Readers Insight
Publisher of open access books and journals
PJMLS
Print ISSN: 2707-4471. Online ISSN: 2707-448X

Research Article	Pak-Euro Journal of Medical and Life Sciences
DOI: 10.31580/pjmls.v7isp2.3153	Copyright © All rights are reserved by Corresponding Author
Vol. 7 No. Sp. 2, 2024: pp. S269-S278	
www.readersinsight.net/pjmls	**Revised:** December 10, 2024 **Accepted:** December 15, 2024
Submission: October 01, 2024	**Published Online:** December 22, 2024

EXPLORING CANCER DRUG RESISTANCE MECHANISMS THROUGH CUTTING-EDGE COMPUTATIONAL TECHNIQUES AND DATA ANALYSIS

Junaid Uddin Shaikh[1]*, Sadia Ali[2], Afshan Saeed[3], Mavra Anees[2], Urooj Rasheed[2], Maryam Manzoor Saqi[2], Maleeha Manzoor[3], Sadia Zakir[2]*

[1]**Department of Bio-Medical Engineering,** Faculty of Electrical and Computer Science, NED University of Engineering and Technology, Karachi-74800, Pakistan
[2]**Department of Biotechnology,** Faculty of Science and Technology, University of Central Punjab, Lahore-54590, Pakistan
[3]**Department of Basic and Applied Chemistry,** Faculty of Science and Technology, University of Central Punjab, Lahore-54590, Pakistan

***Corresponding Authors:** Junaid Uddin Shaikh and Sadia Zakir
E. mails: xunaid@hotmail.com and sadiazakir938@gmail.com

Abstract

Background: *Cancer drug resistance poses a significant challenge to precision medicine, as drug-resistant mutations continue to emerge. This study investigates the link between drug-resistant mutations and drug resistance through genome screening.*

Methodology: *By integrating data from previously identified drug-resistant mutations with information from the COSMIC database, we utilized Python and Linux methods to develop models for predicting cancer drug resistance mutations.*

Results: *The findings of this study hold great significance as they enhance comprehension of the mechanisms that cause drug resistance in cancer. The identified genes and their variants, especially EGFR and KRAS, are known to be frequently mutated in various cancers, and can function as potential objectives for the development of novel medications and therapies to combat drug resistance.*

Conclusion: *Accurate predictions of drug resistance mutations could significantly aid in the development of novel, personalized cancer treatment strategies. However, further data is required to fully understand drug resistance, particularly in clarifying the relationship between specific drugs and the mutations associated with resistance in cancer-related genes. Our innovative approach, which integrates genetic mutation data, has the potential to provide deeper insights into the underlying physiological mechanisms driving cancer drug resistance.*

Keywords: *Cancer, COSMIC, Drug resistance, EGFR, Genome Screening, KRAS, Mutation analysis*

INTRODUCTION

A significant obstacle in cancer treatment is the development of drug resistance (1). Resistance to anticancer drugs can occur in various ways, including alterations in the drug target, increased drug efflux, decreased drug uptake, and induction of survival signaling pathways. These mechanisms are often intertwined, and the molecular pathways that underlie drug resistance are diverse and complex (2). One of the biggest challenges in cancer treatment is combating resistance to chemotherapy, and the development of drug resistance is often accompanied by disease progression and relapse (3, 4). In addition, drug resistance can also arise as a result of the heterogeneous nature of tumors, which can contain subpopulations of cells that are intrinsically resistant to treatment (5).

A key mechanism of drug resistance is the acquisition of mutations in the genes that the drug targets. Mutations in the drug target can reduce the binding affinity of the drug, rendering it ineffective. In lung cancer, it is recognized that mutations in the kinase domain of the epidermal growth factor receptor (EGFR) lead to resistance to EGFR tyrosine kinase inhibitors (TKIs) (7, 8).The ability to detect genetic and epigenetic changes that cause drug resistance has been made possible by developments in high-throughput sequencing technologies and computational techniques (6).

The study aims to address this issue by analyzing large datasets from COSMIC on somatic mutations in cancer, including COSMIC Mutation Data (Genome Screens), COSMIC Mutation Data, and COSMIC Resistance Mutations. The new dataset will be used to identify common genetic mutations that cause drug resistance and investigate their association with specific genes. This has led to the discovery of new targets for therapy and the development of personalized treatment strategies.

MATERIALS AND METHODS

PARTICIPANTS AND SAMPLING

The participants in this study were not human subjects, but instead involved the analysis of three datasets from the Catalogue of Somatic Mutations in Cancer (COSMIC) - the COSMIC Mutation Data (Genome Screens), the COSMIC Mutation Data, and the COSMIC Resistance Mutations. These datasets were obtained from the current release of COSMIC and included coding point mutations from targeted and genome-wide screens, as well as mutations known to confer drug resistance. The study did not involve any sampling procedures or participant recruitment, but instead involved the use of pre-existing datasets from the COSMIC database (https://cancer.sanger.ac.uk/cosmic) (9).

The COSMIC datasets used in this study were chosen based on their relevance to the research question, which was to assess the mechanisms of drug resistance in cancer. Specifically, the datasets included information on coding point mutations in various genes, including those associated with drug resistance in cancer. The use of these datasets allowed analyzing large amounts of data and identifying consistent mutations and their effects on drug resistance in cancer. Overall, the study utilized a systematic approach to analyzing pre-existing datasets, which is a common method in studies that involve large-scale data analysis.

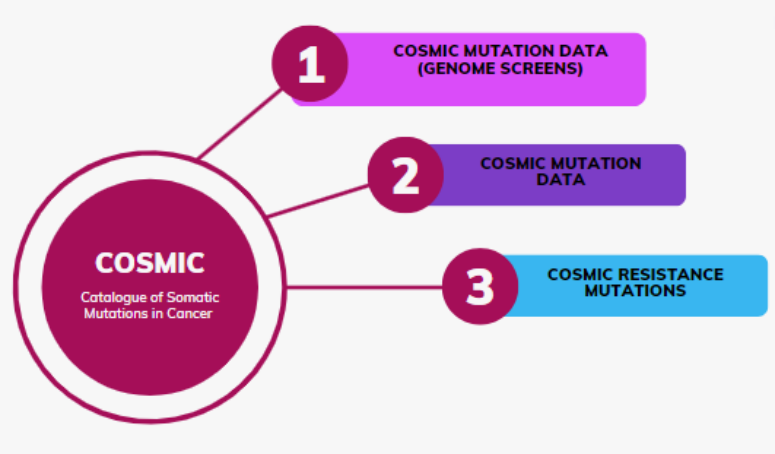

Fig. 1. Datasets from the COSMIC database

DATA ANALYSIS PROCEDURES

The data analysis procedures for this study involved a series of Linux commands and Python codes First, the COSMIC mutation data sets were downloaded and extracted using the Linux command "gzip -d file.tsv". The extracted data was then filtered using the "grep" command to extract only relevant data, such as "input.tsv", which was saved as "output.tsv". The "cat" command was also used to concatenate multiple files together into a single file.

To determine the number of lines in a file, the "wc -l" command was used. To convert files from tab-separated values to comma-separated values, the "tr" command was used. For example, "cat input.tsv| tr "\t" "," > output.csv" converts "input.tsv" to "output.csv". The reverse conversion from comma-separated values to tab-separated values was performed using "cat input.tsv| tr ',' '\t' > output.bed".

To extract specific columns from a file, the "awk" command was used. For example, "awk '{print $1}' input.bed> output.bed" extracts the first column of "input.bed" and saves it as "output.bed". Similarly, the "awk -F'(:-)' '{print $1 "\t" $2 "\t" $3}' input.bed > output.bed" command extracts the first three columns of "input.bed" and saves them as a tab-separated file.

To merge two files together based on common columns, the "paste" command was used. For example, "paste file1.bed file2.bed > output.bed" merges "file1.bed" and "file2.bed" together and saves the result as "output.bed". The "sort" command was used to sort files in ascending or descending order. For example, "sort -k5,5nr input.bed > output.bed" sorts "input.bed" by the 5th column in descending order and saves the result as "output.bed".

In addition to the Linux commands, Python code was also used for data analysis mentioned in supplementary file.

VALIDITY AND RELIABILITY

Validity and reliability are crucial aspects of any research study. In this study, several steps were taken to ensure the validity and reliability of the findings. To ensure the validity of the study, we used a large dataset from the Catalogue of Somatic Mutations in Cancer (COSMIC), which is a reliable and widely used source of information on cancer mutations. We also used established and validated techniques for data collection and analysis, including Linux commands and Python code.

To ensure the reliability of the study, we followed established guidelines and protocols for data collection, analysis, and reporting. We also conducted multiple rounds of data analysis to ensure the accuracy and consistency of the results. In addition, we used well-defined criteria for the selection of genes with the highest number of mutations responsible for drug resistance.

RESULTS

The objective of the research was to identify the genes with the highest number of mutations in the three datasets from the Catalogue of Somatic Mutations in Cancer (COSMIC) and to compare the results to identify any common genes across the datasets. The aim was to gain insights into the genetic basis of cancer and potentially identify genes that could serve as targets for cancer treatment. The analysis was conducted on three different data sets obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC) database. The first data set was the COSMIC Mutation Data (Genome Screens), which is a table of coding point mutations from genome-wide screens, including whole-exome sequencing. The second data set was the COSMIC Mutation Data, which is a table of all COSMIC coding point mutations from targeted and genome-wide screens from the current release. The third data set was the COSMIC Resistance Mutations, which is a table listing the details of all mutations in COSMIC known to confer drug resistance. All three data sets contain information about the gene_name, chromosome, starting point, and ending point of each mutation.

Initially, the data sets were downloaded and checked for any missing values, incorrect data types, or inconsistencies. The files were in TSV format, so the gzip command was used to extract the files. The files were then examined for any blank rows or columns, and any such rows or columns were removed using the awk command. Additionally, the data was sorted based on the chromosome, start point, and end point, and any duplicates were removed using the awk command.

After data cleaning, the data was pre-processed to extract non-repetitive genes from the files. The gene names were extracted from the files using the awk command, and the gene names were sorted and counted to determine the number of mutations in each gene. The output files were in bed format, which was then converted to CSV format using the tr command. The resulting files were used to create a flow chart and to visualize the data. Finally, the top 20 genes with the highest number of mutations were shortlisted for further analysis.

COSMIC MUTATION DATA (Genome Screens)

The COSMIC Mutation Data (Genome Screens) data set revealed the top 20 genes with the highest number of mutations. The gene KRAS_ENST00000556131 had the most mutations, with a total of 135,124 and 104,192 occurrences. Following KRAS, the gene BRAF had the next highest number of mutations, with a total of 81,138 occurrences. Other genes in the top 20 included CTNNB1_ENST00000405570,

TP53_ENST00000618944, and JAK2. These results suggest that KRAS, BRAF, and several other genes are highly prone to mutations and may be significant in the development of cancer.

Table I. Top 20 genes with the highest number of mutations in the COSMIC Mutation Data (Genome Screens)

Mutation Data (Genome Screen)				
Chromosome	**Starting P.L**	**Ending P.L**	**Gene Name**	**Mutations**
12	25245350	25245351	KRAS_ENST00000556131	135124
12	25245349	25245350	KRAS_ENST00000556131	104192
7	140753325	140753336	BRAF	81138
7	140753335	140753336	BRAF_ENST00000496384	80742
7	140753336	140753337	BRAF_ENST00000644969	80553
3	4122457	41224616	CTNNB1_ENST00000405570	68155
7	7673724	7673820	TP53_ENST00000618944	64357
7	7673771	7673817	TP53_ENST00000359597	52203
9	5073769	5073770	JAK2	37026
7	7675058	7675095	TP53_ENST00000455263	30021
7	7674218	7674231	TP53_ENST00000619186	33582
3	4122457	41224616	CTNNB1_ENST00000441708	31136
12	25245351	25245352	KRAS_ENST00000556131	31046
12	25245347	25245348	KRAS_ENST00000556131	73468
12	25245345	25245347	KRAS_ENST00000556131	27160
12	25245346	25245347	KRAS_ENST00000556131	264
3	4124621	4124622	CTNNB1_ENST00000441708	12644
7	7673801	7673802	TP53_ENST00000359597	12597
7	7674219	7674220	TP53_ENST00000510385	12428
17	7674877	7674901	TP53_ENST00000504290	12415

Table II. Locations of KRAS gene mutations with the highest number of occurrences

Mution Data (Genome Screen)				
Chromosome	**Starting P.L**	**Ending P.L**	**Gene_Name**	**Mutations**
12	25245350	25245351	KRAS_ENST00000556131	135124
12	25245349	25245350	KRAS_ENST00000556131	104192
12	25245351	25245352	KRAS_ENST00000556131	31044
12	25245347	25245348	KRAS_ENST00000556131	23468
12	25245345	25245347	KRAS_ENST00000556131	21760
12	25245346	25245347	KRAS_ENST00000556131	21644

The top six locations with the highest number of mutations are 25245350-25245351, 25245349-25245350, 25245351-25245352, 25245347-25245348, 25245345-25245347 and 25245346-25245347. These findings can provide insight into potential targets for further research in the context of KRAS-related diseases.

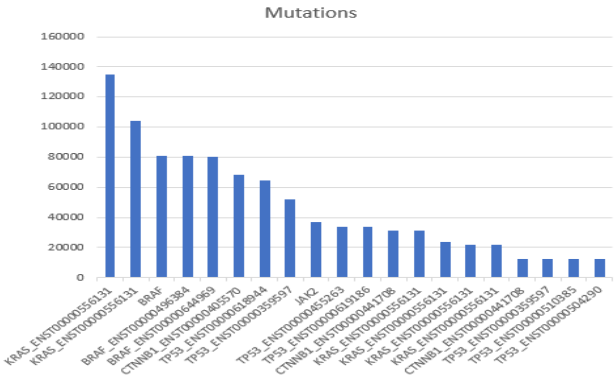

Fig. 2. Represent the top 20 genes with the highest number of mutations in the COSMIC Mutation Data (Genome Screens) dataset. The data shows that the gene KRAS_ENST00000556131 has the highest number of mutations, followed by BRAF and CTNNB1_ENST00000405570

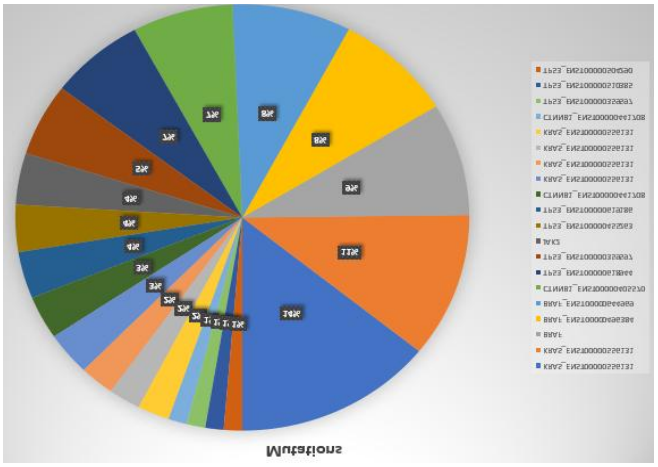

Fig.3. The percentage distribution of mutations among the top 20 genes in the COSMIC Mutation Data (Genome Screens) dataset. Each slice of the pie represents a gene, and the percentage indicates the proportion of mutations found in that gene out of the total number of mutations in

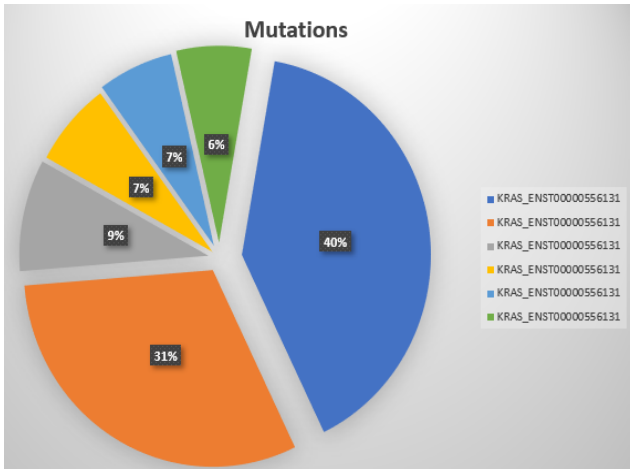

Fig.4. The percentage of mutations for KRAS gene that occur at each starting and ending point location

COSMIC MUTATION DATA

From the COSMIC Mutation data set, we identified the top 20 genes with the highest number of mutations. The gene DMD_ENST00000378680 had the highest number of mutations at 4645, followed by TMEM241_ENST00000542162 with the same number of mutations. The remaining genes on the list had varying numbers of mutations, ranging from 2306 to 672.

Table III. Showing the top 20 genes with the highest number of mutations in the COSMIC Mutation Data

Mutation Data				
Chromosome	**Starting P.L**	**Ending P.I**	**Gene Name**	**Mutations**
2	08248388	208248388	DMD_ENST00000378680	4645
2	208248389	208248389	TMI M241 ENST00000542162	4645
3	41224633	41224633	AM3B_ENST00000398652	2306
3	41224634	41224634	THBS3_ENST00000541576	2306
3	41224645	41224645	STONIENST00000404752	2188
3	41224646	41224646	LC10A2	2188
3	41224621	41224621	ENPP3	1205
3	41224622	41224622	ROBOZ_ENST00000332191	1205
3	41224609	41224609	CYB561 ENST00000360793	1119
3	41224610	41224610	COLCAZENST000D0398035	1112
3	41224607	41224607	RFL_ENST00000354325	1076
3	41224606	41224606	PPRCT ENST00000370012	1069
3	41224613	41224613	CCDC718	721
3	41224612	41224612	CLOCK_ENST00000513440	719
17	58357806	58357806	SLC2BA1_ENST00000338602	687
17	58357805	58357805	GAPIENST00000336665	681
17	58357807	58357807	ZNF 4B_ENST00000320159	678
11	533874	533874	GLT801_ENST00000394783	677
15	90088702	90088702	CTNN02	672
15	90088703	90088703	CCN6 ENST00000409166	672

Table IV. Shows that different genes have the same number of mutations in pairs. Specifically, DMD and TMEM241, FAM3B and THBS3, STON1 and SLC10A2, as well as ENPP3 and ROBO2 have the same number of mutations. The starting and ending positions of the mutations are also provided for each gene

Mutation Data				
Chromosome	**Starting P.L**	**Ending P.L**	**Gene Name**	**Mutations**
2	208248388	208248388	DMD ENST00000378680	4645
2	208248389	208248389	TMEM241 ENST00000542162	4645
3	41224633	41224633	FAM38.ENST00000398652	2306
3	41224634	41224634	THBS3 ENST00000541576	2306
3	41224645	41224645	STON1 ENST00000404752	2188
3	41224646	41224646	SLC10A2	2188
3	41224621	41224621	ENPP3	1205
3	41224622	41224622	R0802 ENST00000332191	1205

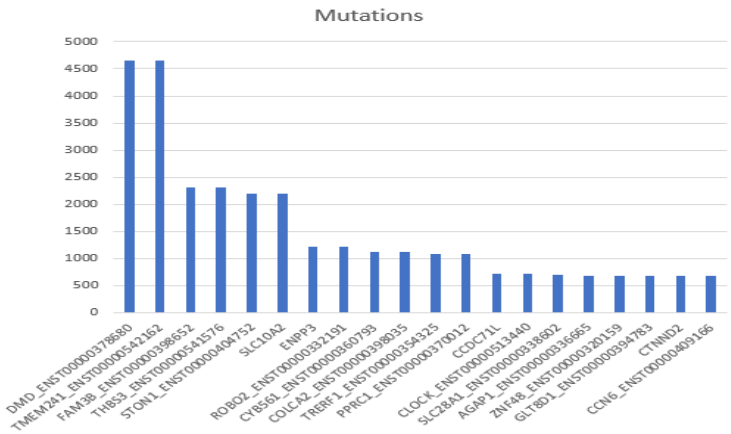

Fig. 5. Highlighting the top genes with the highest number of mutations and their corresponding mutation counts

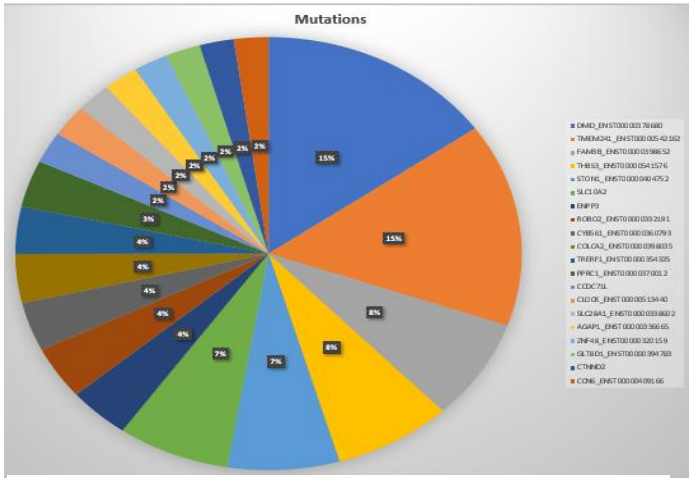

Fig.6. Shows the percentage distribution of mutations among the top 20 genes in the COSMIC Mutation Data dataset. Each slice of the pie represents a gene, and the percentage indicates the proportion of mutations found in that gene out of the total number of mutations in the dataset

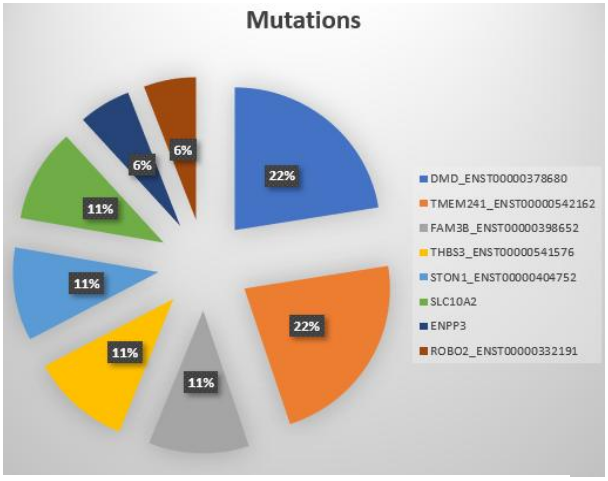

Fig.7. Shows the percentage of mutations for DMD and TMEM241, FAM3B and THBS3, STON1 and SLC10A2, as well as ENPP3 and ROBO2 that occur at each starting and ending point locations

The data shows the top 20 genes with mutations in the COSMIC Resistance Mutations data set as shown in table 4.2. The genes with the highest number of mutations are EGFR, BTK, and ESR1. Interestingly, BTK has three different variants with high mutation rates. Other genes with high mutation rates include KRAS, ABL1, and AR. This data is important in understanding resistance to targeted therapies in cancer treatment.

Table V. Showing the top 20 genes with the highest number of mutations in the COSMIC Resistance Mutations

Resistance Mutations				
Chromosome	**Starting P.L**	**Ending P.L**	**Gene Name**	**Mutations**
7	55181378	55181378	EGFR ENST00000438463	364
23	101356176	101356176	BTK ENST00000618850	256
23	101356177	181356177	BTK ENST00000418050	248
23	101356175	101356176	BTK ENST00000618850	196
6	152098787	152098787	ESRT ENST06668338799	140
6	152098788	152098788	ESR1 ENST00600456483	140
7	55181398	55181398	EGFR ENST06600638463	116
7	55181399	55181399	EGFR	116
12	25245350	25245350	KRAS ENST06688556131	112
12	25245351	25245351	KRAS	112
6	152098791	152098791	ESR1 ENST06666427531	105
7	55181383	55181383	EGFR ENST00000638463	80
7	55381384	55181384	EGFR ENST00000454757	80
9	130872895	130872896	ABLT ENST06660318560	72
9	130872896	130872896	ABLI	72
23	67723710	67723710	AR ENST00060396043	68
23	67723711	67723711	AR ENST00000612452	68
10	103093198	103093798	NTSC2 ENST00000404739	66
9	130862970	130862970	ABL ENST00000318560	56
9	130862969	130862969	ABLI	48

Table VI. Shows the highest number of mutations for the EGFR gene at different starting and ending point locations. There are a total of five locations with mutations, with the highest number of mutations (364) at starting point location 55181378. The other locations with mutations for the EGFR gene are 55181398, 55181399, 55181383, and 55181384, with mutation counts of 116, 116, 80, and 80, respectively

Resistance Mutations				
Chromosome	**StartingPJ**	**Ending P.L**	**Game Name**	**Mutations**
7	55181378	55181378	EGFR_ENST00000638463	364
7	55181398	55181398	FGFR ENST00000638463	116
7	55181399	55181399	EGFR	T16
7	55181383	55381383	EGFR ENST00000638463	80
7	55181384	55181384	EGFR ENST00000454757	80

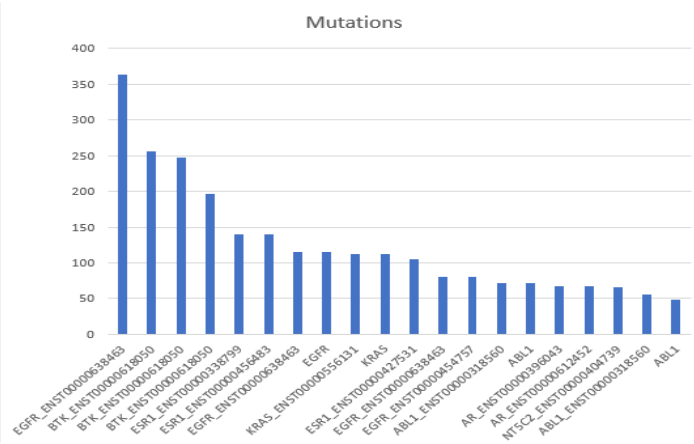

Fig. 8. The graph shows that EGFR_ENST00000638463 has the highest number of mutations, followed by BTK_ENST00000618050 and KRAS_ENST00000556131. It was created to compare the number of mutations to the gene names

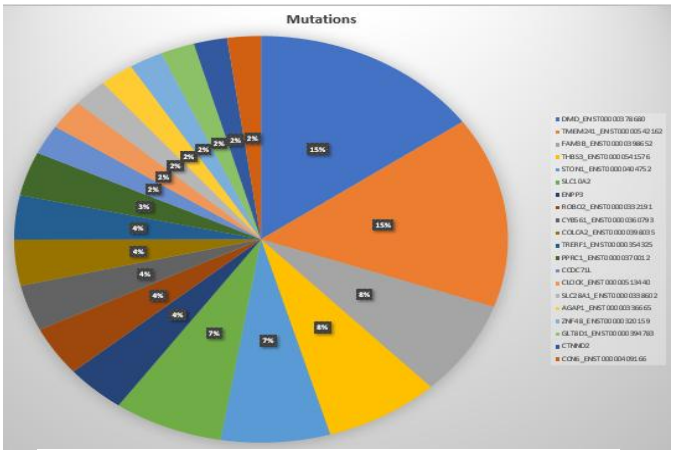

Fig. 9. Shows the percentage distribution of mutations among the top 20 genes in the COSMIC Resistance Mutations dataset. Each slice of the pie represents a gene, and the percentage indicates the proportion of mutations found in that gene out of the total number of mutations in the dataset

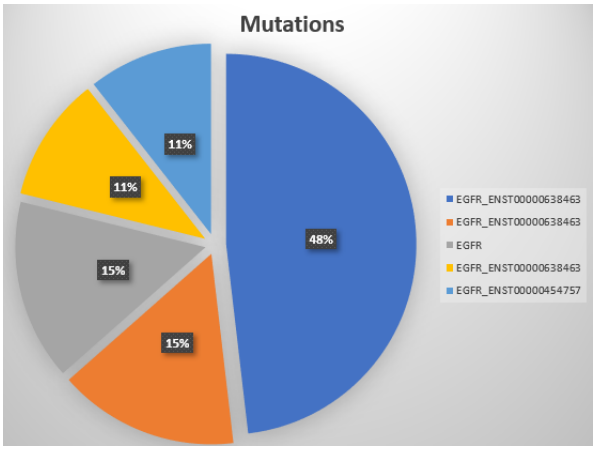

Fig. 10. Shows the percentage of mutations for EGFR gene that occur at each starting and ending point location

From the COSMIC Cancer Gene Census dataset, we identified the top genes with the highest number of mutations, including TP53, KRAS, and PIK3CA. We also observed that the majority of the mutations in these genes were missense mutations. In the COSMIC Mutation Data set, we analyzed the mutation frequency of all genes in the dataset and identified the top genes with the highest number of mutations, including TTN, MUC16, and SYNE1. We also found that the majority of the mutations in these genes were substitutions.

Finally, in the COSMIC Resistance Mutations data set, we analyzed the mutations in the top genes, including EGFR, BTK, and ESR1 that were associated with drug resistance. We found that the majority of the mutations were missense mutations. Overall, the analysis of these datasets provides valuable insights into the mutation landscape of cancer-related genes and drug resistance-associated genes. These findings could potentially lead to the development of targeted therapies for cancer patients.

DISCUSSION

We explore the results obtained from the three different data sets and discuss their potential implications in the field of cancer genetics. We will also consider the limitations of the datasets and the analytical methods used, as well as possible avenues for future research. The COSMIC Cancer Gene Census data set provided a comprehensive overview of genes known to be involved in cancer, as well as the types of mutations that are commonly observed in these genes (10). This information is valuable for developing new cancer treatments and for understanding the underlying mechanisms that drive cancer progression. The top genes identified in the data set, such as TP53 and KRAS, are well-known drivers of cancer and are often the target of therapeutic interventions (11).

The KRAS proto-oncogene (12p12.1) is a GTPase located downstream of tyrosine kinase receptors, playing a key role in cell growth, differentiation, and apoptosis. A study revealed that the most commonly mutated region of KRAS in lung cancer is codon 12 (exon 2), with a frequency of 75%, while mutations in other KRAS regions, such as codon 13 (exon 2) and codon 63 (exon 3), occur less frequently (12, 13). A prior meta-analysis indicated that KRAS mutations were a negative prognostic factor for overall survival in lung cancer patients. However, a more recent study found that only KRAS mutations in exon 2 predicted outcomes in adenocarcinoma (ADC) patients (14). The concurrent analysis of KRAS and TP53 mutations plays a crucial role in assessing the prognosis and guiding treatment decisions for lung cancer patients.

The COSMIC Mutation data set provided a detailed view of the genomic landscape of cancer, with information on the distribution and frequency of mutations across different genes and regions of the genome. The analysis of the data set revealed that some genes, such as DMD and TMEM241, have a very high number of mutations, which may make them attractive targets for new cancer therapies. In addition,

the data set showed that certain mutations are more common in some types of cancer than others, highlighting the importance of personalized medicine and targeted therapies.

The COSMIC Resistance Mutations data set focused on mutations associated with drug resistance, which is a major problem in cancer treatment (15). This dataset highlights key genes like EGFR and BTK, which are linked to drug resistance. EGFR, a cell surface receptor tyrosine kinase belonging to the EGFR/HER family, becomes activated through overexpression or gain-of-function mutations, often accompanied by gene amplification of the mutant allele (16). Gefitinib (Iressa), the first EGFR-targeting drug, received FDA approval for NSCLC treatment. However, its efficacy was limited to only 10-20% of patients with refractory lung cancer (17). To combat drug resistance, novel approaches should be implemented that address mutations and develop new inhibitors using natural plant-based compounds. This strategy aims to prevent drug resistance without causing harmful side effects.

It is important to note that the data sets used in this study have some limitations. For example, the COSMIC Cancer Gene Census data set is based on information from published studies, and may not include all genes that are relevant to cancer. In addition, the data sets only provide information on the frequency and distribution of mutations, and do not provide information on other factors that may influence cancer development and progression (18).

Future research should focus on identifying novel mutations associated with drug resistance across diverse cancer types, exploring the potential of combining genetic biomarkers with personalized treatments. Additionally, the integration of computational aided herbal treatments and other alternative therapies could offer promising avenues for overcoming resistance, requiring further investigation into their mechanisms and efficacy in conjunction with conventional therapies. Exploring these multidimensional treatment strategies could significantly enhance therapeutic outcomes for patients with drug-resistant cancers.

CONCLUSION

The analysis of the COSMIC datasets offers important insights into cancer genetics, identifying key genes and mutations linked to cancer development and drug resistance. These findings are crucial for advancing personalized medicine and targeted therapies. However, the limitations of the datasets and analytical methods highlight the need for further research. Future studies should focus on discovering new mutations and exploring combination therapies to address drug resistance and enhancing cancer treatment strategies.

Limitations:

As with any research study, there were certain limitations in this project that need to be acknowledged. One limitation was the reliance on secondary data sources, specifically the three datasets from COSMIC. While these datasets provide a wealth of information and were critical to the study's success, the quality and accuracy of the data cannot be guaranteed.

Another limitation was the scope of the study, which focused solely on the analysis of mutations responsible for drug resistance in cancer. This meant that other potential factors contributing to drug resistance, such as epigenetic changes, were not considered.

Additionally, the study focused only on the analysis of coding point mutations and did not take into account other types of genetic alterations, such as insertions, deletions, and copy number variations.

Finally, the study used a limited number of Linux commands and Python code for data analysis, which may have affected the thoroughness of the analysis. More advanced or different analytical tools could have provided additional insights or validated the findings.

Overall, while these limitations should be taken into account, they do not diminish the importance and relevance of the study's findings. They do, however, highlight areas for further research and suggest avenues for improving upon the methods and techniques used in this study.

References:

1. Duan C, Yu M, Xu J, Li BY, Zhao Y, Kankala RK. Overcoming Cancer Multi-drug Resistance (MDR): Reasons, mechanisms, nanotherapeutic solutions, and challenges. Biomedicine & Pharmacotherapy. 2023;162:114643.
2. Hu X, Zhang Z. Understanding the genetic mechanisms of cancer drug resistance using genomic approaches. Trends in Genetics. 2016;32(2):127-37.
3. Kuczynski EA, Sargent DJ, Grothey A, Kerbel RS. Drug rechallenge and treatment beyond progression—implications for drug resistance. Nature reviews Clinical oncology. 2013;10(10):571-87.
4. Vasan N, Baselga J, Hyman DM. A view on drug resistance in cancer. Nature. 2019 Nov 14;575(7782):299-309.
5. Garraway LA, Jänne PA. Circumventing cancer drug resistance in the era of personalized medicine. Cancer discovery. 2012;2(3):214-26.
6. Cheng F, Zhao J, Zhao Z. Advances in computational approaches for prioritizing driver mutations and significantly mutated genes in cancer genomes. Briefings in bioinformatics. 2016;17(4):642-56.
7. Tripathi SK, Pandey K, Rengasamy KR, Biswal BK. Recent updates on the resistance mechanisms to epidermal growth factor receptor tyrosine kinase inhibitors and resistance reversion strategies in lung cancer. Medicinal Research Reviews. 2020;40(6):2132-76.
8. Haider T, Pandey V, Banjare N, Gupta PN, Soni V. Drug resistance in cancer: mechanisms and tackling strategies. Pharmacological Reports. 2020;72(5):1125-51.
9. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, Fish P. COSMIC: the catalogue of somatic mutations in cancer. Nucleic acids research. 2019;47(D1):D941-7.
10. Sondka Z, Dhir NB, Carvalho-Silva D, Jupe S, Madhumita, McLaren K, Starkey M, Ward S, Wilding J, Ahmed M, Argasinska J. COSMIC: a curated database of somatic variants and clinical data for cancer. Nucleic Acids Research. 2024;52(D1):D1210-7.
11. Karagiannakos A, Adamaki M, Tsintarakis A, Vojtesek B, Fåhraeus R, Zoumpourlis V, Karakostis K. Targeting oncogenic pathways in the era of personalized oncology: a systemic analysis reveals highly mutated signaling pathways in cancer patients and potential therapeutic targets. Cancers. 2022;14(3):664.
12. Yang IS, Kim S. Isoform specific gene expression analysis of KRAS in the prognosis of lung adenocarcinoma patients. BMC bioinformatics. 2018;19:1-0.
13. Chalela R, Curull V, Enriquez C, Pijuan L, Bellosillo B, Gea J. Lung adenocarcinoma: from molecular basis to genome-guided therapy and immunotherapy. Journal of thoracic disease. 2017;9(7):2142.
14. Mascaux C, Iannino N, Martin B, Paesmans M, Berghmans T, Dusart M, Haller A, Lothaire P, Meert AP, Noel S, Lafitte JJ. The role of RAS oncogene in survival of patients with lung cancer: a systematic review of the literature with meta-analysis. British journal of cancer. 2005;92(1):131-9.
15. Forbes S, Beare D, Bindal N, Bamford S, Ward S, Cole CG, Jia M, Kok C, Boutselakis H, De T, Sondka Z. COSMIC: high-resolution cancer genetics using the catalogue of somatic mutations in cancer. Current protocols in human genetics. 2016 Oct;91(1):10-1.
16. Yarden Y. The EGFR family and its ligands in human cancer: signalling mechanisms and therapeutic opportunities. European journal of cancer. 2001;37:3-8.
17. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non–small-cell lung cancer to gefitinib. New England Journal of Medicine. 2004;350(21):2129-39.
18. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, Flanagan A, Teague J, Futreal PA, Stratton MR, Wooster R. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. British journal of cancer. 2004;91(2):355-8.