# Exploring the Diseasome of COPD and its associated diseases

Solène Grosdidier[1], Antoni Ferrer[2], Rosa Faner[3], Joaquim Gea[2], Janet Piñero[1], Josep Roca[3], Borja Cosío[4], Alvar Agustí[3], Ferran Sanz[1], Laura I Furlong[1]

*[1] Integrative Biomedical Informatics Group, Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Research Institute (IMIM), Universitat Pompeu Fabra, Barcelona, Spain; [2] Servei de Pneumologia, Hospital del Mar, Universitat Pompeu Fabra, CIBERES, Barcelona, Spain; [3] Hospital Clínic, Universitat Barcelona, FISIB, CIBERES, Barcelona, Spain; [4] Hospital Son Espases, CIBERES, Palma, Spain.*

Correspondece: lfurlong@imim.es

Chronic Obstructive Pulmonary Disease (COPD) is a major health problem because of its prevalence (10% of adult population), raising incidence (WHO predicts that it will be the 3rd cause of death in the world by 2020), and associated morbi-mortality and socio-economic cost. Patients with COPD often suffer concomitant disorders (comorbidity) that impact their health-status and prognosis in a significant manner. Several mechanisms linking COPD to its associated disorders have been proposed, such as systemic inflammation, physical inactivity, chronic hypoxia, shared genetic predisposition and decreased synthesis of anti-aging molecules, but more research is needed to understand the links between these diseases and search for common treatable components [1].

It has been proposed that the genetic origin of diseases is correlated with comorbidities [2] meaning that, in some cases, comorbid diseases have a common genetic basis. Otherwise, the gene products associated with both diseases tend to interact together or belong to the same pathway within the cell [3-4]. Understanding these relationships can lead to crucial insights to improve the prevention, diagnosis, prognosis and/or treatment of these diseases. In this study we explored the hypothesis that COPD and concomitant disorders share etiologic factors, such as common genes and/or pathways.

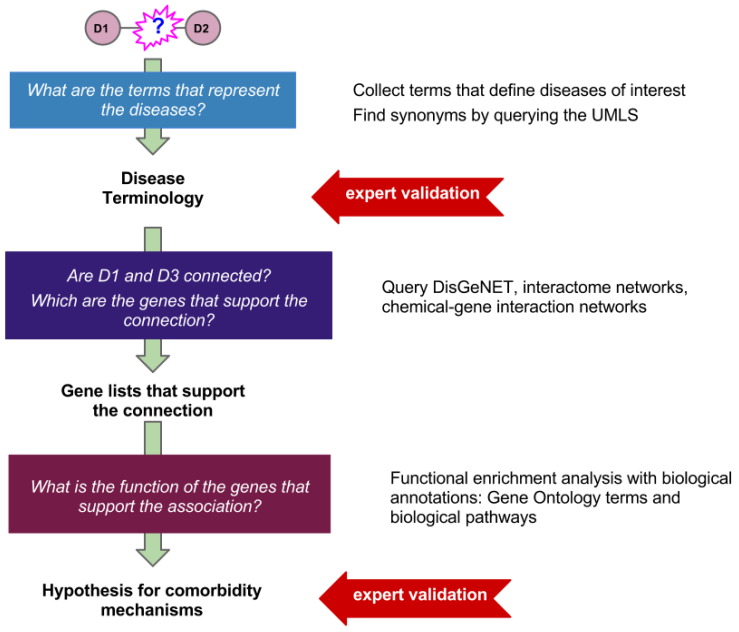


*Figure 1. Schematic representation of the systems biology framework proposed for the exploration of the Diseasome of COPD.*

To evaluate this hypothesis, we developed a systems biology framework to explore the associations between COPD and the most relevant comorbid conditions based on publicly available information on gene-disease associations and molecular interactions, controlled vocabularies for disorders, gene functional annotations and network analysis (Figure 1). Specifically, DisGeNET [5], a database collecting information on diseases and its associated genes for a broad range of disorders, was used to generate gene-disease (GDN) and disease (DN) association networks. In the disease network (DN), two diseases are connected if they either share at least one associated gene in the GDN or if its associated proteins are known to interact. Interactome networks were

derived from large-scale datasets of protein-protein interactions to extract pairs of interacting proteins [6-7]. The strength of association between two disorders was estimated by the Jaccard coefficient. The Unified Medical Language System (UMLS, http://www.nlm.nih.gov/research/umls/) was used to manage diverse disease terminologies used in different scenarios (*i.e.* clinical setting *vs.* biomedical literature). Information about chemical-gene association was obtained from the Comparative Toxicogenomics Database [8]. The CTD database contains annotations on the effect of environmental chemicals on genes and proteins. This dataset was combined with gene-disease associations derived from DisGeNET in chemical-gene-disorder networks (Figure 2). Functional enrichment analysis was carried out to characterize gene lists that support association between diseases by their annotation to Gene Ontology (GO) terms and biological pathways. The framework is implemented as bioinformatics modules and the results are displayed by a web interface to allow clinical researchers to evaluate the intermediate results at specific points of the workflow (red arrows in Figure 1).

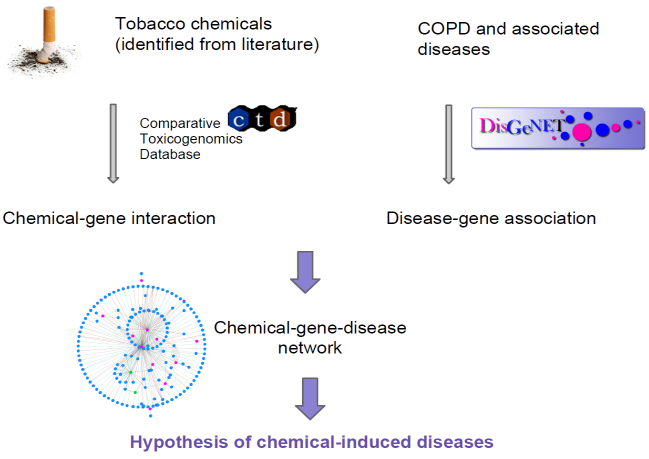


*Figure 2. Development of chemical-gene-disease networks to explore the effect of chemical compounds of tobacco smoke on COPD and its associated diseases.*

When we applied this framework to the particular case of COPD and its association with 21 clinically-relevant associated conditions, including ischemic heart disease, muscle weakness, depression and lung cancer [9], we found that COPD shares genes and protein interactions with 20 of its associated diseases. The genes that support the association between COPD and its associated diseases were further characterized by functional enrichment analysis using GO concepts and biological pathways to detect main biological processes signatures. The functional annotation of shared genes highlighted relevant biological processes such as inflammation, apoptosis and hypoxia, among others. Moreover, in some selected cases, the participation of biological processes related with aging was also found. In addition, we evaluated the contribution of environmental factors such as chemical products present in cigarettes smoke [10-11] by exploring the chemical-gene-disease network for each comorbid disease pair, and found disease-associated proteins that are targeted by several of these compounds. For 106 chemicals from cigarettes smoke we found interactions for 61 of such chemicals with 101 genes that support the association of COPD with comorbid disorders. The gene-chemical-disease network was explored to derive mechanistic hypothesis.

To our knowledge, this is the first approach to systematically explore disease comorbidities of COPD by a combination of data mining and network analysis of chemical-gene-disease associations. Previous studies based on gene expression analysis were focused on the linkage between COPD and muscle abnormalities [12]. This approach generates testable hypothesis for the biological mechanisms underlying COPD and its associated comorbidities. Finally, this framework constitutes a valid approach that can be applied to the study of the mechanisms underlying other prevalent disorders characterized also by the presence of comorbidity.

Acknowledgments: This work was supported by the Instituto de Salud Carlos III FEDER (CP10/00524), COMBIOMED and SAF2011-26908 grants. The Research Programme on Biomedical Informatics (GRIB) is a node of the Spanish National Institute of Bioinformatics (INB).

**References**

1. Barnes JP. Chronic Obstructive Pulmonary Disease: Effects beyond the Lungs. PLoS Med. 2010;7:e1000220

2. Park J, Lee DS, Christakis NA, Barabási AL. The impact of cellular networks on disease comorbidity. Mol

Syst Biol. 2009;5, 262

3. Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabási AL. The implications of human metabolic network topology for disease comorbidity. PNAS. 2008;105:9880-9885

4. Park S, Yang JS, Shin YE, Park J, Jang SK, Kim S. Protein localization as a principal feature of the etiology and comorbidity of genetic diseases. Mol Syst Biol. 2011;7:494

5. Bauer-Mehren A, Rautschka M, Sanz F, Furlong LI. DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. Bioinformatics. 2010;26:2924-2926

6. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, et al. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res. 2011;39:D685-90

7. Schaefer MH, Fontaine JF, Vinayagam A, Porras P, Wanker EE, Andrade-Navarro MA. HIPPIE: Integrating protein interaction networks with experiment based quality scores. PLoS One. 2012;7:e31826

8. Davis A P, Wiegers TC, Rosenstein MC, Murphy CG, Mattingly CJ. The curation paradigm and application tool used for manual curation of the scientific literature at the Comparative Toxicogenomics Database. 2011;bar034

9. Agusti A, Calverley PM, Celli B, Coxson HO, Edwards LD, Lomas DA, et al. Characterisation of COPD heterogeneity in the ECLIPSE cohort. Respiratory Research. 2010;11:122-136

10. Rustemeier K, Stabbert R, Haussmann HJ, Roemer E, Carmines EL. Evaluation of the potential effects of ingredients added to cigarettes. Part 2: chemical composition of mainstream smoke. Food and chemical toxicology. 2002;40:93-104

11. Hoffmann D, Hoffmann I. Tobacco smoke components. Beiträge zur Tabakforschung International. 1998;18:49-52

12. Turan N, Kalko S, Stincone A, Clarke K, Sabah A, Howlett K, et al. A systems biology approach identifies molecular networks defining skeletal muscle abnormalities in chronic obstructive pulmonary disease. PLoS Comput Biol. 2011;7:e1002129