BV, with eosinophils counted in blood	NED	NS	NS
(B, BE) vs (V, VE)	NED	NS	NS

AECOPD, acute exacerbations of COPD; B, bacterial infection; BE, bacterial infection with eosinophils; BV, bacterial/viral co-infection; COPD, chronic obstructive pulmonary disease; DEG, differentially-expressed genes; NED, not enough data to perform the analysis; NS, not significant; V, viral infection; VE, viral infection with eosinophils; vs, versus.

Comparison between pathogen species

The aim of these analyses is to find a set of DEGs that differentiate distinct types of infections detected by PCR in the sputum samples. For instance, we compared the whole blood gene expression in presence/absence of a specific pathogen in sputum (e.g. *H. influenzae*). We performed different analyses by further selecting samples with the absence of other bacteria and/or viruses.

To answer questions involving infection, we applied Models 3 and 4.

Moreover, we ran several analyses for DEG using the variable HL_Season and separately the $Season\_$ variable as $Season$ variable. These comparisons involved presence and interplay between *M. catarrhalis* and *H. influenzae.* However, we are reporting in this document the outcomes using the variable HL_Season in the model.

As shown in supplementary Table 4, there are significant DEG in two cases. The first one is the comparison between samples with *H. influenzae* and no *M. catarrhalis* and those in the opposite condition. There are significant DEG also when comparing co-infection with *H. influenzae* and *M. catarrhalis* and those samples with their absence.

Supplementary Table 4: Number of DEGs (FDR < 0.05) using the variable high/low (HL) season. The analysis [IV] uses three age categories (< 54 years, 54 years to 69 years, ≥ 70 years). The analysis [V] uses two age categories for exacerbations (<65 and ≥ 65 years).

ANALYSIS	ALL
Presence vs absence of Hi	NS
Presence vs absence of Mcat	NS
Presence vs absence of S. pneumoniae	NS
Presence vs absence of P. aeruginosa	NS
Presence vs absence of S. aureus	NS
Presence vs absence of Hi, no bacteria	NS
Presence vs absence of Mcat, no bacteria	NS
Presence vs absence of S. pneumoniae*, no bacteria*	NS
Presence vs absence of P. aeruginosa*, no bacteria*	NS
Presence vs absence of S. aureus*, no bacteria*	NED
Presence vs absence of Hi, no bacteria, no viruses	NS
Presence vs absence of Mcat, no bacteria, no viruses	NS
Presence vs absence of S. pneumoniae*, no bacteria, no viruses*	NS

Presence vs absence of P. aeruginosa*, no bacteria, no viruses*	NS
Presence vs absence of S. aureus*, no bacteria, no viruses*	NED
Difference Hi / Mcat	2122 probes, 1470 genes
Difference Hi / Mcat (Gene-wise)	1329 genes
Difference Hi / Mcat, no other bacteria	1947 probes, 1343 genes
Difference Hi / Mcat, no other bacteria (Gene-wise)	1195 genes
Difference Hi / Mcat, no other bacteria, no viruses	657 probes, 465 genes
Difference Hi / Mcat, no other bacteria, no viruses (Gene-wise)	364 genes
Difference Hi / Mcat, no viruses (Gene-wise) [IV]	603 genes
Difference Hi / Mcat, no viruses (Gene-wise) [V]	603 genes
(Hi^+, $Mcat^+$) vs (Hi^-, $Mcat^-$) (Gene-wise)	2559 genes
(Hi^+, $Mcat^+$) vs (Hi^-, $Mcat^-$) no other bacteria (Gene-wise)	2144 genes
(Hi^+, $Mcat^+$) vs (Hi^-, $Mcat^-$) no other bacteria, no viruses (Gene-wise)	1503 genes
(Hi^+, $Mcat^+$) vs (Hi^-, $Mcat^-$) no viruses (Gene-wise)	1965 genes

DEG, differentially-expressed genes; FDR, false discovery rate; Hi, *Haemophilus influenzae*; Mcat, *Moraxella catarrhalis*; NED, not enough data to perform the analysis; NS, not significant; vs, versus.

Supplementary section 5. Gene set enrichment analysis (GSEA)

For each contrast, we performed the GSEA with the genes passing IQR filtering, comparing the samples grouped by the contrast condition. In case of identification of DEGs, we also applied GSEA with the list of significant genes. Here we report the GSEA analyses corresponding to the identification of DEGs when *H. influenzae* (Hi) and *M. catarrhalis* (Mcat) infections (Supplementary tables 5 and 6) have an alternate occurrence or when they are concomitant (Supplementay tables 7 and 8). Enrichment analyses were repeated by averaging expression level when multiple probes correspond to the same gene.

Supplementary Table 5: Global gene-wise GSEA results for the interplay of Hi and Mcat. The rows are the biological functions from the Hallmark gene-set collection (MSigDB), each column in the table embodies a different analysis.

	(Hi+, Mcat-) vs (Hi-, Mcat+) Gene-wise		
	All (stable + exacerbated)		
Presence of other pathogen	*Any*	*No*	*No*
Presence of virus	*Any*	*Any*	*No*
Pathways and biological processes			
Allograft rejection			
Protein secretion			
E2F targets			
G2M checkpoint			
Cholesterol homeostasis			
DNA repair			
Spermatogenesis			
TNFA signaling via NFKB			
Hypoxia			
Glycolysis			
Complement			
Inflammatory response			
Xenobiotic metabolism			
PI3K/AKT/MTOR signaling			
MTORC1 signaling			
Unfolded protein response			
Estrogen response early			
KRAS signaling DN			
Oxidative phosphorylation			
IL-2/STAT5 signaling			
IL-6/JAK/STAT3 signaling			
Mitotic spindle			
Coagulation			
Heme metabolism			
Interferon gamma response			

Legend:

Not significant	
FDR < 1e-2	
FDR < 1e-4	
FDR < 1e-6	
FDR < 1e-8	
FDR < 1e-10	

Interferon alpha response			
UV response DN			
UV response UP			
Apoptosis			
Epithelial mesenchymal transition			
Estrogen response late			
Androgen response			
Myogenesis			
Peroxisome			

GSEA, gene set enrichment analysis; FDR, false discovery rate; Hi, *Haemophilus influenzae*; Mcat, *Moraxella catarrhalis*; MSigDB, molecular signatures database; vs, versus.

Supplementary Table 6: Gene-wise GSEA results for DEG linked to the interplay of Hi and Mcat. The rows are the biological functions from the Hallmark gene-set collection (MSigDB), each column in the table embodies a different analysis.

	(Hi+, Mcat-) vs (Hi-, Mcat+) Gene-wise		
	All (stable + exacerbated)		
Presence of other pathogen	*Any*	*No*	*No*
Presence of virus	*Any*	*Any*	*No*
Pathways and biological processes			
Allograft rejection			
Protein secretion			
E2F targets			
G2M checkpoint			
Cholesterol homeostasis			
DNA repair			
Spermatogenesis			
TNFA signaling via NFKB			
Hypoxia			
Glycolysis			
Complement			
Inflammatory response			
Xenobiotic metabolism			
PI3K/AKT/MTOR signaling			
MTORC1 signaling			
Unfolded protein response			
Estrogen response early			
KRAS signaling DN			
Oxidative phosphorylation			
IL-2/STAT5 signaling			
IL-6/JAK/STAT3 signaling			
Mitotic spindle			
Coagulation			
Heme metabolism			
Interferon gamma response			

Legend:

Not significant	
FDR < 1e-2	
FDR < 1e-4	
FDR < 1e-6	
FDR < 1e-8	
FDR < 1e-10	

Interferon alpha response			
UV response DN			
UV response UP			
Apoptosis			
Epithelial mesenchymal transition			
Estrogen response late			
Androgen response			
Myogenesis			
Peroxisome			

DEG, differentially-expressed gene; GSEA, gene set enrichment analysis; FDR, false discovery rate; Hi, *Haemophilus influenzae*; Mcat, *Moraxella catarrhalis*; MSigDB, molecular signatures database; vs, versus.

Supplementary Table 7: Global gene-wise GSEA results for the co-occurrence of Hi and Mcat. The rows are the biological functions from the Hallmark gene-set collection (MSigDB), each column in the table embodies a different analysis.

	(Hi+, Mcat-) vs (Hi-, Mcat+) Gene-wise		
	All (stable + exacerbated)		
Presence of other pathogen	*Any*	*No*	*No*
Presence of virus	*Any*	*Any*	*No*
Pathways and biological processes			
Allograft rejection			
Protein secretion			
E2F targets			
G2M checkpoint			
Cholesterol homeostasis			
DNA repair			
Spermatogenesis			
TNFA signaling via NFKB			
Hypoxia			
Glycolysis			
Complement			
Inflammatory response			
Xenobiotic metabolism			
PI3K/AKT/MTOR signaling			
MTORC1 signaling			
Unfolded protein response			
Estrogen response early			
KRAS signaling DN			
Oxidative phosphorylation			
IL-2/STAT5 signaling			
IL-6/JAK/STAT3 signaling			
Mitotic spindle			
Coagulation			
Heme metabolism			
Interferon gamma response			

Legend:

Not significant	
FDR < 1e-2	
FDR < 1e-4	
FDR < 1e-6	
FDR < 1e-8	
FDR < 1e-10	

Interferon alpha response			
UV response DN			
UV response UP			
Apoptosis			
Epithelial mesenchymal transition			
Estrogen response late			
Androgen response			
Myogenesis			
Peroxisome			

GSEA, Gene set enrichment analysis; EX, exacerbation; FDR, false discovery rate; Hi, *Haemophilus influenzae*; Mcat, *Moraxella catarrhalis*; MSigDB, molecular signatures database; vs, versus.

Supplementary Table 8: Gene-wise GSEA results for DEG linked to the co-occurrence of Hi and Mcat. The rows are the biological functions from the Hallmark gene-set collection (MSigDB), each column in the table embodies a different analysis.

	(Hi+, Mcat-) vs (Hi-, Mcat+) Gene-wise		
	All (stable + exacerbated)		
Presence of other pathogen	*Any*	*No*	*No*
Presence of virus	*Any*	*Any*	*No*
Pathways and biological processes			
Allograft rejection			
Protein secretion			
E2F targets			
G2M checkpoint			
Cholesterol homeostasis			
DNA repair			
Spermatogenesis			
TNFA signaling via NFKB			
Hypoxia			
Glycolysis			
Complement			
Inflammatory response			
Xenobiotic metabolism			
PI3K/AKT/MTOR signaling			
MTORC1 signaling			
Unfolded protein response			
Estrogen response early			
KRAS signaling DN			
Oxidative phosphorylation			
IL-2/STAT5 signaling			
IL-6/JAK/STAT3 signaling			
Mitotic spindle			
Coagulation			
Heme metabolism			
Interferon gamma response			

Legend:

Not significant	
FDR < 1e-2	
FDR < 1e-4	
FDR < 1e-6	
FDR < 1e-8	
FDR < 1e-10	

Interferon alpha response			
UV response DN			
UV response UP			
Apoptosis			
Epithelial mesenchymal transition			
Estrogen response late			
Androgen response			
Myogenesis			
Peroxisome			

DEG, differentially-expressed gene; GSEA, gene set enrichment analysis; FDR, False discovery rate; Hi, *Haemophilus influenzae*; Mcat, *Moraxella catarrhalis*; MSigDB, molecular signatures database; vs, versus.

Supplementary section 6. Content of the 17 gene communities identified by MNGCD for exacerbation and stable conditions

Group	Community	Entrez_ID	SYMBOL	GENE NAME	GENE TYPE
ST	616	5610	EIF2AK2	eukaryotic translation initiation factor 2 alpha kinase 2	protein-coding
ST	616	3437	IFIT3	interferon induced protein with tetratricopeptide repeats 3	protein-coding
ST	616	24138	IFIT5	interferon induced protein with tetratricopeptide repeats 5	protein-coding
ST	616	55281	TMEM140	transmembrane protein 140	protein-coding
ST	996	9636	ISG15	ISG15 ubiquitin like modifier	protein-coding
ST	996	4599	MX1	MX dynamin like GTPase 1	protein-coding
ST	535	23586	DDX58	DExD/H-box helicase 58	protein-coding
ST	535	4502	MT2A	metallothionein 2A	protein-coding
ST	535	83666	PARP9	poly(ADP-ribose) polymerase family member 9	protein-coding
ST	535	5359	PLSCR1	phospholipid scramblase 1	protein-coding
ST	535	7453	WARS1	tryptophanyl-tRNA synthetase 1	

ST	644	2038	EPB42	erythrocyte membrane protein band 4.2	protein-coding
ST	644	6622	SNCA	synuclein alpha	protein-coding
ST	644	7504	XK	X-linked Kx blood group	protein-coding
ST	599	1870	E2F2	E2F transcription factor 2	protein-coding
ST	599	26268	FBXO9	F-box protein 9	protein-coding
ST	599	23770	FKBP8	FKBP prolyl isomerase 8	protein-coding
ST	599	3029	HAGH	hydroxyacylglutathione hydrolase	protein-coding
ST	599	84333	PCGF5	polycomb group ring finger 5	protein-coding
ST	599	5305	PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase type 2 alpha	protein-coding
ST	1914	6521	SLC4A1	solute carrier family 4 member 1 (Diego blood group)	protein-coding
ST	1914	7145	TNS1	tensin 1	protein-coding
ST	261	759	CA1	carbonic anhydrase 1	protein-coding
ST	261	55363	HEMGN	hemogen	protein-coding
ST	261	7111	TMOD1	tropomodulin 1	protein-coding
ST	261	10107	TRIM10	tripartite motif containing 10	protein-coding
ST	875	2993	GYPA	glycophorin A (MNS blood group)	protein-coding
ST	875	6710	SPTB	spectrin beta, erythrocytic	protein-coding
ST	315	6352	CCL5	C-C motif chemokine ligand 5	protein-coding
ST	315	925	CD8A	CD8a molecule	protein-coding
ST	315	1521	CTSW	cathepsin W	protein-coding
ST	315	4818	NKG7	natural killer cell granule protein 7	protein-coding
ST	315	5196	PF4	platelet factor 4	protein-coding
ST	315	5450	POU2AF1	POU class 2 homeobox associating factor 1	protein-coding

ST	315	81539	SLC38A1	solute carrier family 38 member 1	protein-coding
EX	182	6790	AURKA	aurora kinase A	protein-coding
EX	182	332	BIRC5	baculoviral IAP repeat containing 5	protein-coding
EX	182	83879	CDCA7	cell division cycle associated 7	protein-coding
EX	182	1058	CENPA	centromere protein A	protein-coding
EX	182	55165	CEP55	centrosomal protein 55	protein-coding
EX	182	2553	GABPB1	GA binding protein transcription factor subunit beta 1	protein-coding
EX	182	3832	KIF11	kinesin family member 11	protein-coding
EX	182	4171	MCM2	minichromosome maintenance complex component 2	protein-coding
EX	182	51203	NUSAP1	nucleolar and spindle associated protein 1	protein-coding
EX	182	9232	PTTG1	PTTG1 regulator of sister chromatid separation, securin	protein-coding
EX	182	6241	RRM2	ribonucleotide reductase regulatory subunit M2	protein-coding
EX	182	7153	TOP2A	DNA topoisomerase II alpha	protein-coding
EX	182	7298	TYMS	thymidylate synthetase	protein-coding
EX	182	11130	ZWINT	ZW10 interacting kinetochore protein	protein-coding
EX	78	246	ALOX15	arachidonate 15-lipoxygenase	protein-coding
EX	78	80150	ASRGL1	asparaginase and isoaspartyl peptidase 1	protein-coding
EX	78	25825	BACE2	beta-secretase 2	protein-coding
EX	78	100133941	CD24	CD24 molecule	protein-coding
EX	78	928	CD9	CD9 molecule	protein-coding
EX	78	1053	CEBPE	CCAAT enhancer binding protein epsilon	protein-coding
EX	78	56994	CHPT1	choline phosphotransferase 1	protein-coding
EX	78	79843	FAM124B	family with sequence similarity 124 member B	protein-coding

EX	78	2263	FGFR2	fibroblast growth factor receptor 2	protein-coding
EX	78	2624	GATA2	GATA binding protein 2	protein-coding
EX	78	117245	PLAAT5	phospholipase A and acyltransferase 5	
EX	78	8739	HRK	harakiri, BCL2 interacting protein	protein-coding
EX	78	9173	IL1RL1	interleukin 1 receptor like 1	protein-coding
EX	78	3568	IL5RA	interleukin 5 receptor subunit alpha	protein-coding
EX	78	10365	KLF2	Kruppel like factor 2	protein-coding
EX	78	10215	OLIG2	oligodendrocyte transcription factor 2	protein-coding
EX	78	5376	PMP22	peripheral myelin protein 22	protein-coding
EX	78	22836	RHOBTB3	Rho related BTB domain containing 3	protein-coding
EX	78	6196	RPS6KA2	ribosomal protein S6 kinase A2	protein-coding
EX	78	2030	SLC29A1	solute carrier family 29 member 1 (Augustine blood group)	protein-coding
EX	78	55512	SMPD3	sphingomyelin phosphodiesterase 3	protein-coding
EX	78	6652	SORD	sorbitol dehydrogenase	protein-coding
EX	78	201305	SPNS3	sphingolipid transporter 3 (putative)	protein-coding
EX	532	55601	DDX60	DExD/H-box helicase 60	protein-coding
EX	532	51191	HERC5	HECT and RLD domain containing E3 ubiquitin protein ligase 5	protein-coding
EX	532	9636	ISG15	ISG15 ubiquitin like modifier	protein-coding
EX	532	4599	MX1	MX dynamin like GTPase 1	protein-coding
EX	532	91543	RSAD2	radical S-adenosyl methionine domain containing 2	protein-coding
EX	532	54739	XAF1	XIAP associated factor 1	protein-coding
EX	611	5610	EIF2AK2	eukaryotic translation initiation factor 2 alpha kinase 2	protein-coding

EX	611	2634	GBP2	guanylate binding protein 2	protein-coding
EX	611	3428	IFI16	interferon gamma inducible protein 16	protein-coding
EX	611	3437	IFIT3	interferon induced protein with tetratricopeptide repeats 3	protein-coding
EX	611	24138	IFIT5	interferon induced protein with tetratricopeptide repeats 5	protein-coding
EX	611	84628	NTNG2	netrin G2	protein-coding
EX	611	55281	TMEM140	transmembrane protein 140	protein-coding
EX	530	23586	DDX58	DExD/H-box helicase 58	protein-coding
EX	530	2633	GBP1	guanylate binding protein 1	protein-coding
EX	530	4502	MT2A	metallothionein 2A	protein-coding
EX	530	83666	PARP9	poly(ADP-ribose) polymerase family member 9	protein-coding
EX	530	5359	PLSCR1	phospholipid scramblase 1	protein-coding
EX	530	219285	SAMD9L	sterile alpha motif domain containing 9 like	protein-coding
EX	530	6398	SECTM1	secreted and transmembrane 1	protein-coding
EX	530	6772	STAT1	signal transducer and activator of transcription 1	protein-coding
EX	530	7130	TNFAIP6	TNF alpha induced protein 6	protein-coding
EX	530	7453	WARS1	tryptophanyl-tRNA synthetase 1	
EX	940	3433	IFIT2	interferon induced protein with tetratricopeptide repeats 2	protein-coding
EX	940	54625	PARP14	poly(ADP-ribose) polymerase family member 14	protein-coding
EX	99	118932	ANKRD22	ankyrin repeat domain 22	protein-coding
EX	99	80830	APOL6	apolipoprotein L6	protein-coding
EX	99	116071	BATF2	basic leucine zipper ATF-like transcription factor 2	protein-coding
EX	99	29126	CD274	CD274 molecule	protein-coding
EX	99	26270	FBXO6	F-box protein 6	protein-coding

EX	99	257019	FRMD3	FERM domain containing 3	protein-coding
EX	99	64135	IFIH1	interferon induced with helicase C domain 1	protein-coding
EX	99	113730	KLHDC7B	kelch domain containing 7B	protein-coding
EX	99	4061	LY6E	lymphocyte antigen 6 family member E	protein-coding
EX	99	4343	MOV10	Mov10 RISC complex RNA helicase	protein-coding
EX	99	57674	RNF213	ring finger protein 213	protein-coding
EX	99	64108	RTP4	receptor transporter protein 4	protein-coding
EX	99	710	SERPING1	serpin family G member 1	protein-coding
EX	1931	6622	SNCA	synuclein alpha	protein-coding
EX	1931	7504	XK	X-linked Kx blood group	protein-coding

Acronyms

AECOPD: Acute exacerbations of COPD

BMI: Body mass index

COPD: Chronic obstructive pulmonary disease

DEG: Differentially-expressed genes

EX: Exacerbation

FC: Fold change

FDR: False discovery rate

GSEA: Gene set enrichment analysis

IQR: Interquartile range

LIMMA: Linear models for microarrays data

MNGCD: Multi-network gene community detection

MSigDB: Molecular signatures database

QC: Quality control

ST: Stable

1. References

1 De Domenico, M., Lancichinetti, A., Arenas, A. & Rosvall, M. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Physical Review X* **5**, 011027 (2015).

2 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

3 Gerstein, M. B. et al. Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91-100 (2012).

4 Chou, C. H. et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **46**, D296-D302 (2018).

5 Alonso-López, D. et al. APID interactomes: providing proteome-based interactomes with controlled quality for multiple species and derived networks. *Nucleic Acids Res.* **44**, W529-535 (2016).

6 Wani, N. & Raza, K. Integrative approaches to reconstruct regulatory networks from multi-omics data: A review of state-of-the-art methods. *Comput. Biol. Chem.* **83**, 107120 (2019).

7 Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1118-1123 (2008).

8 Lancichinetti, A. & Fortunato, S. Consensus clustering in complex networks. *Sci. Rep.* **2**, 336 (2012).

9 Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).