**Title:**

Health Bioinformatics research performing bioinformatics data analysis for biomarkers discovery in lung cancer to find out potential genes

**Abstract:**  
  
**1. Introduction:**

First paragraph = Lung Cancer

Second = Bioinformatics techniques why biomarkers are important

Third = what we are going to do in our research

**2. Methodology:**

**2.1 Microarray Data source of Lung Cancer**

The research involved the acquisition of gene expression data from the Gene Expression Omnibus (GEO), a repository managed by the National Center for Biotechnology Information (NCBI). The dataset selected for analysis in this study was identified by its accession number, GSE229260. The GSE229260 dataset encompasses samples, comprising 3 samples from patients with small cell lung cancer (SCLC), and 3 samples from normal lung tissue. For the purpose of this study, a distinction was made between control and treatment parameters. The treatment group included 3 samples of tumor tissue, including both NSCLC and SCLC cases. The control group consisted of 3 samples of adjacent normal lung tissue. These samples were selected for their relevance to the research objectives.

**2.2 Data Processing and Differential Gene Expression Analysis**

Data processing involved the development of a custom R pipeline using the 'affy' package (Version 1.74.0) from Bioconductor (<http://bioconductor.org/packages/release/bioc/html/affy.html>). The pipeline included functions for exploratory oligonucleotide array analysis and background adjustment to ensure data comparability and integrity. Normalization and expression measures were calculated using the Robust Multi-array Average (RMA) method. Differential gene expression analysis between the control and treatment groups was conducted using the 'limma' package (Version 3.26.8) from Bioconductor (<http://bioconductor.org/packages/release/bioc/html/limma.html>) Differentially expressed genes (DEGs) were selected based on a log2 (fold change) absolute value greater than P.05. Probes lacking corresponding platform annotation files were excluded from the analysis if they did not match a gene symbol.

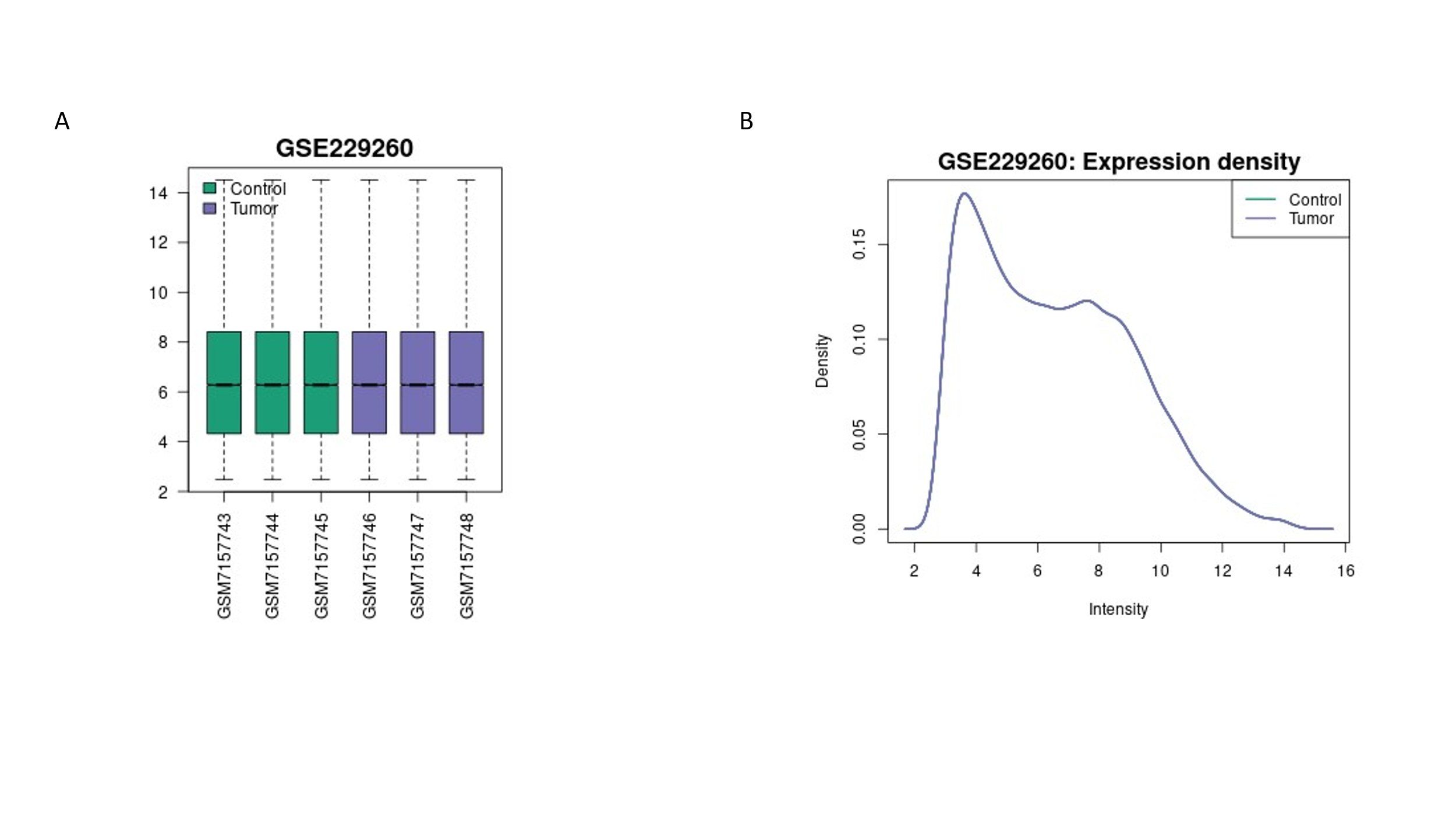
**2.3**

**2.4**

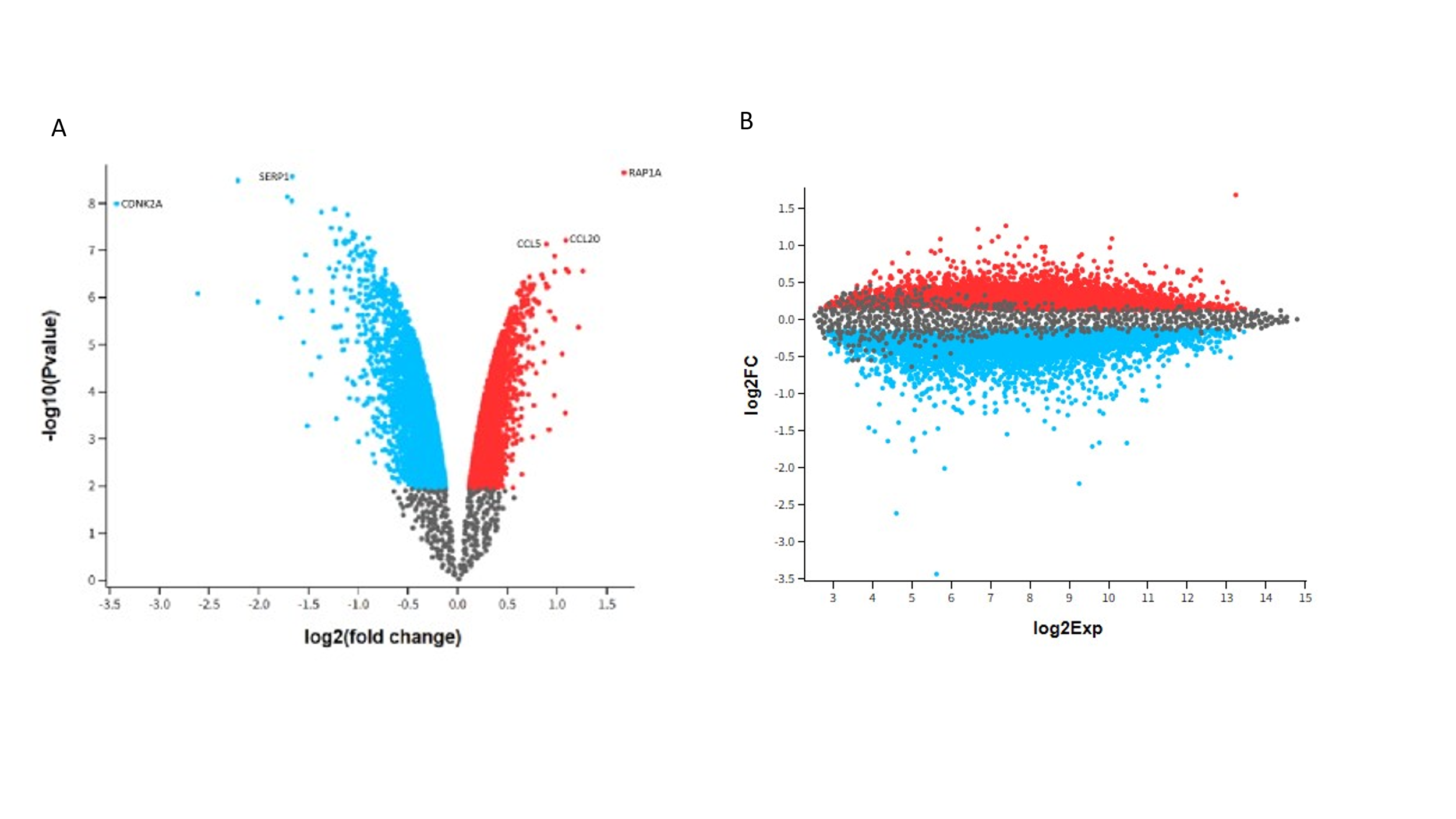
**2.5**

**Results:**

**2.1 Data Source**



**2.2 DEG’s Identification**



**Table 1:**  
DEG’s Table of biomarkers in lung cancer

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ID | adj.P.Val | P.Value | t | B | logFC | GB\_ACC | Gene.Symbol |
| 11762533\_a\_at | 5.42E-05 | 2.27E-09 | 4.34E+01 | 11.08532 | 1.67 | AB051846.1 | RAP1A |
| 11720306\_at | 5.42E-05 | 2.72E-09 | -4.22E+01 | 10.99434 | -1.66 | BX350905 | SERP1 |
| 11718293\_a\_at | 8.49E-05 | 1.03E-08 | -3.45E+01 | 10.22734 | -3.43 | U38945.1 | CDKN2A |
| 11724828\_at | 0.000147 | 6.09E-08 | 2.63E+01 | 8.971929 | 1.09 | NM\_004591.2 | CCL20 |
| 11753810\_a\_at | 0.000147 | 7.35E-08 | 2.56E+01 | 8.824288 | 8.91E-01 | AF043341.1 | CCL5 |

**2.3**

**2.4**

**2.5**

**Discussion and Conclusion:**

First Paragraph = Lung cancer and biomarker

2nd paragraph = Identified biomarkers

3rd = biomarkers identified previously

Unique identifiers:

ENSG00000116473

ENSG00000271503

ENSG00000115009

ENSG00000120742

ENSG00000147889