Identification of Common Molecular Signatures Between Severe Asthma And Lung Cancer

Mirudhula Mukundan, Aditi Sarathy

1 Background and Purpose

Lung cancer is a growing concern worldwide as the incidence of cases is ever-increasing. Common perception indicates smoking to be the likely reason but that does not explain for the 25% of the cases attributed to non-smoking related lung cancer[8]. Some meta-studies have linked severe cases of asthma to lung cancer [6, 3, 4], but these studies are mostly associated with conducting risk analysis on asthmatic and cancer patients, being observed over the course of several years, indicating a dearth in research based on gene expression data.

Although there are studies which have identified the molecular signatures and associated pathways of asthma[1] and lung cancer[9] independently, very few [5] have reported common molecular signatures associated with both lung cancer and severe case of asthma. Hence the main purpose of our study will be to identify a panel of common gene signatures of deferentially expressed genes in non small-cell lung cancer (NSCLC) patients and severe asthmatic (SA) patients which act as bio-markers for lung cancer detection. These may act as a fundamental set of genes that may be responsibe for affecting molecular pathways implicated by lung cancer.

2 Data

For our study we will use two sets of gene expression data-sets for lung cancer and asthma, one set on which we will perform our analysis and one set that will be used for validation

1. For Analysis :

- NSCLC : GSE68793
- Asthma (Bronchial Epithelial Cells) : GSE63142

2. For Validation :

- NSCLC : GSE29013
- Asthma (Epithelial brushings of central and peripheral airways) : GSE64913

Our main purpose for choosing the asthma data sets is to be able to filter on the smokers, number of years smoked and non-smokers along with categorizing the samples based on asthma severity. The datasets we chose for lung cancer will help us categorize the samples based on the different stages of cancer.

3 Methods and Analysis

Following data pre-processing and filtering of our analysis dataset, our main goal will be to identify the differentially expressed genes (either using Gene Set Enrichment Analysis (GSEA) [7] or Bioconductor package [2] from R) for lung cancer and asthma samples individually with respect to healthy individuals. The common probes will be selected as our final gene signature panel, after conducting Fisher's exact test in order to validate their significance, and this will be validated for their classification accuracy against our second pair of dataset by conducting classification algorithms like Support Vector Machine

(SVM) or Logistic Regression (LR). A low classification accuracy would indicate that our gene set to indeed be common signatures for both the diseases as it would mean that the features are indeed shared between them.
An additional analysis using Gene Ontology may be conducted for pathway enrichment analysis that could help us examine the common pathways being affected. The main programming languages we will be using are Python and/or R.

References

[1] Bigler, J., Boedigheimer, M., Schofield, J. P. R., Skipp, P. J., Corfield, J., Rowe, A., Sousa, A. R., Timour, M., Twehues, L., Hu, X., Roberts, G., Welcher, A. A., Yu, W., Lefaudeux, D., Meulder, B. D., Auffray, C., Chung, K. F., Adcock, I. M., Sterk, P. J., and Djukanović, R. A severe asthma disease signature from gene expression profiling of peripheral blood from u-biopred cohorts. *American Journal of Respiratory and Critical Care Medicine 195*, 10 (2017), 1311–1320. PMID: 27925796.

[2] Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology 5*, 10 (2004), 1–16.

[3] Qu, Y.-L., Liu, J., Zhang, L.-X., Wu, C.-M., Chu, A.-J., Wen, B.-L., Ma, C., Yan, X.-y., Zhang, X., Wang, D.-M., et al. Asthma and the risk of lung cancer: a meta-analysis. *Oncotarget 8*, 7 (2017), 11614.

[4] Rosenberger, A., Bickeboeller, H., McCormack, V., Brenner, D. R., Duell, E. J., Tjønneland, A., Friis, S., Muscat, J. E., Yang, P., Wichmann, H.-E., et al. Asthma and lung cancer risk: a systematic investigation by the international lung cancer consortium. *Carcinogenesis 33*, 3 (2012), 587–597.

[5] Salameh, L., Bhamidimarri, P. M., Saheb Sharif-Askari, N., Dairi, Y., Hammoudeh, S. M., Mahdami, A., Alsharhan, M., Tirmazy, S. H., Rawat, S. S., Busch, H., Hamid, Q., Al Heialy, S., Hamoudi, R., and Mahboub, B. In silico bioinformatics followed by molecular validation using archival ffpe tissue biopsies identifies a panel of transcripts associated with severe asthma and lung cancer. *Cancers 14*, 7 (2022).

[6] Salameh, L., Mahboub, B., Khamis, A., Alsharhan, M., Tirmazy, S. H., Dairi, Y., Hamid, Q., Hamoudi, R., and Al Heialy, S. Asthma severity as a contributing factor to cancer incidence: A cohort study. *PloS one 16*, 5 (2021), e0250430.

[7] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences 102*, 43 (2005), 15545–15550.

[8] Sun, S., Schiller, J. H., and Gazdar, A. F. Lung cancer in never smokers—a different disease. *Nature Reviews Cancer 7*, 10 (2007), 778–790.

[9] Taniwaki, M., Daigo, Y., Ishikawa, N., Takano, A., Tsunoda, T., Yasui, W., Inai, K., Kohno, N., and Nakamura, Y. Gene expression profiles of small-cell lung cancers: molecular signatures of lung cancer. *International journal of oncology 29*, 3 (2006), 567–575.