A comprehensive (biological and computational) investigation on the role of microRNA::mRNA regulations performed in chronic obstructive pulmonary disease and lung cancer

Jingshan Huang†
Computer Science Department, School of Computing
University of South Alabama
Mobile, Alabama 36688, U.S.A.

Dejing Dou
Computer and Information Science Department
University of Oregon
Eugene, Oregon 97403, U.S.A.

Jun She
Department of Pulmonary Medicine, Affiliated Zhongshan Hospital of Fudan University, Shanghai, 200032, China
Division of Epidemiology, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, U.S.A.

Andrew H. Limper
Pulmonary and Critical Care Medicine
Mayo Clinic College of Medicine
Rochester, Minnesota 55905, U.S.A.

Yanan Yang
Pulmonary and Critical Care Medicine
Mayo Clinic College of Medicine
Rochester, Minnesota 55905, U.S.A.

Ping Yang
Division of Epidemiology
Mayo Clinic College of Medicine
Rochester, Minnesota 55905, U.S.A.

***Abstract*—Chronic obstructive pulmonary disease (COPD) and lung cancer (LC) are two serious diseases that present a major health problem worldwide. However, genetic contribution to both diseases remains unclear, including various regulation mechanisms at genetic level resulting in the progression from COPD to LC. In this paper, we describe our comprehensive methodologies, which seamlessly integrate both biological (conducted in "wet labs") and computational (based on domain ontologies and semantic technologies) approaches, to investigate the important role of microRNA::mRNA regulations performed in COPD and LC. We discovered two genes, RGS6 and PARK2, that are strongly associated with the risk of developing either COPD or LC or both; additionally, we also identified two sets of microRNAs that are computationally predicted to regulate RGS6 and PARK2, respectively. These microRNAs can be further biologically verified in the future and serve as novel biomarkers in COPD and/or LC.**

***Keywords*—chronic obstructive pulmonary disease, COPD, lung cancer, microRNA, target gene, bio-ontology, semantic search.**

I. Introduction

Chronic obstructive pulmonary disease (COPD) and lung cancer (LC) combined present a major global health problem. COPD is a major cause of severe long-term disability and ranks third as the leading cause of death in the United States. The mortality rate of COPD has steadily increased over the past ten years, accounting for approximately 5% of all deaths in the states [1]. LC has been the most common malignant disease with high morbidity and mortality for several decades worldwide [2–4]. The overall 5-year survival rate of LC is 16%, resulting in more cancer death in the states than any other malignancy [5]. In addition, COPD coexists in 40% to 70% of patients with LC and is increasingly recognized as a predecessor of LC development [6]. Indeed, COPD is a strong and independent risk factor for developing LC, even after controlling for smoking exposure. Further, accumulating evidence [7–9] indicates that the presence of COPD independently increases the risk of LC as much as ten-fold.

Whereas past investigations have enhanced our understanding of the links between COPD and the development of LC, further studies are desperately needed because the mechanisms that regulate these two diseases still remain largely unclear. In particular, new approaches that can further facilitate our understanding of the genetic, genomic, and epigenetic bases of these disease processes will be required to (1) better explore the associations between these two deadly diseases, (2) identify various regulation mechanisms at genetic level that lead to the progression from COPD to LC, (3) more readily define patients at greatest risk, and (4) create additional strategies to prevent and treat these two serious diseases.

Towards this end, it is necessary to develop more advanced methodologies that are capable of integrating both biological and computational/bioinformatics approaches in a seamless manner; it will then be possible for us to explore a more accurate representation of biological processes that regulate both diseases and the progression from COPD to LC. That said, we report in this paper our on-going efforts to effectively combine biological experiments (conducted in "wet labs") and software analysis (based upon domain ontologies and semantic technologies), so that to better investigate the important role performed by regulations between microRNAs (miRs) and messenger RNAs (mRNAs) in COPD and LC.

†Corresponding author (Email: huang@southalabama.edu)

978-1-5090-1610-5/16/$31.00 ©2016 IEEE

The rest of this paper is organized as follows. Section II summarizes state-of-the-art research in COPD, LC, and semantic technologies, respectively; Section III describes our comprehensive methodologies; Section IV reports our current findings along with discussion; and finally, Section V concludes with important future work.

II. Related Work

A. Related work in COPD and LC

COPD and LC have attracted substantial research attention over the past few decades due to the potential links between these two diseases, as well as shared versus unique biological mechanisms.

While effective screening methods have remained challenging, computer tomography (CT) has demonstrated promising as a useful means of screening for cancer. The National Lung Cancer Screening Trial (NLCST) showed reduced mortality in a high-risk population that received screening by low-dose CT [10]. However, this method was associated with an unacceptably high level of false positive findings (around 95%), resulting in extensive cost and repeated imaging and procedures to follow patients that ultimately proved to have benign disease. NLCST illustrates the on-going need to better understand risk factors for LC development so that clinicians can more precisely define individuals that are at increased risk.

The relationship between obstructive airway disease and LC has been increasingly well defined, but the clinical entity of COPD includes more than airflow obstruction alone. Specifically, COPD encompasses a broad variety of clinical phenotypes, and various phenotypes are characterized by differing degrees of airflow obstruction, emphysema, and chronic bronchitis [11]. Until recent years, the relationship between emphysematous destruction of lung tissue and LC was not entirely clear. With the development of advanced imaging techniques, CT has become the preferred diagnostic approach for non-invasive detection and quantification of emphysema. Based on visual assessment of CT scans, using semi-quantitative emphysema assessments has allowed investigators to demonstrate that the presence of emphysema itself is also an independent risk factor for developing LC [12–14]. It was shown that patients with emphysematous lung tissue demonstrated on CT scans were roughly three-fold more likely to develop LC compared to individuals without any evidence emphysema. In fact, the association of LC with emphysema persisted even in those individuals lacking airflow obstruction on lung function testing.

A number of theories have been proposed [15] to help understand how COPD and LC can be linked at the pathogenic level. One possibility involves neutrophil elastase, a protease capable of degrading elastin. Alpha-1 anti-trypsin serves as the native anti-protease that binds and thereby neutralizes neutrophil elastase. If neutrophil elastase acts in an unopposed fashion, such as occurs in patients with alpha-1 anti-trypsin deficiency, emphysematous destruction of the lung is the consequence. Another possibility involves the family of enzymes known as matrix metalloproteinases (MMPs), which are responsible for the degradation of collagen and other components of the extracellular matrix. Alveolar macrophages and neutrophils are known [15–17] to release increased amounts of MMPs as a result of elevated levels of inflammatory cytokines including IL-1β and TNF-α resulting from activation of the nuclear factor NF-kβ signaling pathway. Further, IL-1β and TNF-α have been shown [16] to be present at elevated levels in the sputum of patients with COPD. Therefore, it appears that the inflammatory environment present in the lungs of patients with COPD likely supports the development of LC.

B. Related work in semantic technologies

In biological and biomedical investigation, it, more often than not, requires an ability to integrate and analyze large amounts of data distributed across different sources such as gene-related information, various phenotypic data, PubMed publications, and so forth. These data sources usually have quite heterogeneous semantics (intended biological meanings) among one another. Thus, it is necessary to seek assistance from domain ontologies and semantic technologies, which have been proven, particularly within biological and biomedical research, to promote more precise communication among scientists, enable more effective information retrieval and integration, and better facilitate knowledge acquisition out of original data.

1) Related work in bio-ontologies: As a foundation of semantic technologies, ontologies have been widely utilized in biological, biomedical, and clinical research. We briefly describe relevant biological and biomedical ontologies (short for bio-ontologies) as follows. Gene Ontology (GO) [18] is by far the most successful and widely used bio-ontology, consisting of three independent sub-ontologies: biological processes, molecular functions, and cellular components. The GO has been utilized to annotate gene products of model organisms including Homo sapiens. RNA Ontology (RNAO) [19] is an Open Biological and Biomedical Ontologies (OBO) [20] foundry reference ontology to catalogue the molecular entities composing primary, secondary, and tertiary components of RNA. The goal of the RNAO project is to enable integration and analysis of diverse RNA datasets. Non-Coding RNA Ontology (NCRO) [21] [22] is an OBO candidate reference ontology in non-coding RNA (ncRNA) domain. The NCRO was developed to provide a precisely defined ncRNA controlled vocabulary, which can fill a specific and highly needed niche in unification of ncRNA biology. Ontology for MIcroRNA Target (OMIT) [23–25] is an application ontology in OBO Library. This miR domain-specific ontology aims to provide the community with common data elements and data exchange standards in the miR research, especially the research exploring miR target genes.

2) Related work in semantic search: Our investigation in this paper is closely related to semantic search, which is a research field that intends to improve the access to contents by considering the semantics behind the search process [26]. In other words, semantic search goes beyond keyword-based search by considering contextual semantics of words, the intend of the user, and the search space. In general, semantic search requires the use of structured knowledge in the modeling and interpretation of queries. One popular way to represent structured knowledge in ontologies is by formal logic, such as first order logic, description logic, and frame logic. Because ontologies can help improve the search by query expansion, ontology-based semantic search can be

treated as one sub-area of semantic data mining [27]. One main idea in many semantic search systems (e.g., [28–32]) is, the original set of query keywords can be expanded by drawing on synonyms and other relationships (e.g., subclass and parthood) that are not part of the query. For example, in the work by Chauhan et al. [32], the original query was first expanded by considering synonyms, then terms with high semantic similarity were chosen from the ontology to be integrated to the search query, and the semantic similarity used for the query expansion was computed by the distance among concepts in the ontology, the position in the hierarchy, and the number of upper classes. Another way to implement semantic search is to use ontologies to translate keyword-based search into formal semantic queries. For example, Tran et al. [26] used a set of models (mental, user, system, and query) to capture information, such as thought entities, language primitives, knowledge representation (KR) primitives, and query elements. These models were then combined with a set of assumptions to redefine original queries, filling the gap between terms with structural information from an ontology. That is, each term within the query was considered a property of another term.

III. Methodologies

Our methodologies consist of three steps.

1) *Step One: Validation study based on genome-wide single nucleotide polymorphism based analysis (SNP-GWA).* Subjects will be recruited from four different groups: LC-only, LC-and-COPD, COPD-only, and controls. We will then conduct a comprehensive validation study on these subjects to tease apart genes among candidate genes that have been associated with COPD, or LC, or both. To be more specific, we will perform redundancy analysis of linkage disequilibrium, and the biological roles will be elucidated by transcript expression quantitative trait loci (eQTL), expression difference between tumor and normal lung, and pathway analyses, along with allele-specific risks assessed by both odds ratio (OR) and 95% confidence interval (CI).
2) *Step Two: Semantic integration and search through OmniSearch software tool.* OmniSearch [25] [33] [34] was developed based upon domain ontologies and semantic technologies, aiming to handle the significant challenge of miR-related data integration and semantic search query. Three miR target prediction databases (miRDB [35], TargetScan [36], and miRanda [37]) and one miR validated target database (miRTarBase [38]) are integrated in OmniSearch. This software tool provides with users a "one-stop" visit that enables a convenient, side-by-side comparison among prediction results from numerous miR target prediction databases, as well as providing access to other valuable, relevant data sources such as GO annotations and PubMed publications. Our previous research has demonstrated that OmniSearch has many advantages over conventional miR target search [25] [33], especially with regard to the accuracy (effectiveness) and efficiency of software output. For genes resulted from the analysis in *Step One*, we will utilize OmniSearch to obtain a list of miRs that have been computationally predicted to regulate these genes.
3) *Step Three: Biological validation.* We will conduct wet-lab biological experiments to validate the regulation mechanism of computationally predicted miR::mRNA pairs output from *Step Two*. Respective 3'-UTR luciferase reporter constructs will be generated for all mRNAs in the list of miR::mRNA pairs obtained from *Step Two*, we will then perform reporter assays to confirm the suppression of target genes (i.e., target mRNAs) by their computationally paired miRs. Following that, we will assess the role of these miR::mRNA pairs in the malignant transformation of lung epithelial cells using soft agar arrays. To be more specific, we will transfect the validated miRs into BEAS2B and HBEC3 lung epithelial cells to determine whether or not they promote the soft agar colony formation.

IV. Results and Discussion

A. SNP-GWA analysis results

Our study was conducted in a Caucasian population with 2,523 subjects, including LC-only (n = 612), LC-and-COPD (n = 612), COPD-only (n = 537), and controls (n = 762). We tested a total of 4,491 SNPs out of 304 candidate genes and identified the following eight candidate genes:

- LC-with-COPD: RGS6, PARK2, FOXP1, and IRF2;
- LC-from-COPD: RGS6, PARK2, COL4A3, and MACRQD2;
- COPD-only: RGS6 and IL10; and
- LC-only: CCND1.

Among these eight genes, 11 SNPs were validated. A SNP in RGS6 was inversely associated with COPD without LC (OR = 0.47; 95% CI, 0.31-0.72) or with LC (OR = 0.40; 95% CI, 0.27-0.60), supported by eQTLs (P = 0.02). A SNP in PARK2 was inversely associated with LC that developed from COPD patients (OR = 1.65; 95% CI, 1.17-1.78), with lower PARK2 transcript levels in tumor than normal lung tissues (P = 0.01).

Fig. 1 demonstrates the following results:

- Fig. 1A: RGS6 and PARK2 are associated with both COPD+LC vs. controls (healthy) and COPD-only; whereas RGS6 by itself is also associated with COPD vs. controls.
- Fig. 1B: IL10 and RGS6 are related to immunity cell function from healthy to COPD; PARK2, COL4A3, MACROD2, and CCND1 are related to cellular development from COPD to LC; CASP7, EGFR, and TERT are related to cell death and survival from healthy to LC; and RGS6, PARK2, FOXP1, and IRF2 are related to microtubule destabilization from healthy to COPD+LC.
- Fig. 1C: RGS6 expression by rs2238233 genotype.
- Fig. 1D: Impaired mitoses including multipolar spindles, misalignment, and abnormal microtubule in PARK2 knockout (KO) cells.[1]

[1]PARK2-related pilot work that is closely related to the research reported in this paper was published in our previous research [39].

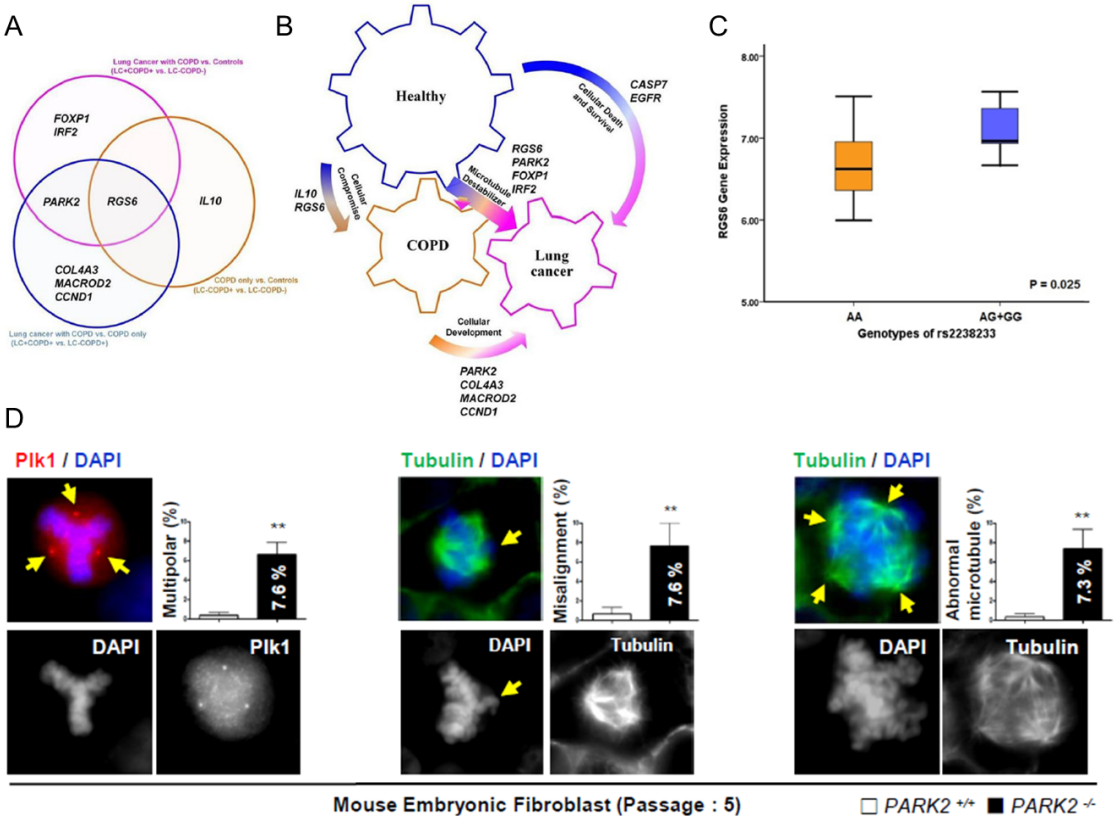

Fig. 1. PARK2 and RGS6 associate with the progression from COPD to lung cancer.

TABLE I. A SUBSET (33 OUT OF 159) OF OMNISEARCH RESULTS FOR GENE RGS6

miR name	Score from miRDB	Score from TargetScan	Score from miRanda
hsa-miR-137	53.56	22	0.7924
hsa-miR-128-3p	89.03	97	0
hsa-miR-216a-3p	61.57	96	0
hsa-miR-221-3p	73.11	18	0
hsa-miR-222-3p	87.19	18	0
hsa-miR-3681-3p	61.57	97	0
hsa-miR-190b	0	0	0.3285
hsa-miR-199a-5p	0	0	0.8531
hsa-miR-199b-5p	0	0	0.8496
hsa-let-7c-5p	0	17	0
hsa-miR-1184	61.90	0	0
hsa-miR-1199-5p	73.55	0	0
hsa-miR-1275	85.19	0	0
hsa-miR-1291	58.87	0	0
hsa-miR-302e	0	0	0.1404
hsa-miR-30a-5p	0	49	0
hsa-miR-3180-5p	58.15	0	0
hsa-miR-339-5p	0	0	0.1029
hsa-miR-33b-3p	51.03	0	0
hsa-miR-363-5p	80.59	0	0
hsa-miR-3679-5p	66.05	0	0
hsa-miR-3689d	82.26	0	0
hsa-miR-4458	0	16	0
hsa-miR-4459	56.41	0	0
hsa-miR-4500	0	18	0
hsa-miR-4508	59.05	0	0
hsa-miR-4520-2-3p	63.75	0	0
hsa-miR-4524b-3p	62.87	0	0
hsa-miR-455-5p	0	0	0.3287
hsa-miR-4651	99.74	0	0
hsa-miR-4663	74.28	0	0
hsa-miR-520c-3p	0	0	0.1377
hsa-miR-615-3p	0	0	0.1390

B. OmniSearch analysis results

The friendly user interface of OmniSearch allows the search for putative target genes given a miR, as well as the search for putative miRs given a target gene. Users can also

TABLE II. A SUBSET (23 OUT OF 72) OF OMNISEARCH RESULTS FOR GENE PARK2

miR name	Score from miRDB	Score from TargetScan	Score from miRanda
hsa-miR-320a	60.55	0	0.8533
hsa-miR-320b	60.55	0	0.8533
hsa-miR-320c	60.55	0	0.8533
hsa-miR-320d	60.55	0	0.8533
hsa-miR-107	0	0	0.1918
hsa-miR-1227-5p	85.38	0	0
hsa-miR-125a-3p	0	0	0.2020
hsa-miR-129-5p	0	0	0.1907
hsa-miR-1299	64.64	0	0
hsa-miR-140-5p	0	0	0.4894
hsa-miR-146b-5p	0	0	0.2156
hsa-miR-17-3p	50.00	0	0
hsa-miR-181a-5p	0	85	0
hsa-miR-199a-5p	0	0	0.1383
hsa-miR-199b-5p	0	0	0.1369
hsa-miR-29a-5p	60.40	0	0
hsa-miR-3154	52.84	0	0
hsa-miR-384	0	0	0.1375
hsa-miR-4311	88.69	0	0
hsa-miR-4755-5p	77.92	0	0
hsa-miR-496	0	0	0.7973
hsa-miR-6791-5p	60.97	0	0
hsa-miR-543	0	0	0.9246

TABLE III. STATISTICS OF OMNISEARCH RESULTS FOR BOTH RGS6 AND PARK2

	RGS6	PARK2
Total number of putative targeting miRs	159	72
Number of targeting miRs predicted by all three databases	1	0
Number of targeting miRs predicted by any two databases	5	4
Number of targeting miRs predicted by only one database	153	68

perform additional analysis, e.g., DAVID and PANTHER, on returned targets, and the interface provides a convenient way to save search results in different formats (e.g., tab-delimited text and CSV). In this study, the OmniSearch software returned a total of 159 and 72 computationally putative miRs for genes RGS6 and PARK2, respectively. Some of these results are demonstrated in Table I and Table II, with their statistics summarized in Table III.

Based on the above results, we further analyzed as follows:

- For RGS6: Our focus was placed on hsa-miR-137, hsa-miR-221-3p, and hsa-miR-222-3p. In particular, hsa-miR-137 was predicted by all three prediction databases but not yet biologically validated — this information was provided in the OmniSearch tool. The other two miRs are also of high importance because they were predicted by two different prediction databases, and, similarly, not yet validated.
- For PARK2: Our focus was placed on four miRs: hsa-miR-320a, hsa-miR-320b, hsa-miR-320c, and hsa-miR-320d. These four miRs were all predicted by at least two different prediction databases but not yet biologically validated.
- An immediate piece of future work is to conduct biological validation on those miRs reported above. Upon completion of validation experiments, we will be able to identify novel biomarkers in COPD and LC.

V. CONCLUSIONS

COPD and LC together present a major global health problem, but genetic regulation mechanisms on both diseases remains unclear. We described in this paper our comprehensive methodologies to seamlessly integrate both biological and computational approaches in order to fully investigate the important role of miR::mRNA regulations performed in COPD and/or LC, as well as in the progression from COPD to LC. We discovered that RGS6 and PARK2 are strongly associated with the risk of developing either COPD or LC or both. Further, we utilized our methodologies to successfully identify a set of miRs that are computationally predicted to regulate these two genes. Once the biological functions of these miRs are verified by reporter and soft agar assays, they can highly likely serve as novel biomarkers in COPD and/or LC. In particular, our current findings indicated that hsa-miR-137::RGS6 appeared to be a rather promising pair for further investigation.

Besides the biological validation to be performed (discussed in Section IV-B), another interesting piece of future work is to incorporate into our methodologies the analysis of next-generation (NG) sequencing data, which have profoundly altered our understanding of biology, human diversity, and disease, thus revolutionizing genomic and transcriptomic research over the past decade.

ACKNOWLEDGMENT

Research reported in this paper was partially supported by the the National Institutes of Health (NIH), under the Award Numbers U01CA180982, R01CA80127, R01CA84354, and R01HL107612. The views contained in this paper are solely the responsibility of the authors and do not represent the official views, either expressed or implied, of the NIH or the U.S. Government.

REFERENCES

[1] R. Lozano, M. Naghavi, and K. Foreman, "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010," *Lancet*, vol. 380, pp. 2095–2128, 2012.

[2] H. Wang, L. Dwyer-Lindgren, K. Lofgren, J. Rajaratnam, J. Marcus, A. Levin-Rector, C. Levitz, A. Lopez, and C. Murray, "Age-specific and sex-specific mortality in 187 countries, 1970-2010: a systematic analysis for the Global Burden of Disease Study 2010," *Lancet*, vol. 380, pp. 2071–2094, 2012.

[3] N. Zhong, C. Wang, W. Yao, P. Chen, J. Kang, S. Huang, B. Chen, C. Wang, D. Ni, Y. Zhou, S. Liu, X. Wang, D. Wang, J. Lu, J. Zheng, and P. Ran, "Prevalence of chronic obstructive pulmonary disease in China - A large, population-based survey," *Am J Respir Crit Care Med*, vol. 176, pp. 753–760, 2007.

[4] J. She, P. Yang, Q. Hong, and C. Bai, "Lung Cancer in China Challenges and Interventions," *Chest*, vol. 143, pp. 1117–1126, 2013.

[5] American Cancer Society, "Cancer Facts and Figures 2012," *American Cancer Society*, 2012.

[6] R. Young, R. Hopkins, T. Christmas, P. Black, P. Metcalf, and G. Gamble, "COPD prevalence is increased in lung cancer, independent of age, sex and smoking history," *Eur Respir J*, vol. 34, pp. 380–386, 2009.

[7] D. Skillrud, K. Offord, and R. Miller, "Higher risk of lung cancer in chronic obstructive pulmonary disease. A prospective, matched, controlled study," *Annals of Internal Medicine*, vol. 105, 1986.

[8] R. Loganathan, D. Stover, W. Shi, and E. Venkatraman, "Prevalence of COPD in women compared to men around the time of diagnosis of primary lung cancer," *Chest*, vol. 129, pp. 1305–1312, 2006.

[9] A. Mohan, C. Mohan, A. Pathak, R. Pandey, and R. Guleria, "Impact of chronic obstructive pulmonary disease on respiratory status and quality of life in newly diagnosed patients with lung cancer," *Respirology*, vol. 12, pp. 240–247, 2007.

[10] D. Aberle, A. Adams, C. Berg, W. Black, J. Clapp, R. Fagerstrom, I. Gareen, C. Gatsonis, P. Marcus, and J. Sicks, "Reduced lung-cancer mortality with low-dose computed tomographic screening," *The New England journal of medicine*, vol. 365, no. 5, pp. 395–409, 2011.

[11] N. Muller and H. Coxson, "Chronic obstructive pulmonary disease. 4: Imaging the lungs in patients with chronic obstructive pulmonary disease," *Thorax*, vol. 57, no. 11, pp. 982–985, 2002.

[12] D. Wilson, J. Weissfeld, A. Balkan, J. Schragin, C. Fuhrman, S. Fisher, J. Wilson, J. Leader, J. Siegfried, and S. Shapiro, "Association of radiographic emphysema and airflow obstruction with lung cancer," *American journal of respiratory and critical care medicine*, vol. 178, no. 7, pp. 738–744, 2008.

[13] J. de Torres, G. Bastarrika, J. Wisnivesky, A. Alcaide, A. Campo, L. Seijo, J. Pueyo, A. Villanueva, M. L. MD, and U. Montes, "Assessing the relationship between lung cancer risk and emphysema detected on low-dose ct of the chest," *Chest*, vol. 132, no. 6, pp. 1932–1938, 2007.

[14] Y. Li, S. Swensen, L. Karabekmez, R. Marks, S. Stoddard, R. Jiang, J. Worra, F. Zhang, D. Midthun, and M. de Andrade, "Effect of emphysema on lung cancer risk in smokers: A computed tomography-based assessment," *Cancer prevention research*, vol. 4, no. 1, 2011.

[15] A. Houghton, "Mechanistic links between COPD and lung cancer," *Nat Rev Cancer*, vol. 13, no. 4, pp. 233–245, April 2013.

[16] P. Barnes, "The cytokine network in chronic obstructive pulmonary disease," *American journal of respiratory cell and molecular biology*, vol. 41, no. 6, pp. 631–638, 2009.

[17] L. Hohberger, D. Schroeder, B. Bartholmai, P. Yang, C. Wendt, P. Bitterman, O. Larsson, and A. Limper, "Correlation of regional emphysema and lung cancer: a lung tissue research consortium-based study," *J Thorac Oncol*, vol. 9, no. 5, pp. 639–645, May 2014.

[18] M. Ashburner, C. Ball, J. Blake, D. Botstein, H. Butler, J. Cherry, A. Davis, K. Dolinski, S. Dwight, J. Eppig, M. Harris, D. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. Matese, J. Richardson, M. Ringwald, G. Rubin, and G. Sherlock, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat Genet*, vol. 25, no. 1, pp. 25–29, May 2000.

[19] R. Hoehndorf, C. Batchelor, T. Bittner, M. Dumontier, K. Eilbeck, R. Knight, C. Mungall, J. Richardson, J. Stombaugh, E. Westhof, C. Zirbel, and N. Leontis, "The RNA Ontology (RNAO): An ontology for integrating RNA sequence and structure data," *Applied Ontology*, vol. 6, no. 1, pp. 53–89, January 2011.

[20] OBO Library. [Online]. Available: http://obofoundry.org

[21] J. Huang, K. Eilbeck, B. Smith, J. Blake, D. Dou, W. Huang, D. Natale, A. Ruttenberg, J. Huan, M. Zimmermann, G. Jiang, Y. Lin, B. Wu, H. Strachan, N. de Silva, M. Kasukurthi, V. Jha, Y. He, S. Zhang, X. Wang, Z. Liu, G. Borchert, , and M. Tan, "The Development of Non-Coding RNA Ontology," *Int. J. Data Mining and Bioinformatics*, vol. 15, no. 3, pp. 214–232, June 2016.

[22] J. Huang, K. Eilbeck, B. Smith, J. Blake, D. Dou, W. Huang, D. Natale, A. Ruttenberg, J. Huan, M. Zimmermann, G. Jiang, Y. Lin, B. Wu, H. Strachan, Y. He, S. Zhang, X. Wang, Z. Liu, G. Borchert, , and M. Tan, "The Non-Coding RNA Ontology (NCRO): A comprehensive resource for the unification of non-coding RNA biology," *J Biomed Semantics*, vol. 7, no. 24, May 2016.

[23] J. Huang, J. Dang, X. Lu, D. Dou, J. Blake, W. Gerthoffer, and M. Tan, "An Ontology-Based MicroRNA Knowledge Sharing and Acquisition Framework," in *Proc. BHI Workshop at 2012 IEEE International Conference on Bioinformatics and Biomedicine (BIBM-2012)*, October 2012, pp. 16–23.

[24] J. Huang, J. Dang, G. Borchert, K. Eilbeck, H. Zhang, M. Xiong, W. Jiang, H. Wu, J. Blake, D. Natale, and M. Tan, "OMIT: Dynamic, Semi-Automated Ontology Development for the microRNA Domain," *PLOS ONE*, vol. 9, no. 7, July 2014.

[25] J. Huang, F. Gutierrez, H. Strachan, D. Dou, W. Huang, B. Smith, J. Blake, K. Eilbeck, D. Natale, Y. Lin, B. Wu, N. de Silva, X. Wang, Z. Liu, G. Borchert, M. Tan, and A. Ruttenberg, "OmniSearch: A semantic search system based on the Ontology for MIcroRNA Target (OMIT) for microRNA-target gene interaction data," *J Biomed Semantics*, vol. 7, no. 25, May 2016.

[26] T. Tran, P. Cimiano, S. Rudolph, and R. Studer, "Ontology-based interpretation of keywords for semantic search," *The Semantic Web*, vol. 4825, pp. 523–536, 2007.

[27] D. Dou, H. Wang, and H. Liu, "Semantic Data Mining: A Survey of Ontology-based Approaches," in *Proc. IEEE International Conference on Semantic Computing*, 2015, pp. 244–251.

[28] W. Premerlani and M. Blaha, "An Approach for Reverse Engineering of Relational Databases," *Commun. ACM Journal*, vol. 37, no. 5, pp. 42–49, 1994.

[29] L. Stojanovic, N. Stojanovic, and R. Volz, "Migrating data-intensive web sites into the Semantic Web," in *Proc. ACM symposium on Applied computing*, 2002, pp. 1100–1107.

[30] P. Verheyden, J. Bo, and R. Meersman, "Semantically Unlocking Database Content Through Ontology-Based Mediation," in *Proc. SWDB 2004*, 2004, pp. 109–126.

[31] L. Lubyte and S. Tessaris, "Extracting Ontologies from Relational Databases," in *Proc. 2007 Description Logics*, 2007, pp. 122–126.

[32] R. Chauhan, R. Gouda, R. Sharma, and A. Chauhan, "Domain ontology based semantic search for efficient information retrieval through automatic query expansion," in *Proc. 2013 International Conference on Intelligent Systems and Signal Processing (ISSP)*, March 2013, pp. 397–402.

[33] J. Huang, F. Gutierrez, D. Dou, J. Blake, K. Eilbeck, D. Natale, B. Smith, Y. Lin, X. Wang, Z. Liu, M. Tan, and A. Ruttenberg, "A semantic approach for knowledge capture of microRNA-target gene interactions," in *Proc. BHI Workshop at 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM-2015)*, November 2015, pp. 975–982.

[34] OmniSearch. [Online]. Available: http://omnisearch.soc.southalabama.edu/ui/

[35] miRDB. [Online]. Available: http://mirdb.org/miRDB/

[36] TargetScan. [Online]. Available: http://www.targetscan.org

[37] miRanda. [Online]. Available: http://www.microrna.org

[38] miRTarBase. [Online]. Available: http://mirtarbase.mbc.nctu.edu.tw/

[39] S. Lee, J. She, B. Deng, J. Kim, M. de Andrade, J. Na, Z. Sun, J. Wampfler, J. Cunningham, Y. Wu, A. Limper, M. Aubry, C. Wendt, P. Biterman, P. Y. P, and Z. Lou, "Multiple-level validation identifies PARK2 in the development of lung cancer and chronic obstructive pulmonary disease," *Oncotarget*, January 2016.