



Bioinformatics-Guided Profiling of ABCG2 and GJA1 to Inform Biomedical Engineering of Nanoparticle Treatment for Advanced Non-Small Cell Lung Cancer

ABSTRACT

Introduction

Non-small cell lung cancer (NSCLC) is the most common Lung Cancer type, accounting for 85 percent lung cancer patients. The disease is more frequently diagnosed at advanced stages, for which treatments have poor survival rates. Addressing these treatment limitations can be done using bioinformatics to guide biomedically-engineered treatments. Depletion of the LDOC1 protein is linked to heightened NSCLC tumor growth. Therefore, this research aimed to find biomarkers through bioinformatics analysis of LDOC1-depleted samples, and identify biomedical engineering strategies to target those biomarkers for advanced-stage NSCLC treatment.

Methods

Transcriptomic dataset ID number GSE235829 was obtained from NCBI GEO Data Sets, including data set included samples of human A549 cell lines, obtained through expression profiling by array. The samples were grouped into “Less LDOC1 with mutation”, “Less LDOC1”, and the control group “Vector control of Less LDOC1”. DAVID was used for functional and enrichment analysis of Gene Ontology and Reactome Pathways. STRING was then used to reveal gene-protein interactions, and DGIdb was used to discover gene-drug interactions.

Results

21004 genes were identified (fig. 2d) and narrowed down to 50 genes using p-values. The most crucial categories and genes in DAVID analysis were transmembrane transport, plasma membrane, and the GJA1 gene. STRING analysis revealed GJA1-ABCG2 associations. In DGIdb, ABCG2 was shown to have the most drug interactions, one drug being paclitaxel, which can be nanoparticle-bound.

Conclusion

Together, these results identified ABCG2 as a nanoparticle-based treatment candidate. GJA1 could also be studied as a marker of tumor communication and treatment response.

INTRODUCTION

Non-Small Cell Lung Cancer (NSCLC) is the most common and frequent sub-type of Lung Cancer [1]. Lung Cancer is the leading cause of all cancer or cancer-related deaths worldwide [1]. This is mainly due to a global increase in tobacco use [1]. Smoking causes 85 to 90 percent of all Lung Cancer cases [1]. Lung cancer involves 2 main subtypes: NSCLC and SCLC (Small Cell Lung Cancer) [1]. The main difference is the frequency of the disease, with NSCLC making up 85% of Lung Cancer patients, and SCLC making up 15% [1].

The main challenge of NSCLC is that diagnosis often only occurs when the disease is in its later, more advanced stages [1]. This makes treatment more challenging most of the time, as it becomes harder to help the patient survive [1]. This is shown in the platinum-based chemotherapy that is traditionally used for treating advanced lung cancer, for which the survival rates are low (33% of patients survive after 1 year and 11% after 2 years) [1,2]. Due to this, newer research has gone into molecularly targeted therapy and treatment, one example being the study of treatment pathways regarding EGFR mutations [1,2]. This is a crucial field because it allows the treatment to target highly specific genes that cause the cancer, and make the treatment specific to the patient [1, 2, 3]. Molecular research of genes allows for significant biomarkers and genes to be pinpointed, which opens up lines for drug development and biomedical engineering pathways that will revolutionize treatment for the disease [1, 2, 3]. The molecularly targeted research done in this study therefore adds significant knowledge to progress the development of specific treatments [4].

The samples used in this research involve depletion of a protein called LDOC1 [4]. LDOC1 is a protein that controls the H2Bub1's chemical tag's ability to bind to the H2B histone, as the chemical tag regulates the histone's ability to control transcription [4]. When activated by H2Bub1, H2B permits transcription of many cancer-suppressing genes by unraveling DNA [4]. H2Bub1 cannot bind to H2B and allow it to regulate transcription without LDOC1, so when LDOC1 is depleted, it causes abnormally constricted DNA compaction, preventing gene expression and thus leading to cancer [4]. There is also a lab-created mutation called H2BK120R which prevents the H2B histone from ever binding to DNA, causing the same effect as LDOC1-depletion [4].

NSCLC is quite common and difficult to survive [1]. It is often driven by environmental factors such as tobacco smoking, second-hand smoking, carcinogenic (cancer-causing) chemicals, HIV infection, alcohol, and a family history [1]. It is classified into 3 subtypes: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma [1]. For early stages of NSCLC, surgical excision/resection is the standard recommended treatment [1]. For later stages of NSCLC, platinum-based chemotherapy has always been the standard treatment despite poor survival rates and many concerns [1]. However, the methods of immunotherapy, molecularly targeted therapy, and radiotherapy are all being used and researched more and more as potentially better treatment types for NSCLC [1].

Bioinformatics being introduced to cancer overall has been a major breakthrough for researching the disease [5]. The development of high-throughput technologies has allowed the investigation of genetic variation, including gene expression and protein levels [5]. This has caused the development of precision oncology, a field which needs biomarkers to be identified and validated

[5]. An example is the discovery of biomarkers to help inhibit the EGFR gene, which helps prevent NSCLC in patients with an EGFR mutation [5]. Some important functions of high-throughput machinery and bioinformatics tools include DNA sequencing, RNA sequencing, large data repositories, data manipulation, and data analysis [5].

The biggest challenge about cancer as a whole is how vast, diverse, and varied it is [6]. Cancer can come in all different parts of the body, with different tumours having different characteristics, and the cells within tumours varying widely [6]. Tumor Heterogeneity—the fact that there are differences between cancer cells within a tumor, between tumors in the same person, and between people—is a serious problem that makes cancer incredibly difficult to treat [6]. Overall, it is very hard to develop centralized treatments for cancer, and the treatments cannot always fully remove the cancer cells [6]. In NSCLC, the biggest challenge comes in the fact that diagnosis often only happens in the later, more complicated stages of it, and the treatments that are given for it once diagnosed are poorly effective [1, 2].

The overall goal of this study is to identify genes that can be used as biomarkers for treating Non-Small Cell Lung Cancer, to fuel biomedical engineering research for improvement of treatment of advanced NSCLC. The study utilized samples of cells depleted of the LDOC1 protein, with and without the H2BK120R mutation, and samples of normal cells, to determine if there any specific genes that show significant differential expression between these groups and can be used as biomarkers for treatment for this disease. The study also used functional analysis and Drug-Gene analysis to develop an understanding of possible biomedical engineering strategies for the disease.

This study hypothesizes that when comparing samples of cells depleted of the LDOC1 protein, samples of cells that are LDOC1-depleted with a H2BK120R mutation, and a control group of normal cells, there are specific genes that show significant differential expression and can be considered as biomarkers to be used for treatment for this disease.

This research is important because research of the LDOC1 protein, the H2B histone, and H2Bub1 chemical tag and their links to cancer by comparing different samples, reveals important genetic markers that can be used as a stepping stone for biomedical engineering advancements, as it provides crucial knowledge that will allow for novel treatments to be developed [1, 2, 4].

METHODS

Data Collection and Analysis of GEO2R Data

In this study, the dataset titled “Transcriptomic effect of LDOC1 and H2Bub1 in NSCLC” (GEO accession: [GSE235829](#) on Non-small cell lung cancer (NSCLC) was collected from NCBI GEO2R by using the keywords “NLSCLC” and “LDOC1” in the query in the search bar [7].

Then the datasets was defined or categorized into groups “Less LDOC1 with mutation” (experimental group: samples with the H2BK120R mutated packaging protein, that also have less of LDOC1, and force a dramatic reduction in overall cell-wide H2Bub1 levels due to the mutation), “Less LDOC1” (experimental group: cells with less LDOC1, causing more cell-wide H2Bub1 but less H2Bub1 bound to chromatin) And "Vector control of Less LDOC1" (control group: carrying a vector with a non-targeting RNA sequence instead of a sequence targeting

LDOC1, meaning it has no effect on the level of LDOC1 in the sample), And analyzed using the no-code GEO2R bioinformatics tool that uses R programming language [7].

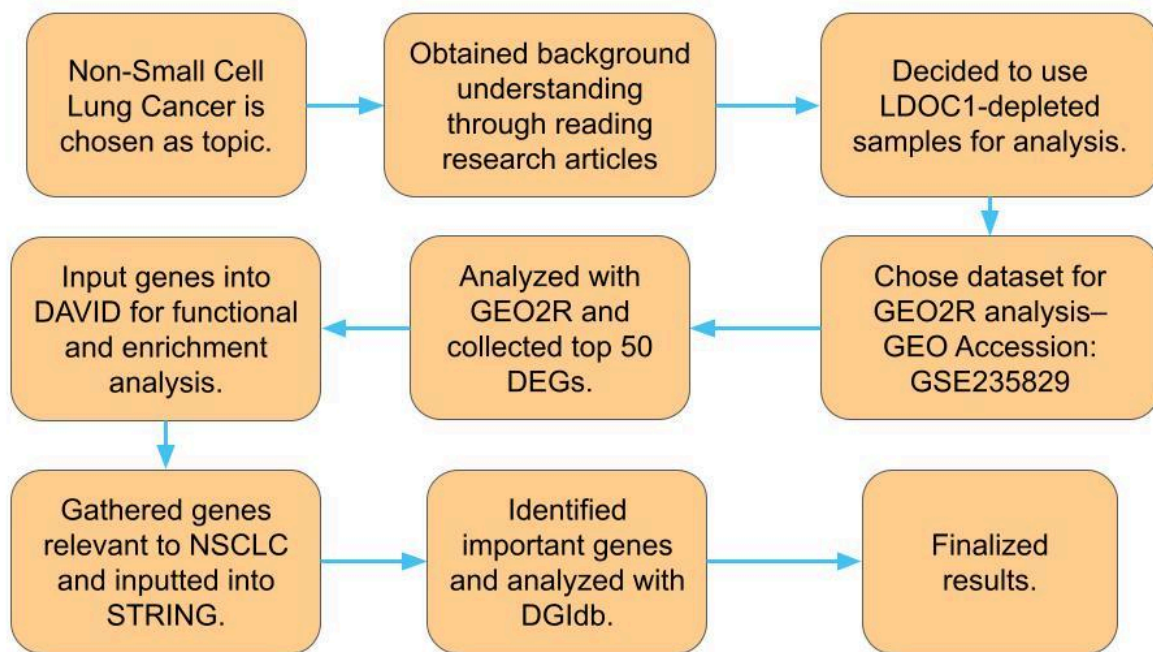


Figure 1: Research Methodology: The steps and bioinformatics strategies used in this study.

Identification of the Top Differentially Expressed Genes

To identify the most significant differentially expressed genes (the top 30 or 40), statistical analysis was applied. This process used the p-value only to prioritize the most important genes based on their differential expression across samples. The process involved taking the full gene table from GEO2R [7], inputting it into Google Sheets, and finding only the 50 genes with the lowest P-Values from that list, which were then copied to another spreadsheet in that Google Sheet [7].

Functional and Enrichment Analysis using DAVID, KEGG, and GO Bioinformatics Tools

The Database for Annotation, Visualization and Integrated Discovery (DAVID), KEGG, and GO bioinformatics tools and databases were utilized to analyze the functions of these top genes [8, 9, 10]. These tools helped uncover the potential roles of the genes in NSCLC tumor development and expansion, and how significant the genes were in contributing to that. The tools also helped reveal the amount of genes (enrichment of genes) involved in these different functions.

Then from there, the genes from the important functions were input into the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), where gene-protein interactions were discovered [11]. Finally, the genes with the most interactions there were analyzed using the Drug-Gene Interaction Database (DGldb), to find recorded drug interactions. This also revealed

which drugs have regulatory approval and which are not approved, and the type of interaction [12].

RESULTS

NCBI's GEO2R Bioinformatics Tool was used first to identify Differentially Expressed Genes (DEGs) [7]. DEGs are genes that vary in their expression among different groups of samples. Identifying them is key in bioinformatics research for biomedical engineering of treatments for diseases, as the most significant DEGs can be used as biomarkers for the disease being researched.

Many genes were expressed differently between the three groups. When comparing the experimental groups (the Less LDOC1 group and the Less LDOC1 with Mutation group) to the control group (the Vector Control Less LDOC1 group), there were many genes that had differential expression. However, when comparing the two experimental groups with each other, there was no differential expression for any of the genes.

When comparing two groups in GEO2R, the blue dots in the volcano plot shown represent genes that are down-regulated, meaning the gene is expressed less in the control group compared to the experimental group, and the red dots in the volcano plot represent genes that are up-regulated, meaning the gene is expressed more in the control group compared to the experimental group. The black dots represent genes that are expressed the same amount in each group, meaning that the expression of those genes are not affected by the experiment.

Given by the results of the analysis done in GEO2R, the study produced 21004 genes total, as demonstrated in the Venn Diagram (Fig. 2D). There were also 248 DEGs that overlapped in both experimental-control comparisons in the Venn Diagram (Fig. 2D). Due to the analysis revealing 0 DEGs when comparing the two experimental groups, no DEGs overlapped in all three comparisons.

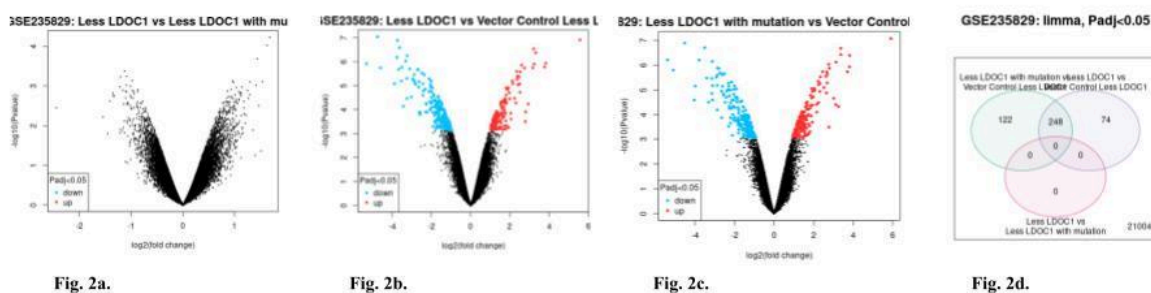


Figure 2: Differentially Expressed Genes: Volcano plots show the up/down-regulation of each DEG when comparing two groups, while the venn diagram shows the DEGs for each of the three comparisons made, and any DEGs that are discovered in multiple comparisons (overlapping DEGs). **2a.** Volcano plot comparing Less LDOC1 vs Less LDOC1 with mutation. **2b.** Volcano plot comparing Less LDOC1 vs Vector Control Less LDOC1. **2c.** Volcano plot comparing Less LDOC1 with mutation vs Vector Control Less LDOC1. **2d.** Venn Diagram.

Identification of (30 or 40 or 50) Statistically Significant Differentially Expressed Genes (DEGs)

The P-Value was used to narrow down the genes from 21004 to 50 top DEGs. These top 50 were the DEGs with the smallest P-Value. Out of the 50 DEGs, 29 were Down-regulated (higher expression in control group than experimental groups), and 21 were Up-regulated (lower expression in control group than experimental groups). Top 50 DEGs

Potential Functions and Enrichment of the Identified Genes and/or pathways

The Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics tool was used to determine potential functions and biological involvements of the genes, as well as enrichment of the DEGs [8].

The KEGG Pathways did not present any data in DAVID, prompting investigation into Reactome Pathways. The 3 highest enriched Reactome Pathways were Developmental Biology (10 genes, p-value = $3.13\text{E-}02$), Axon guidance (5 genes, p-value = $7.33\text{E-}02$), and Nervous system development (5 genes, p-value = $8.34\text{E-}02$). All 3 pathways were quite connected to NSCLC and cancer, as problems during fetal development can lead to NSCLC tumor development, and because axon guidance molecules can be hijacked to fuel NSCLC tumor growth. However, the axon-guidance result has $p = 0.0733$, and nervous-system development has $p = 0.0834$. These are not statistically significant at the conventional $p < 0.05$ threshold. Developmental biology is the only reactome pathway that is nominally significant, at $p = 0.0313$.

Developmental Biology was therefore the only reactome pathway which was significant and included in further analysis. The genes involved in this pathway were GREM1, KRT19, ANKRD30A, ITGB4, DPYSL3, SCN9A, NRCAM, GFRA1, UNC5D, ABCG2. Within these, DPYSL3, SCN9A, NRCAM, GFRA1, UNC5D, stood out as being common in all 3 reactome pathways mentioned.

DAVID analysis also highlighted the most enriched genes in each Gene Ontology (GO), and focused on the 3 most enriched categories in each Ontology category [10]. It revealed the most enriched BPs to be Signal Transduction, Transmembrane transport, and Cell Communication. All of these processes demonstrated a statistically significant P-Value. Signal Transduction had a P-Value of $2.99\text{E-}02$, Transmembrane transport had a P-Value of $1.59\text{E-}03$, and Cell Communication showed the most significant P-Value of $7.80\text{E-}05$. Cell communication stood out as a highly significant process to NSCLC on pure statistical merit. Analysis of these biological processes regarding their importance to NSCLC showed all three as significant to NSCLC, with Transmembrane Transport pointed out as being especially key. In Cell Components, DAVID analysis identified plasma membrane (p-value = $1.18\text{E-}05$), membrane (p-value = $1.87\text{E-}04$), and extracellular region (p-value = $2.46\text{E-}02$) as being the most enriched. All three of these components proved to be statistically significant (p-value > 0.05), prompting further analysis to eliminate redundancies or insignificant genes. The membrane component was pinpointed as including too many genes also present in the plasma membrane. Membrane was less specific and contained less genes than plasma membrane, and so was finally eliminated from further consideration. The extracellular region was analyzed and also removed from the research due to not having significant connections to NSCLC tumor formation or continuation.

Finally, GO analysis in DAVID highlighted the three most enriched Molecular Functions, which were Metal ion binding (p value = $1.79\text{E-}02$), Transmembrane transporter activity (p value =

1.50E-05), and Signaling receptor binding (6.64E-02). Metal ion binding had a significant p value, with several connections to cancer in general. In-depth analysis of it showed involvement in immune cell activation and multiplication for both the innate and adaptive immune system, especially cells that can identify tumours and eliminate them. This indicated potential roles in improving general immunotherapy. Metal ion binding also is able to induce tumor cell death by affecting cell metabolism, in ways such as ferroptosis. Metal ions are used in nanomedicines that can prevent tumor growth or kill tumors. Overall, metal ion binding shows potential in improving cancer immunotherapy [13]. Transmembrane transporter activity was found to be an unnecessary category due to many similarities with the Transmembrane Transport biological process, and was eliminated from further study, for similar reasons as the removal of the membrane cellular component. Signaling receptor binding also possessed a p-value under the 0.05 threshold, making it not significant and eliminating it from further study. Totally, 5 Gene Ontologies were selected for further research, after eliminating all irrelevant/insignificant ontologies. This included 3 BPs: Signal Transduction (8 genes, $p = 2.99\text{E-}02$), Transmembrane transport (7 genes, $p = 1.59\text{E-}03$), and Cell Communication (5 genes, $p = 7.80\text{E-}05$). The BPs included the genes: LGALS3BP, GREM1, NR4A1, GJA1, PDE1C, OLFML3, TNFSF15, UNC5D, SLC45A4, SLC2A14, SCN9A, SLC2A3, KCNQ5, ABCG2, NPY4R, ITGB4, and NTRK3, with GJA1 being present in all 3 BPs. There was also the CC of Plasma Membrane (27 genes, $p = 1.18\text{E-}05$), which included the genes SLC45A4, SNAP25, NPY4R, ITGB4, SLC2A3, GLIPR1, GJA1, VSIG1, SCN9A, NRCAM, GLUL, SLC2A14, TNFSF15, SUSD2, NTRK3, SLC6A14, FZD8, GFRA1, UNC5D, MYO1D, KRT19, ADORA2B, CEACAM6, ADAM12, KCNQ5, SLC29A3, ABCG2. Finally, there was the MF of Metal ion binding (10 genes, $p = 1.79\text{E-}02$), which included the genes PDE1C, ITGB4, GDA, NTRK3, ADAM12, SLC6A14, NEK10, NPTX1, GLUL, LHX8. The genes NTRK3 and ITGB4 were present in all the 3 types of GOs at least once.

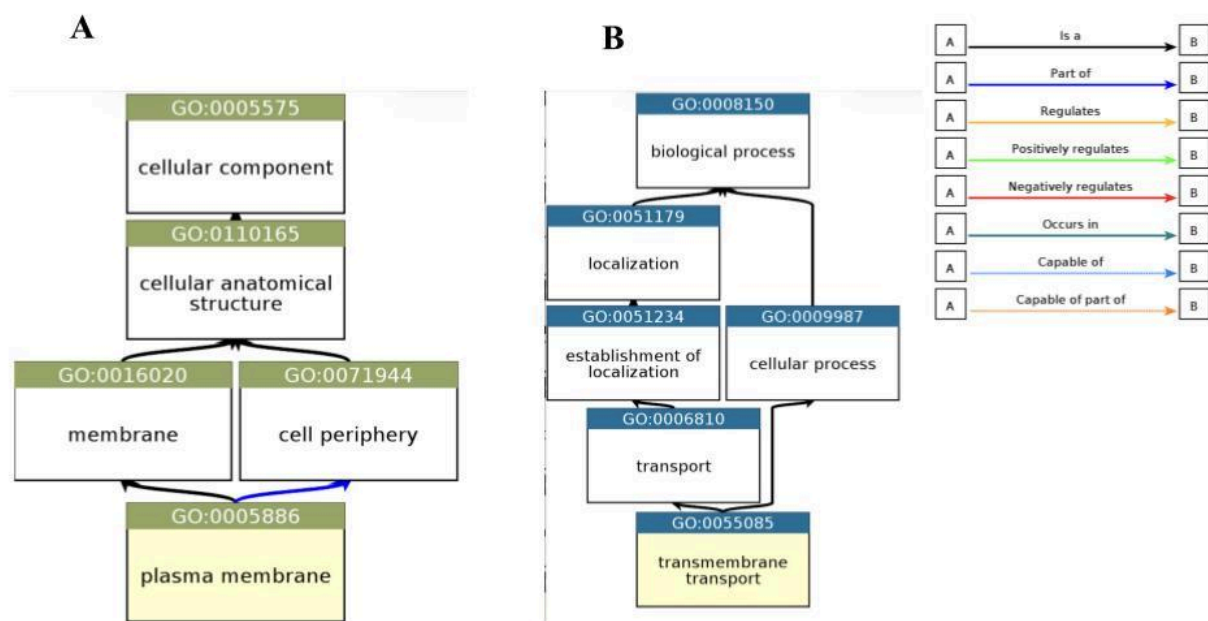


Figure 3: Ancestor charts of 2 significant GO categories. A. Plasma membrane cellular component, which included ABCG2, GJA1, NTRK3, and KRT19. B. Transmembrane Transport biological process, which included ABCG2 and GJA1.

Table 1: DAVID Categories and genes for each one. All of these 5 categories are statistically significant and are enriched with many genes.

Type of GO/Pathway	Name of GO/Pathway	P-Value	Genes
BP	Signal Transduction	2.99E-02	LGALS3BP, GREM1, NR4A1, GJA1 , PDE1C, OLFML3, TNFSF15, UNC5D.
BP	Transmembrane transport	1.59E-03	SLC45A4, SLC2A14, GJA1 , SCN9A, SLC2A3, KCNQ5, ABCG2 .
BP	Cell Communication	7.80E-05	GJA1 , NPY4R, ITGB4, TNFSF15, NTRK3.
CC	Plasma Membrane	1.18E-05	SLC45A4, SNAP25, NPY4R, ITGB4, SLC2A3, GLIPR1, GJA1 , VSIG1, SCN9A, NRCAM, GLUL, SLC2A14, TNFSF15, SUSD2, NTRK3, SLC6A14, FZD8, GFRA1, UNC5D, MYO1D, KRT19, ADORA2B, CEACAM6, ADAM12, KCNQ5, SLC29A3, ABCG2 .
MF	Metal ion binding	1.79E-02	PDE1C, ITGB4, GDA, NTRK3, ADAM12, SLC6A14, NEK10, NPTX1, GLUL, LHX8.
Reactome	Developmental Biology	3.13E-02	GREM1, KRT19, ANKRD30A, ITGB4, DPYSL3, SCN9A, NRCAM, GFRA1, UNC5D, ABCG2 .

All 38 genes totally involved in the 5 categories were inputted into STRING [11]. This showed the network of gene-protein interactions formed with these specific genes. NRCAM was revealed to have the most protein interactions (3), including a connection to NTRK3, which had 2 interactions. ABCG2, GLUL, GJA1, and KRT19 each had 2 interactions, which also highlighted an association between ABCG2 and GJA1, as well as ABCG2 and KRT19. SLC2A14 and SLC2A3 also demonstrated a connection. There were 9 gene-protein interactions total. Despite showing some significance in the DAVID functional and enrichment analysis, ITGB4 showed no interactions.

From here, the genes NTRK3, GLUL, ABCG2, GJA1, CEACAM6, KRT19, and NRCAM were inputted into DGIdb, based on their significance in DAVID functional and enrichment analysis, and STRING gene-protein interaction analysis [12]. The ABCG2 gene had the most recorded drug interactions (74), which were all inhibitors. Out of these 74 interactions, 58 drugs were approved and only 16 were not approved. One of the approved drugs was paclitaxel, a drug that can be nanoparticle albumin-bound (known as nab-paclitaxel) [14]. NTRK3 had the second most drug interactions (58) and GJA1 had the third most (15). KRT19 had only 2 interactions. Despite having the highest number of protein interactions, NRCAM showed 0 drug interactions.

Summary Table of Key Selected Genes

Key Genes	Gene Function	Most Enriched Biological Pathway	Connection to Biomedical-Engineering Strategy for NSCLC
ABCG2	ATP-binding cassette transporter ABCG2, extrudes a wide variety of materials from cells.	Transmembrane transport, Plasma Membrane	An albumin-based nanoparticle delivering paclitaxel to regulate ABCG2, and may also increase paclitaxel retention inside NSCLC cells.
GJA1	Gap junction alpha-1 protein; Gap junction protein that acts as a regulator of bladder capacity.	Transmembrane transport, Plasma Membrane	To be studied as a marker of tumor communication and treatment response.

DISCUSSION

Summary of Findings

In summary, GEO2R produced many DEGs when comparing LDOC1-depleted cells and normal cells, providing insight into the impact that LDOC1-depletion has on cells. GEO2R also showed many DEGs when comparing normal cells to LDOC1-depleted cells that also carry the H2BK120R mutation. However, when comparing the two experimental groups (LDOC1-depleted samples to LDOC1-depleted cells that include the H2BK120R mutation), there were no DEGs highlighted. The 21004 DEGs given by GEO2R analysis were further analyzed and narrowed down to 50 DEGs, based on the genes with the lowest P-Value. Next, DAVID functional and enrichment analysis was used to identify the amount of genes involved in various functions and pathways. Here, KEGG analysis highlighted Developmental Biology as an important reactome pathway. GO analysis in DAVID revealed Signal Transduction, Transmembrane transport, Cell Communication, Plasma Membrane, and Metal Ion Binding to all be significant Gene Ontologies, with Transmembrane Transport and Plasma Membrane being especially key. DAVID analysis spotlighted 38 DEGs total as worthwhile candidates for further evaluation. Using STRING to identify gene-protein interactions produced a network of the genes. This included ABCG2-GJA1 and ABCG2-KRT19 connections being shown. NRCAM was also shown to have the most interactions (3). Key genes that were pinpointed in STRING and in DAVID functional and enrichment analysis, were inputted into DGIdb for Drug-Gene interaction analysis. Here, the genes NTRK3, GLUL, ABCG2, GJA1, CEACAM6, KRT19, and NRCAM were analyzed, and ABCG2 was shown to have the most drug interactions (74). One of the approved drugs for ABCG2 was paclitaxel, which can be nanoparticle albumin-bound (nab-paclitaxel) [14]. NTRK3 had the second most drug interactions (58) and GJA1 had the third most (15).

Interpretation of Results

The GEO2R results show that LDOC1 depletion certainly has an impact on gene expression in lung cancer cells. It also demonstrates that adding the H2BK120R mutation to cells already LDOC1-depleted does not affect gene expression, since comparison of LDOC1-depleted samples with and without the mutation showed no differential expression in the results. The reason reactome pathways were analyzed instead of KEGG pathways themselves, is because there was found to be no significant KEGG pathway data. Therefore, the developmental biology reactome pathway was selected, which showed some links to NSCLC tumorigenesis [15]. On the topic of GO analysis results, Signal Transduction and transmembrane transport have many connections as biological processes. The dysregulation of signaling pathways is a way in which NSCLC cells can proliferate and make use of signal transduction. SLC, ABC, and MUC transporters in transmembrane transport are linked to NSCLC tumor growth, malignancy, and drug resistance [16, 17]. Cell communication is also quite important in NSCLC, as it is the basic process that cancer uses to spread. GJA1 was involved in all three of these biological processes. Regarding cellular components, the plasma membrane is highly significant in NSCLC, as NSCLC tumors often have abnormally quick membrane resealing and repairing times. The speed of membrane resealing and repairing can be used as an indicator of the danger a tumor poses [16]. There is also evidence of Epithelial membrane proteins (such as EMP2) suppressing NSCLC cell-growth by inhibiting specific pathways [17]. Metal ion binding has a variety of important factors in NSCLC, including inducing cell death, and activating immune cells against tumor cells [13].

The ABCG2-GJA1 and ABCG2-KRT19 associations identified in STRING analysis were logical, as all 3 genes were involved in the plasma membrane. GJA1 and ABCG2 were also both recognized in the Transmembrane Transport Biological Process. GJA1 and ABCG2 had 2 interactions each. NRCAM having the most interactions was an indicator of the gene's importance, but did not have a larger importance to the study as the gene was not very significant in any of the other analyses. In DGIdb, ABCG2 was shown to have the most drug interactions, including one with paclitaxel. Paclitaxel has been used in treatments as a nanoparticle albumin-bound drug. This shows potential avenues of drug development that are nanoparticle-based, especially considering that ABCG2 is a plasma membrane protein and a transmembrane transporter, and nanoparticles can allow drugs to reach and/or pass through the plasma membrane more easily [14, 20]. ABCG2 is also one of the ABC transmembrane transporters, which have many links to NSCLC. GJA1 had 15 drug interactions. Looking at the functions of ABCG2 and GJA1, ABCG2 produces a membrane protein that can remove certain anticancer drugs from cells. GJA1 produces connexin 43, which allows neighboring cells to communicate through gap junctions [21].

These results prioritized ABCG2 as a druggable nanoparticle-treatment target and GJA1 as a potential biomarker of tumor-cell communication. Overall, ABCG2 and GJA1 are shown as strong potential biomarkers and therapeutic targets for NSCLC tumor growth.

Comparison with Previous Studies

The findings revolve around two crucial genes which are involved in the plasma membrane and transmembrane transport. One of these genes is an ABCG2 transporter

(ABCG2), and its recognized significance in the results of this study is similar to previous studies on ABC and SLC transporters in Lung Cancer [17]. This study's findings concerning the ABCG2 gene also aligned with previous wet research on ABCG2 gene expression in NSCLC, especially with regards to how both this study and other research identified it as having lower transcriptional levels in affected cells [22]. As for GJA1, previous studies have identified it as being a therapeutic target for NSCLC, and how inhibiting the gene can play a tumor suppressor role, which is all very similar to this study's discoveries [23].

Implications

ABCG2 can be identified as a gene with high therapeutic potential, and as a candidate for nanoparticle-related biomedical engineering strategies and possibilities. The nanoparticle albumin-bound paclitaxel (nab-paclitaxel) drug can be used to regulate ABCG2 and prevent transport across the membrane [14, 20]. The nanoparticle is suited to helping the drug get into and through the membrane, potentially improving the success of the drug. Overall, the bioinformatics analysis done in this research prioritizes ABCG2 as a key candidate for biomedical engineering research involving nano-particle based treatment development. GJA1 can also be recognized as a strong biomarker and research target for NSCLC tumor growth, as well as cancer cell communication and signaling.

Limitations

One limitation of this research is related to the nature of bioinformatics analysis. Due to using bioinformatics datasets, the research lacks in-vitro treatment development. The research requires further research in a laboratory to test the potential biomedical engineering strategies and develop a real treatment. Another limitation is regarding the format of the data in GEO2R. Although the results from GEO2R analysis classified ABCG2 as downregulated in experimental groups, a numerical fold-change value was unavailable for the full dataset; therefore, the magnitude of its decrease could not be determined.

Future Directions

There are various avenues of future research. The identified genes can first be tested by scientists in the laboratory or clinical trials to determine if they are suitable for biomedical engineering treatments to be based on them. This can help to confirm the significance of ABCG2 and GJA1 to NSCLC, and validate them as biomarkers. These biomarker genes can then be paired with the biomedical engineering guidance provided in the research to develop therapeutic plans that will address the challenges of NSCLC today, and/or create specific novel drugs or mechanisms to treat advanced NSCLC.

REFERENCES

1. Alduais Y, Zhang H, Fan F, Chen J, Chen B. Non-small cell lung cancer (NSCLC): A review of risk factors, diagnosis, and treatment. *Medicine* [Internet]. 2023 Feb 22;102(8):e32899.

2. Diebels I, Van Schil PEY. Diagnosis and treatment of non-small cell lung cancer: Current advances and challenges. *Journal of Thoracic Disease* [Internet]. 2022 June 30 [cited 2026 Aug 26];14(6):1753–7.
3. Riely GJ, Marks J, Pao W. KRAS mutations in non-small cell lung cancer. *Proceedings of the American Thoracic Society*. 2009 Apr 15;6(2):201–5.
4. Huang HN, Hung PF, Tsai YT, Liu ET, Cha TL, Chen YP, et al. LDOC1 connects histone H2B monoubiquitination to tumor cell plasticity in non-small cell lung cancer. *Cell Communication and Signaling* [Internet]. 2026 Jan 3 [cited 2026 Aug 26];24(1).
5. Clark, A. J., & Lillard, J. W., Jr (2024). A Comprehensive Review of Bioinformatics Tools for Genomic Biomarker Discovery Driving Precision Oncology. *Genes*, 15(8), 1036.
6. El-Sayes N, Vito A, Mossman K. Tumor heterogeneity: A great barrier in the age of cancer immunotherapy. *Cancers* [Internet]. 2021 Feb 15 [cited 2026 Aug 26];13(4):806.
7. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Research* [Internet]. 2012 Nov 26;41(D1):D991–5.
8. Huang Da W, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. DAVID bioinformatics resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Research* [Internet]. 2007 July;35(suppl_2):W169–75.
9. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* [Internet]. 1999 Jan 1;27(1):29–34.
10. Gene Ontology Consortium . The gene ontology (GO) database and informatics resource. *Nucleic Acids Research* [Internet]. 2004 Jan 1;32(90001):258D–261.
11. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguéz P, et al. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research* [Internet]. 2010 Nov 2;39(Database):D561–8.
12. Cannon M, Stevenson JG, Stahl K, Basu RK, Coffman AC, Kiwala S, et al. DGIdb 5.0: Rebuilding the drug–gene interaction database for precision medicine and drug discovery platforms. *Nucleic Acids Research* [Internet]. 2023 Nov 11;52(D1).
13. Gao Y, Liu S, Huang Y, Li F, Zhang Y. Regulation of anti-tumor immunity by metal ion in the tumor microenvironment. *Frontiers in Immunology* [Internet]. 2024 June 10;15.
14. Kundranda M, Niu J. Albumin-bound paclitaxel in solid tumors: Clinical development and future directions. *Drug Design, Development and Therapy* [Internet]. 2015 July;9:3767.
15. Dong J, Kislinger T, Jurisica I, Wigle DA. Lung cancer: Developmental networks gone awry? *Cancer Biology & Therapy* [Internet]. 2009 Feb 15;8(4):312–8.
16. Li X, Chen Y, Lan R, Liu P, Xiong K, Teng H, et al. Transmembrane mucins in lung adenocarcinoma: Understanding of current molecular mechanisms and clinical applications. *Cell Death Discovery* [Internet]. 2025 Apr 10;11(1).
17. Li P, Dong M, Chen B, Luo J, Jiang H, Lin N. SLC and ABC transporters in lung cancer: Orchestrators of malignancy and therapeutic resistance. *Comprehensive Physiology* [Internet]. 2026 Feb 27;16(2).

18. Xia X, Yang H, Au D, Lai S, Lin Y, Cho W. Membrane repairing capability of non-small cell lung cancer cells is regulated by drug resistance and epithelial-mesenchymal-transition. *Membranes* [Internet]. 2022 Apr 15;12(4):428.
19. Ma Y, Schröder DC, Nenkov M, Rizwan MN, Abubrig M, Sonnemann J, et al. Epithelial membrane protein 2 suppresses non-small cell lung cancer cell growth by inhibition of MAPK pathway. *International Journal of Molecular Sciences* [Internet]. 2021 Mar 14;22(6):2944.
20. Holder JE, Oliveira E, Lodeiro C, Trim CM, Byrne LJ, Emília Bértolo, et al. The use of nanoparticles for targeted drug delivery in non-small cell lung cancer. *Frontiers in Oncology* [Internet]. 2023 Mar 9;13.
21. Xiong X, Chen W, Chen C, Wu Q, He C. Analysis of the function and therapeutic strategy of connexin 43 from its subcellular localization. *Biochimie* [Internet]. 2023 Aug 22;218:1–7.
22. Jeleń A, Żebrowska-Nawrocka M, Łochowski M, Szmajda-Krygier D, Balcerczak E. ABCG2 gene expression in non-small cell lung cancer. *Biomedicines* [Internet]. 2024 Oct 19;12(10):2394.
23. Luo J, Jin Y, Li M, Dong L. Tumor suppressor miR-613 induces cisplatin sensitivity in non-small cell lung cancer cells by targeting GJA1 retraction in /10.3892/MMR.2024.13291. *Molecular Medicine Reports* [Internet]. 2021 Mar 18;23(5).