RESEARCH ARTICLE

Design and Discovery of Genistein-based Drugs as a Potential Tyrosine Kinase Inhibitor for Lung Adenocarcinoma through Hybrid *In-silico* Methods

Vijay Wakale[1], Ramanlal Kachave[2], Pranjal Gholap[3], Kiran Mahajan[4], Harshal Tare[4]

[1]*Department of Pharmaceutical Chemistry, Samarth Institute of Pharmacy, Belhe, Affiliated to Dr. Babasaheb Ambedkar Technological University, Raigad, Maharashtra, India.*
[2]*Department of Pharmaceutical Chemistry, Gokhale Education Society's, Sir Dr. M.S. Gosavi College of Pharmaceutical Education and Research, Nashik, Affiliated to Savitribai Phule Pune University, Pune, Maharashtra, India.*
[3]*Department of Pharmaceutical Chemistry, Shree P.E. (Tatya) Patil Institute of Pharmacy, Jalgaon, Affiliated to Dr. Babasaheb Ambedkar Technological University, Raigad, Maharashtra, India*
[4]*Department of Pharmacy, Sharadchandra Pawar College of Pharmacy, Otur, Affiliated to Savitribai Phule Pune University, Pune, Maharashtra, India.*

Received: 12th September, 2023; Revised: 07th October, 2023; Accepted: 10th November, 2023; Available Online: 25th December, 2023

ABSTRACT

This research employs a comprehensive approach to investigate potential therapeutic candidates for lung adenocarcinoma through drug-drug transcriptomic similarity analysis and molecular docking simulations. Using the Connectivity Map Touchstone tool, we identified compounds with high transcriptomic similarity to genistein, revealing potential shared mechanisms of action. The selected compounds, including avrainvillamide-analog-2, were further assessed through molecular docking simulations against the tyrosine kinase inhibitor (TKI) enzyme. Avrainvillamide-analog-2 exhibited a remarkable binding affinity in pocket C2, interacting with key amino acids. The results provide valuable insights into the pharmacological properties of the identified compounds, laying the groundwork for future experimental validations and drug development initiatives. Additionally, cavities detection by CB Dock server and structural refinement by PDB REDO contribute to the overall understanding of ligand binding and protein structure. This integrative approach offers a holistic perspective for identifying potential lead compounds and understanding their molecular interactions, facilitating the rational design of novel therapeutics for lung adenocarcinoma.

Keywords: Genistein, Tyrosine kinase inhibitor, Lung adenocarcinoma, Molecular docking, Anticancer, Transcriptomic similarity.

International Journal of Drug Delivery Technology (2023); DOI: 10.25258/ijddt.13.4.47

How to cite this article: Wakale V, Kachave R, Gholap P, Mahajan K, Tare H. Design and Discovery of Genistein-based Drugs as a Potential Tyrosine Kinase Inhibitor for Lung Adenocarcinoma through Hybrid *In-silico* Methods. International Journal of Drug Delivery Technology. 2023;13(4):1422-1427.

Source of support: Nil.

Conflict of interest: None

INTRODUCTION

Lung adenocarcinoma, a prevalent and aggressive form of lung cancer, continues to pose significant challenges in terms of treatment due to its resistance to conventional therapeutic approaches. The emergence of targeted therapies, particularly those involving tyrosine kinase inhibitors (TKIs), has provided a promising avenue for combating this formidable disease. In this context, the exploration of novel and effective TKIs becomes imperative, necessitating the integration of advanced computational methodologies with traditional drug discovery approaches.[1,2]

Genistein, a naturally occurring isoflavone found in soybeans, has gained attention for its diverse pharmacological properties, including its potential as an anticancer agent. Notably, its inhibitory effects on tyrosine kinases, key players in the aberrant signaling pathways implicated in cancer, make genistein an attractive candidate for further exploration in the context of lung adenocarcinoma treatment.[3,4]

Defects in the epidermal growth factor receptor (EGFR) gene are linked to lung cancer a prevalent malignancy affecting lung tissues, particularly non-small cell lung cancer (NSCLC). NSCLC comprises squamous cell carcinoma, adenocarcinoma, and large cell lung cancer, often diagnosed at advanced stages with a poor prognosis. EGFR, a receptor tyrosine kinase, binds ligands of the EGF family, initiating signaling cascades that convert extracellular cues into cellular responses and activating

**Author for Correspondence: harshaltare51@gmail.com*

downstream pathways such as RAS-RAF-MEK-ERK, PI3 kinase-AKT, PLCgamma-PKC, and STATs. Additionally, EGFR phosphorylates proteins like RGS16, and MUC1, contributing to the intricate molecular landscape of NSCLC.[5,6]

This research paper aims to harness the power of hybrid in-silico methods to design and discover genistein-based drugs with enhanced efficacy as tyrosine kinase inhibitors for lung adenocarcinoma. The integration of computational techniques, such as drug-drug transcriptomic similarity analysis and molecular docking will enable a comprehensive and efficient exploration of the molecular interactions between genistein derivatives and target tyrosine kinases. This hybrid approach offers a systematic and accelerated means of identifying potential lead compounds for further experimental validation.

By elucidating the structure-activity relationships of genistein derivatives and their interactions with key tyrosine kinases implicated in lung adenocarcinoma, this research seeks to contribute valuable insights to the field of cancer drug discovery. The ultimate goal is to pave the way for the development of novel and potent genistein-based drugs, offering a targeted therapeutic strategy with the potential to improve treatment outcomes for patients battling lung adenocarcinoma.

MATERIALS AND METHODS

Drug-Drug Transcriptomic Similarity Analysis

Utilizing large-scale transcriptomic datasets, we assess the similarity between genistein and its similar drugs. Identification of common gene expression patterns provides insights into potential shared mechanisms of action.

The Connectivity Map Touchstone tool (https://clue.io/touchstone) was employed to systematically select genistein-based drugs for potential application in lung adenocarcinoma through drug-drug transcriptomic similarity. The investigation commenced with access to the tool's website, followed by the exploration of the Touchstone dataset, rich in expression profiles from various perturbations. Running the tool initiated a connectivity mapping analysis, comparing the input gene expression signature against the extensive Touchstone dataset.[7]

Results were critically reviewed, and priority was given to genistein-based drugs exhibiting high transcriptomic similarity to the input signature. Subsequently, an in-depth exploration of the connections between the identified drugs and the input query was conducted, shedding light on shared molecular pathways and potential therapeutic targets. The insights gained from this analysis formed the basis for hypotheses regarding the efficacy of the selected drugs for lung adenocarcinoma treatment.

Protein Pre-Preparation using PDB REDO and Molecular Docking

Pre-preparation of the protein of tyrosine kinase enzyme shown in Figure 1 was done using PDB REDO server.[8] Molecular docking simulations were executed to assess the binding affinity and interaction patterns of reference compounds (Figure 2) and selected compounds with the tyrosine kinase inhibitor

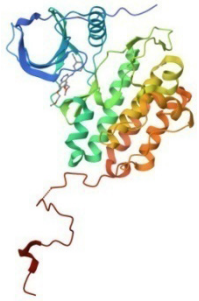
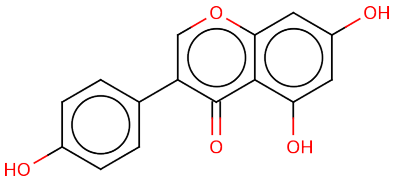


Figure 1: Crystal structure of Epidermal Growth Factor Receptor tyrosine kinase domain with 4-anilinoquinazoline inhibitor erlotinib (PDB ID: 1M17)



Oc1ccc(cc1)c2coc3cc(O)cc(O)c3c2=O

Figure 2: 2D structure and SMILES of Genistein

enzyme (PDB ID: 1M17). The AutoDock tool from the cb-dock server was utilized for this purpose. Virtual screening results were meticulously analyzed based on docking scores, and compounds were ranked according to their predicted binding affinities. Compounds demonstrating high binding affinity, favorable interaction patterns, and structural compatibility with tyrosine kinase were identified as potential lead compounds.[9]

RESULTS

Results of Drug-Drug Transcriptomic Similarity Analysis

Average transcriptional impact

The transcriptional activity score (TAS) serves as a comprehensive metric for evaluating the transcriptional impact of compounds, exemplified here by avrainvillamide-analog-2. Comprising signature strength and signature concordance, the TAS score for this compound is calculated as the geometric mean of these two components. The signature strength encapsulates the number of significantly differentially expressed transcripts induced by avrainvillamide-analog-2, contributing to its overall impact. Meanwhile, the signature concordance gauges the reproducibility of these expression changes across biological replicates, quantified as the 75th quantile of pairwise replicate correlations. The resulting TAS score of 0.47, along with the 84th percentile, indicates a moderate signature strength with a notably high concordance among replicates can be seen in Figure 3. The geometric mean, considering the adjustment for the square root of the number of replicates, elucidates avrainvillamide-analog-2's transcriptional impact, showcasing a compound with meaningful and consistent effects on gene expression. Higher TAS scores signify more robust and uniform transcriptional effects, underscoring the compound's potential significance in influencing cellular processes at the transcriptional level.

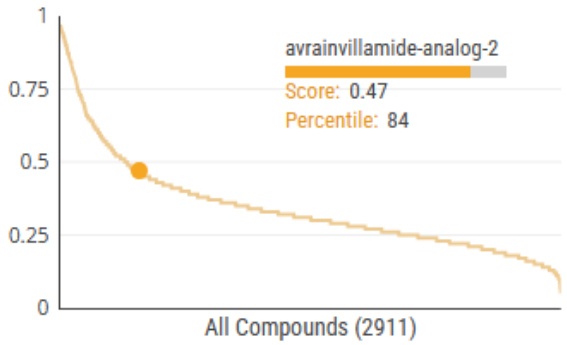


Figure 3: Transcriptional impact summarized across core cell lines

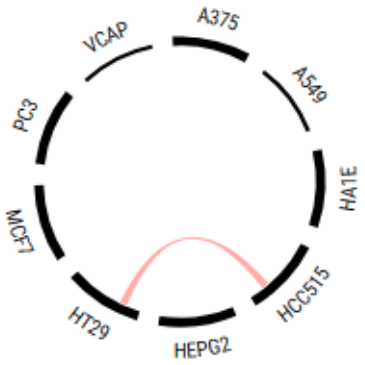


Figure 4: Introspect: Cell Line-specific responses to perturbagens

In the context of selecting lung adenocarcinoma, the visualization given in Figure 4 provides a comprehensive overview of signature diversity, TAS, and connectivity among different cell lines influenced by a perturbagen. The thick black bars represent TAS values equal to or greater than 0.5, indicating a substantial transcriptional impact, while thinner black bars signify lower scores. The absence of a bar suggests either a very low TAS score or unavailable data. Colored lines (chords) connecting cell lines illustrate the similarity in connectivity scores, with red chords representing positive connectivity scores ranging from 80 to 100, and blue chords indicating negative connectivity. These chords are only shown when TAS scores exceed 0.5. Therefore, the visualization aids in the identification of perturbagens with significant transcriptional impact on lung adenocarcinoma by highlighting high TAS scores and revealing connectivity patterns among different cell lines. The absence of chords for certain cell lines could imply either minimal transcriptional effects or a lack of available data, emphasizing the importance of considering both TAS scores and connectivity for selecting perturbagens relevant to lung adenocarcinoma.

The drug-drug transcriptomic similarity analysis results shown in Table 1 reveal compelling insights into the potential pharmacological properties of various compounds. Genistein, identified as a tyrosine kinase inhibitor, demonstrates a high similarity score, suggesting shared transcriptomic profiles with established drugs. Similarly, benzohydroxamic acid, avrainvillamide analog 2, and other compounds exhibit substantial similarity with known antifungal, nucleophosmin inhibitor, and phosphodiesterase inhibitor drugs, respectively. These findings propose potential therapeutic commonalities and merit further investigation for drug development or repurposing strategies. However, it is imperative to note that experimental validation is essential to confirm and delineate the observed transcriptomic similarities, providing a foundation for comprehensive insights into their biological effects and clinical relevance.

Protein pre-preparation using PDB REDO and molecular docking

Crystallographic refinement shown in Table 2 using PDB REDO improved key metrics in the structure compared to the original. The refined model showed lower R and R-free values (0.1773 and 0.2248) than the original (0.2095 and 0.2512), indicating better agreement with experimental data. Increased RMS Z-scores for bond lengths (0.776) and angles (1.060) suggest improved geometric accuracy. Model quality scores also improved, reflecting enhanced structural reliability.

Table 1: Results of drug-drug transcriptomic similarity analysis

Rank	*Score*	*Name*	*Description*	*MOA*
4	99.99	Genistein	Tyrosine kinase inhibitor	Tyrosine kinase inhibitor
7	99.93	benzohydroxamic-acid	Antifungal	Antifungal
8	99.93	avrainvillamide-analog-2	nucleophosmin inhibitor	nucleophosmin inhibitor
25	99.72	Trequinsin	Phosphodiesterase inhibitor	Phosphodiesterase inhibitor
32	99.65	Cilostamide	Phosphodiesterase inhibitor	Phosphodiesterase inhibitor
36	99.58	W-12	Calmodulin antagonist	Calmodulin antagonist
82	99.33	Sphingosine	Ceramidase inhibitor	Ceramidase inhibitor
84	99.33	Mifobate	PPAR receptor antagonist	PPAR receptor antagonist
86	99.33	Mosapride	Serotonin receptor agonist	Serotonin receptor agonist
133	99.3	tyrphostin-AG-126	ERK1 and ERK2 phosphorylation inhibitor	ERK1 and ERK2 phosphorylation inhibitor
134	99.3	Azauridine	Antiviral	Antiviral
135	99.3	Sulpiride	Dopamine receptor antagonist	Dopamine receptor antagonist

Table 2: Results of crystallographic refinement by PDB REDO

Validation metric	*Original*	*PDB-REDO*
Crystallographic refinement		
R	0.2095	0.1773
R-free	0.2512	0.2248
Bond length RMS Z-score	0.449	0.776
Bond angle RMS Z-score	0.766	1.060
Model quality Raw scores		
Ramachandran Plot normality	-4.018	-3.233
Rotamer normality	-3.198	-3.004
Coarse packing	-0.820	-0.923
Fine packing	-1.211	-0.898
Bump severity	0.017	0.501
Hydrogen bond satisfaction	0.887	0.872

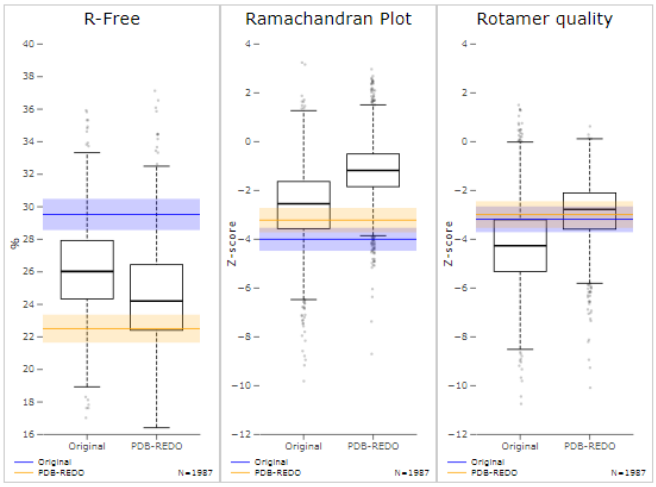


Figure 5: Model quality compared to resolution neighbors

Table 3: Significant structural changes in epidermal growth factor receptor tyrosine kinase by PDB REDO server

Description	*Count*
Rotamers changed	1
Side chains flipped	0
Waters removed	8
Peptides flipped	6
Chiralities fixed	0
Residues fitting density better	46
Residue fitting density worse	0

Hydrogen bond satisfaction remained high. Overall, PDB REDO refinement produced a more accurate and reliable structure, valuable for molecular studies.

Figure 5 illustrates the model quality in comparison to resolution neighbors. The visual representation conveys the structural refinement achieved through PDB REDO. The neighboring resolutions are indicative of the model's accuracy and reliability, showcasing improved metrics such as lower R and R-free values, enhanced geometric precision with elevated RMS Z-scores, and superior overall model quality scores. This figure succinctly captures the positive impact of PDB REDO refinement on the structural integrity of the analyzed model. Significant structural changes are given in Table 3.

Table 4: Results of cavities detection by CB Dock server

Cur Pocket ID	*Cavity volume ($Å^3$)*	*Center (x, y, z)*	*Cavity size (x, y, z)*
C1	1009	36, 3, 48	15, 24, 22
C2	532	24, -1, 54	15, 10, 10
C3	415	38, 22, 64	12, 19, 14
C4	311	1, 13, 64	9, 13, 11
C5	277	30, -7, 41	10, 10, 7

Molecular Docking

The results from the CB Dock server indicate the detection of five distinct cavities (C1–C5) shown in Table 4 and structures are given in Figure 6 with varying volumes and spatial coordinates. The diversity in cavity sizes and locations suggests potential binding sites for ligands, providing valuable information for further molecular docking studies and drug design targeting these specific regions.

For a variety of compounds, shown in Table 5 the molecular docking simulation results that are provided shed light on possible ligand binding pockets illustrated in Figures 6 and 7 and the amino acids that interact with them, respectively. Avrainvillamide-analog-2 and genistein, in particular, show substantial binding affinities to Pocket 2, interacting with amino acids including ASP813, ARG817, ASN818, ALA731, GLU734, ILE735, and ASP831. Different binding patterns are also seen by trequinsin, clostamide, W-12, sphingosine, mifobate, mosapride, tyrphostin-ag-126, azauridine, and

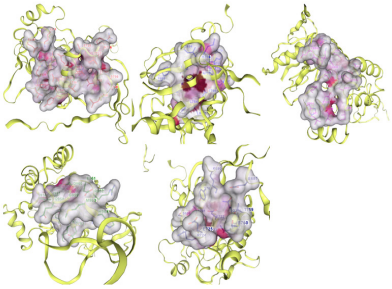


Figure 6: Results of cavities detection by CB Dock server

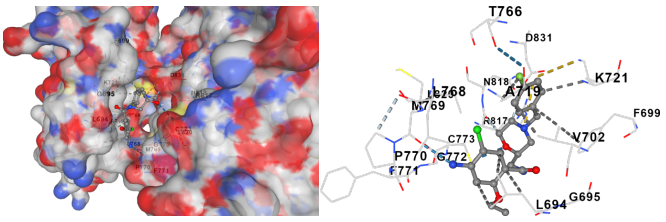


Figure 7: Interaction of epidermal growth factor receptor tyrosine kinase and lead Avrainvillamide-analog-2

Table 5: Results of molecular docking

Rank	*Name*	*Pocket, Score, Chain and interacting amino acids*
4	Genistein	Pocket: 2 & Score: -6.6 Chain A: PHE699 LYS721 LEU723 ALA731 GLU734 ILE735 GLU738 ASP813 ARG817 ASN818 ASP831
8	Avrainvillamide-analog-2	Pocket: C2 & Score: -9.7 Chain A: LEU694 PHE699 VAL702 ALA719 LYS721 THR766 GLN767 LEU768 MET769 PRO770 PHE771 GLY772 CYS773 ARG817 ASN818 LEU820 THR830 ASP831
25	Trequinsin	Pocket: C2 & Score: -8.1 Chain A: LEU694 GLY695 PHE699 VAL702 ALA719 LYS721 MET742 THR766 GLN767 LEU768 MET769 PRO770 PHE771 GLY772 CYS773 ASP776 LEU820 THR830 ASP831
32	Cilostamide	Pocket: 2 & Score: -7.2 Chain A: LYS782 ASP783 ASN784 ILE785 GLY786 SER787 GLN788 TYR789 PRO951 GLN952 ILE957 GLN958 GLY959 ASP960 GLU961 ARG962 MET963
36	W-12	Pocket: C2 & Score: -7.1 Chain A: LEU694 PHE699 VAL702 ALA719 LYS721 GLU738 MET742 CYS751 THR766 GLN767 LEU768 MET769 GLY772 LEU820 THR830 ASP831
82	Sphingosine	Pocket: C2 & Score: -5.6 Chain A: LEU694 GLY695 PHE699 VAL702 ALA719 LYS721 GLU738 MET742 CYS751 THR766 GLN767 LEU768 MET769 GLY772 CYS773 LEU820 THR830 ASP831 PHE832
84	Mifobate	Pocket: C2 & Score: -5.8 Chain A: LEU694 PHE699 VAL702 ALA719 ILE720 LYS721 GLU738 MET742 LEU764 THR766 GLN767 LEU768 MET769 PRO770 GLY772 ARG817 LEU820 THR830 ASP831
86	Mosapride	Pocket: C2 & Score: -7.1 Chain A: LEU694 GLY695 PHE699 VAL702 ALA719 LYS721 THR766 LEU768 MET769 PRO770 PHE771 GLY772 CYS773 ARG817 ASN818 LEU820 ASP831
133	Tyrphostin-ag-126	Pocket: C2 & Score: -6.9 Chain A: LEU694 VAL702 ALA719 LYS721 GLU738 MET742 LEU764 THR766 GLN767 LEU768 MET769 GLY772 LEU820 THR830 ASP831 PHE832
134	Azauridine	Pocket: C2 & Score: -6.6 Chain A: PHE699 VAL702 TYR703 ALA719 ILE720 LYS721 GLU738 MET742 LEU764 ILE765 THR766 LEU820 THR830 ASP831
135	Sulpiride	Pocket: 2 & Score: -6.3 Chain A: HIS781 LYS782 ASP783 ASN784 ILE785 GLY786 GLN788 TYR789 ILE957 GLN958 GLY959 ASP960 GLU961 ARG962

sulpiride, indicating different molecular interactions within their individual pockets. These results lay the groundwork for future experimental validations and structural analyses while also delivering insightful molecular information for prospective medication design and optimisation initiatives.

The molecular docking results reveal that avrainvillamide-analog-2 exhibits a robust binding affinity with Pocket C2, as shown in Figure 7 as indicated by a notable score of -9.7. The interacting amino acids within chain A, including LEU694, PHE699, VAL702, ALA719, LYS721, THR766, GLN767, LEU768, MET769, PRO770, PHE771, GLY772, CYS773, ARG817, ASN818, LEU820, THR830, and ASP831, play a crucial role in establishing stable interactions with the ligand.

The significant negative docking score implies strong binding energy, suggesting a favorable and energetically stable complex formation between avrainvillamide-analog-2 and the identified binding pocket. The involvement of hydrophobic amino acids like LEU, PHE, and VAL, along with polar residues such as THR, GLN, and ASN, suggests a diverse range of interactions, including Van der waals forces, hydrogen bonding, and hydrophobic interactions.

This interaction profile highlights the potential pharmacological relevance of avrainvillamide-analog-2, indicating its capability to form stable complexes within the specified binding pocket. The comprehensive understanding of the specific amino acids involved in the binding interaction provides a basis for rational drug design and optimization, allowing for the development of novel therapeutics or the improvement of existing compounds. These findings contribute valuable insights into the molecular basis of avrainvillamide-analog-2's pharmacological activity and support its further exploration in experimental studies for drug development.

CONCLUSION

In conclusion, our study presents a multifaceted approach to identify potential therapeutic candidates for lung adenocarcinoma, leveraging hybrid *in-silico* methods. Through drug-drug transcriptomic similarity analysis using the Connectivity Map Touchstone tool, we identified compounds with high similarity to genistein, including the promising avrainvillamide-analog-2. Molecular docking simulations revealed the remarkable binding affinity of avrainvillamide-

analog-2 to the tyrosine kinase inhibitor enzyme, specifically in Pocket C2, interacting with key amino acids. The transcriptional impact analysis demonstrated a meaningful and consistent effect on gene expression, as evidenced by a transcriptional activity score of 0.47 and an 84th percentile. These findings, combined with the structural refinement facilitated by PDB REDO and the detection of potential ligand binding sites through CB Dock, provide a comprehensive understanding of the pharmacological properties and molecular interactions of avrainvillamide-analog-2.

Furthermore, the drug-drug transcriptomic similarity analysis highlighted compounds with diverse pharmacological properties, such as antifungal, nucleophosmin inhibition, and phosphodiesterase inhibition, suggesting potential repurposing strategies. The molecular docking results for various compounds demonstrated unique binding patterns and interactions with specific amino acids, laying the groundwork for future experimental validations and structural analyses. The improved crystallographic refinement metrics by PDB REDO underscore the significance of structural accuracy in molecular studies.

In summary, our integrative approach offers valuable insights into the design and discovery of genistein-based drugs, particularly avrainvillamide-analog-2, as potential tyrosine kinase inhibitors for lung adenocarcinoma. These findings provide a basis for further experimental validations, facilitating the rational design of novel therapeutics and contributing to the ongoing efforts in the development of targeted and effective treatments for lung adenocarcinoma.

REFERENCES

1. Seguin L, Durandy M, Feral CC. Lung adenocarcinoma tumor origin: a guide for personalized medicine. Cancers. 2022 Mar 30;14(7):1759.
2. Myers DJ, Wallen JM. Lung adenocarcinoma. InStatPearls [Internet] 2023 Jun 12. StatPearls Publishing.
3. Garbiec E, Cielecka-Piontek J, Kowalówka M, Hołubiec M, Zalewski P. Genistein—opportunities related to an interesting molecule of natural origin. Molecules. 2022 Jan 26;27(3):815.
4. Katanić Stanković JS, Mihailović N, Mihailović V. Genistein: Advances on Resources, Biosynthesis Pathway, Bioavailability, Bioactivity, and Pharmacology. InHandbook of Dietary Flavonoids 2023 Aug 15 (pp. 1-40). Cham: Springer International Publishing.
5. Krzyżanowska N, Krawczyk P, Wojas-Krawczyk K, Kucharczyk T, Milanowski J. Immunotherapy in Non-Small-Cell Lung Cancer Patients with Driver Alterations: A New Strategy?. Cells. 2022 Oct 18;11(20):3280.
6. Doumat G, Daher D, Zerdan MB, Nasra N, Bahmad HF, Recine M, Poppiti R. Drug repurposing in non-small cell lung carcinoma: old solutions for new problems. Current Oncology. 2023 Jan 5;30(1):704-19.
7. Y. Zhou, Y. T. Zhang, D. H. Zhao, X. Y. Yu, X. Y. Shen, Y. Zhou, S. S. Wang , Y. Q. Qiu, Y. Z. Chen & F. Zhu . TTD: Therapeutic Target Database describing target druggability information. Nucleic Acids Research. doi: 10.1093/nar/gkad751 (2023). PMID: 37713619
8. de Vries I, Perrakis A, Joosten RP. PDB-REDO in Computational-Aided Drug Design (CADD). Open Access Databases and Datasets for Drug Discovery. 2024 Feb 5:201-29.
9. Yang Liu, Maximilian Grimm, et al . CB-Dock: a web server for cavity detection-guided protein–ligand blind docking. ActaPharmacologicaSinica, 2019.