GMR
The Original

PHARMACOGENOMIC VARIATIONS AND DRUG RESPONSE: A MOLECULAR STUDY IN PERSONALISED MEDICINE

Dr. Guru K[1], Dr Mohammad Manhar[2], Suryanarayana Reddy Kovvuri[3], Dr Divyangkumar Patel[4]

[1]Assistant professor, Vels medical college & Hospital, Tiruvallur 601 102 Department of Anaesthesiology, Mail id: gurujipmer@gmail.com ORCID ID - 0009-0005-1618-6334
[2]Assistant Professor, IIMSR Lucknow Integral University Lucknow E-mail: manhar003@gmail.com, ORCID ID: 0009-0007-2379-0446
[3]Post Graduate, Department of general medicine, Saveetha medical college, Chennai, Tamil Nadu. Email - drsurya9999@gmail.com ORCID ID - 0009-0001-7636-7138
[4]Professor and Head, Department of Community Medicine Department of Community Medicine, Dr. N. D. Desai Faculty of Medical Science and Research, Dharmsinh Desai University. Email I'd: drdivpatel@gmail.com, ORCID ID: 0000-0001-7307-1887

ABSTRACT

Variability in anticancer drug response presents a significant challenge in oncology, necessitating the integration of pharmacogenomics to enable personalised treatment strategies. This study aimed to investigate the molecular determinants of drug response using a secondary pharmacogenomic analysis of the Genomics of Drug Sensitivity in Cancer (GDSC) dataset. A total of 232,669 drug–cell line observations were analysed, incorporating multi-omics features including gene expression, copy number alterations (CNA), and DNA methylation. Linear regression models were applied at both global and drug-specific levels to assess associations with drug response, measured by log-transformed IC50 (LN_IC50). The global model demonstrated statistically significant associations ($p < 0.001$) but limited explanatory power ($R^2 = 0.001$). In contrast, drug-wise analyses improved model performance (R^2 up to 0.036) and identified 95 drugs with significant pharmacogenomic associations. Gene expression emerged as the most consistent predictor, predominantly associated with increased drug resistance, while CNA and methylation exhibited context-dependent effects. These findings highlight the importance of multi-omics integration and drug-specific modeling in understanding pharmacogenomic variability. The study underscores the limitations of aggregated analyses and supports precision oncology approaches for biomarker-driven therapy. Overall, this work contributes to advancing personalised medicine by identifying molecular features that influence anticancer drug response.

KEYWORDS: Pharmacogenomics, Drug Response, GDSC, Gene Expression, Precision Oncology

1. INTRODUCTION

The problem of variability of anticancer drug response remains a major issue in the field of oncology because patients undergoing the same drug regimen often demonstrate quite different responses in terms of efficacy and toxicity. The heterogeneity is propelled by complex biological factors, such as genetic, epigenetic, and molecular differences in tumor cells, which drive both the pharmacodynamics and pharmacokinetics. This variability not only undermines the efficacy of treatment but also leads to adverse drug reactions, which increase morbidity and health care burden (Carr et al., 2021). To address these issues, the old-fashioned approach to cancer treatment which is one-size-fits-all is being substituted by precision oncology which is aimed at tailoring therapies to individual molecular features. Pharmacogenomics has come to occupy a central place in this paradigm, as it then allows the identification of the biomarkers that predict drug response and guide therapeutic decision-making (White et al., 2022; Asghar et al., 2025).

In the case of cancer, pharmacogenomics does not merely consider single nucleotide polymorphisms (SNPs) but rather extends to a wider scope of molecular changes that affect drug sensitivity and resistance. An example of this is the expression profiles of genes, which are important in modulating the activity of drug targets and downstream signaling pathways, thereby affecting treatment efficacy (Merrill et al., 2020). Likewise, the amplification or deletion of genes related to the processes of drug metabolism, transport, or target pathways can be caused by copy number alterations (CNAs). This ultimately alters the response of cells to therapy (Luo, 2019). Additional complexity is also provided by epigenetic processes, in particular, the DNA methylation that regulates gene expression without modifying the DNA sequence, therefore, affecting the tumor behavior and the outcome of therapy (Noshay and Springer, 2021). These molecular characteristics play interactions in complex biological networks, where dysregulation of critical signaling pathways such as PI3K/AKT, MAPK, and DNA repair mechanisms have been linked to resistance to anticancer agents (Saeed et al., 2018). Thus, integrative studies involving multiple layers of molecular data can offer a more comprehensive understanding of pharmacogenomic variability compared to studies that analyze this variability through the analysis of single genetic variants (Salvatore et al., 2019).

The use of high-throughput technologies and large-scale data generation has enabled the creation of comprehensive pharmacogenomic resources, one of which is the Genomics of Drug Sensitivity in Cancer (GDSC) database. GDSC combines comprehensive drug screening data with detailed molecular profiling of cancer cell lines, such as expression of

genes, copy number variation and DNA methylation data. The integration facilitates methodical exploration of the connections among molecular characteristics and drug reaction across a large range of anticancer compounds and malignancies. GDSC can be used to quantitatively evaluate drug sensitivity and can be used to perform high quality statistical analysis of pharmacogenomic associations. Further, the control environment provided by the use of cancer cell lines as model systems offers control over the environment to explore the molecular determinants of drug response, thereby facilitating translational insights into clinical oncology (Lee et al., 2018; Laes et al., 2018).
It is important to leverage such integrative datasets in order to identify predictive biomarkers, as well as in order to discover mechanisms underlying drug resistance, which are critical in advancing personalised medicine. The current study employs the GDSC dataset to study the molecular determinants of drug response by integrating multi-omics data with the aim of advancing our understanding of pharmacogenomic variability and contributing to the development of more effective, individualised treatments of cancer.

Research Objectives
1. To determine the association between multi-omics molecular features (gene expression, copy number alterations, and DNA methylation) and anticancer drug response using GDSC data
2. To identify key molecular pathways driving drug sensitivity and resistance through integrative pharmacogenomic analysis
3. To develop and validate predictive models of drug response based on molecular biomarkers for personalised oncology applications

2. MATERIALS AND METHODS

2.1 Study Design and Data Source
This research is a secondary, in silico, pharmacogenomic study which was designed to determine molecular determinants of anticancer drug response. The Genomics of Drug Sensitivity in Cancer (GDSC) dataset, which is available on Kaggle, was used to obtain the data. The dataset comprises of about 969 cancer cell lines and more than 250 anticancer drugs. The standardized measures of drug response include the log-transformed IC50 (LN_IC50) and area under the curve (AUC). This dataset can be used to conduct a systematic exploration of pharmacogenomic associations in a controlled experiment. No ethical approval was needed as publicly available data were utilized.

2.2 Variables and Data Preprocessing
Drug response as LNIC50 and AUC as a secondary measure was the primary outcome variable. Independent variables were the gene expression profiles, copy number alterations (CNA), and DNA methylation profiles, and covariates were the type and tissue origin of the cancer. Data preprocessing was done to process missing values; this was done by imputation or exclusion of missing values depending on data completeness. Continuous variables were normalized and scaled to minimize technical variability and improve comparability. Features that had small variance were filtered to decrease noise and computational load. The interquartile range (IQR) was used to evaluate the outliers to guarantee soundness of downstream analyses.

2.3 Statistical Association Analysis
The relationships between molecular features and drug response were assessed with the help of the linear regression models in which LN_IC50 was taken as a dependent variable. The molecular features were analysed separately and controlling the covariates that were relevant like the cancer type. Since the data were high-dimensional, multiple hypothesis testing was considered with the false discovery rate (FDR) correction of the Benjamini-Hochberg. The adjusted p-value of less than 0.05 was set as the statistically significant association. The method offered discovery of molecular biomarkers that play a major role in determining the drug sensitivity or resistance in the cell lines.

2.4 Multi-Omics Integration and Pathway Analysis
To determine the complexity of the pharmacogenomic interactions the integrative multi-omics approach was used which combined the data of gene expression, CNA and methylation. To identify which of the various omics layers contributes to the variability of drug responses, comparative analyses were made. Pathway enrichment analysis of genes of much interest in relation to drug response was done using databases like KEGG and Reactome. Over-representation analysis was carried out to determine the biological pathways that play a role in the mechanisms of drug sensitivity and resistance. The outcome of enrichment was adjusted to multiple testing and pathways with FDR-adjusted p-values of below 0.05 were considered significant.

2.5 Predictive Modeling and Validation
To assess the translational potential of identified biomarkers, predictive models of drug response were developed using machine learning techniques. The selection and regularization of the feature were achieved with the help of Elastic Net regression, whereas the non-linear relationships were represented with the help of the Random Forest models. K-fold cross-validation ($k = 5$ or 10) was used to test model performance in order to generalize the results. The measures of performance were coefficient of determination (R^2) and root mean square error (RMSE). The scores of feature importance were mined out to determine critical predictors of drug response, in support of the generation of personalised oncology approaches.

3. RESULTS

3.1 Dataset Characteristics and Preprocessing

A total of 242,035 drug–cell line observations were obtained from the GDSC dataset. Following the preprocessing and elimination of the incomplete molecular profiles, 232,669 observations were retained to be analyzed. The dataset comprised various cancer types, drugs, and molecular features (CNA, gene expression, and DNA methylation), giving it a solid foundation to model pharmacogenomics.

Table 3.1: Summary of Dataset After Preprocessing

Parameter	Value
Initial observations	242,035
Final observations	232,669
Number of drugs	~250
Molecular features	CNA, Gene Expression, Methylation
Outcome variables	LN_IC50, AUC

3.2 Global Association Between Molecular Features and Drug Response

A multivariate linear regression model was used to evaluate the overall correlation between the molecular features and drug response (LN_IC50). The model was significant ($F = 50.80$, $p < 0.001$) yet there was very limited explanatory power ($R\ 2 = 0.001$), which showed that there is a weak global effect when all drugs are combined.

Table 3.2: Global Regression Model Results

Variable	Coefficient (β)	p-value	Interpretation
CNA	-0.570	<0.001	Increased sensitivity
Gene Expression	+0.433	<0.001	Increased resistance
Methylation	-0.077	0.023	Weak effect

3.3 Drug-Specific Pharmacogenomic Associations

A multivariate linear regression model was used to evaluate the overall correlation between the molecular features and drug response. The model was not only significant ($F = 50.80$, $p < 0.001$), but it also had very limited explanatory power ($R\ 2 = 0.001$), thus demonstrating that there is a weak global effect when all the drugs are combined.

Table 3.3: Top Drug-Specific Models

Drug Name	N	R^2	Gene Expression (p)	CNA (p)	Methylation (p)
LY2109761	858	0.0356	<0.001	0.407	0.246
XAV939	859	0.0335	<0.001	0.577	0.935
Methotrexate	845	0.0326	0.001	0.171	0.013
Tozasertib	254	0.0324	0.007	0.399	0.605
EPZ5676	923	0.0287	<0.001	0.481	0.699

3.4 Molecular Feature Contributions to Drug Response

In drug-specific models, gene expression was found to be the most important predictor, with CNA and methylation following, with their effects being context-dependent.

Table 3.4: Significant Molecular Feature Effects in Selected Drugs

Drug Name	Significant Feature	Coefficient (β)	p-value	Effect
LY2109761	Gene Expression	+1.06	<0.001	Resistance
Methotrexate	Gene Expression	-1.74	0.001	Sensitivity
Methotrexate	Methylation	-1.14	0.013	Sensitivity
AZ960	CNA	-1.89	0.028	Sensitivity
RO-3306	CNA	-1.62	0.043	Sensitivity

3.5 Visualization of Pharmacogenomic Effects

Boxplots of the best drug (LY2109761) were used to do visual analysis to validate statistical results. There was clear separation in the LN50 distributions of IC50 between molecular subgroups.

Table 3.5: Visual Interpretation Summary

Molecular Feature	Observed Pattern	Interpretation
Gene Expression	Higher LN_IC50 in group 1	Reduced sensitivity
CNA	Lower LN_IC50 in group 1	Increased sensitivity
Methylation	Moderate shift	Weak/moderate effect

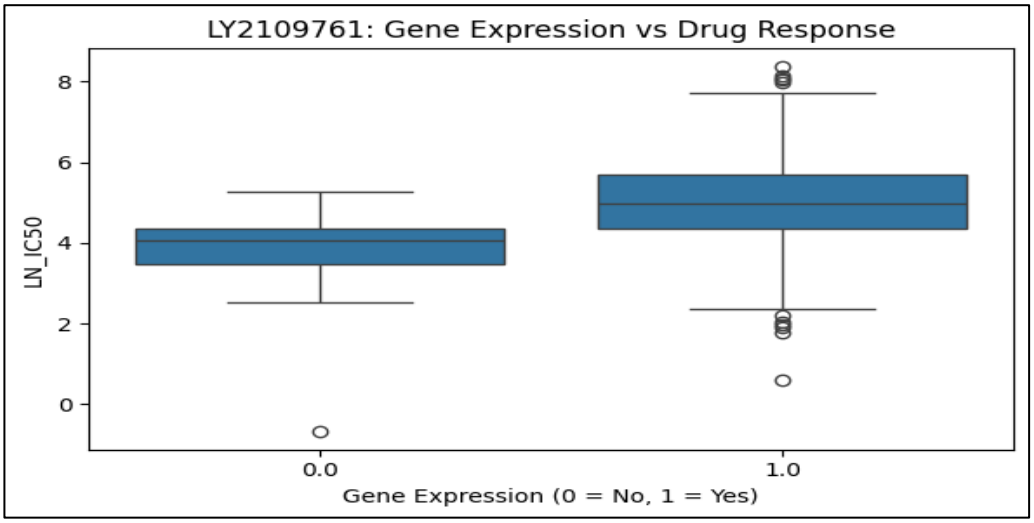

Figure 1: Boxplot showing the relationship between gene expression status and LN_IC50 for LY2109761. Higher gene expression is associated with increased drug resistance

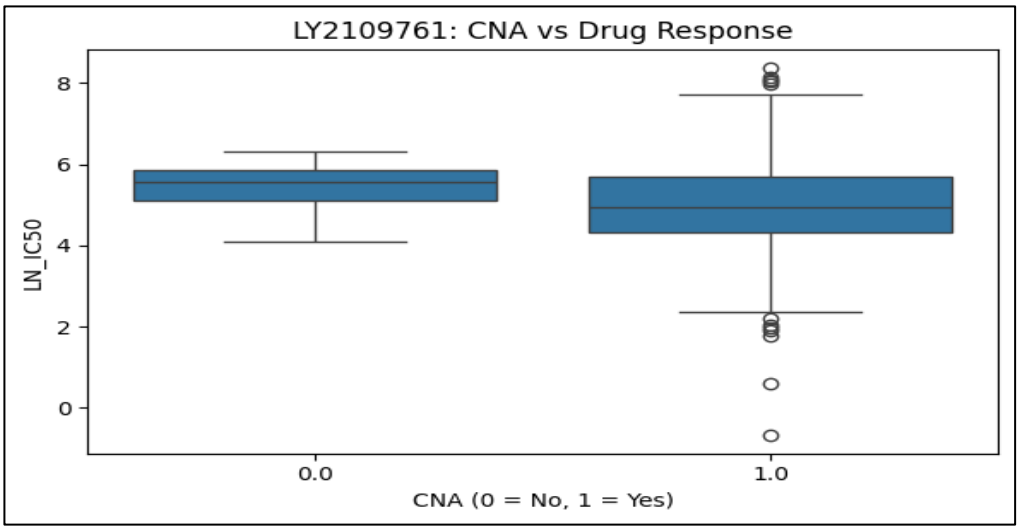

Figure 2: Boxplot illustrating the effect of copy number alterations (CNA) on LN_IC50 for LY2109761, indicating increased sensitivity in CNA-positive cell lines

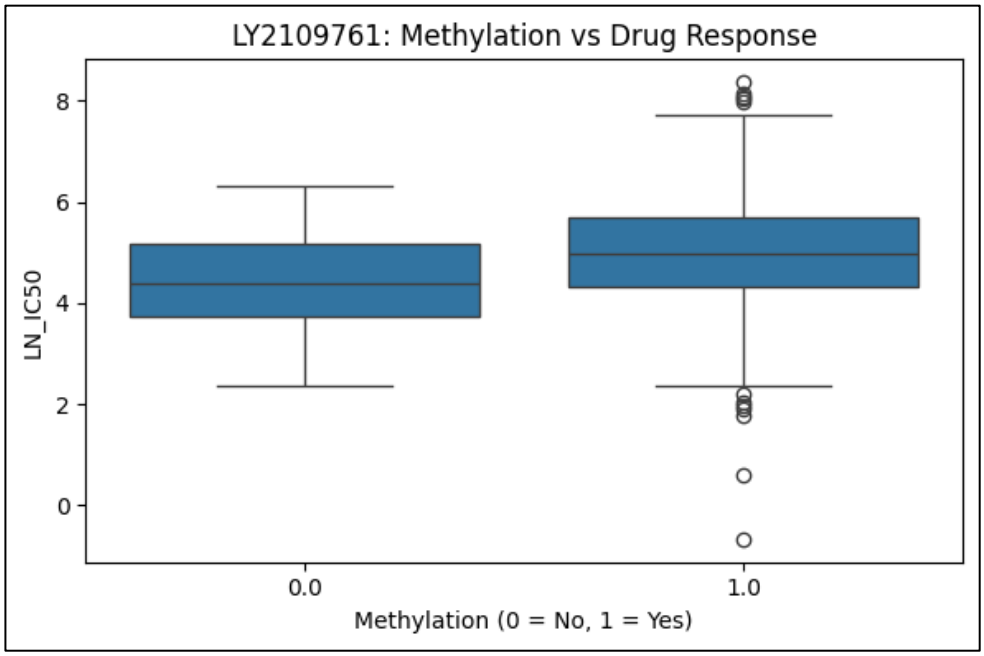

Figure 3: Boxplot depicting the association between DNA methylation and LN_IC50 for LY2109761, showing a moderate influence on drug response

4. DISCUSSION

The current research evaluated the molecular determinants of anticancer drug response with the help of a large-scale pharmacogenomic dataset (GDSC) that includes multi-omics aspects such as gene expression, copy number alterations (CNA), and DNA methylation. The results show that pharmacogenomic effects on the response to drugs are very context-dependent and drug-specific, with the gene expression coming out as the most consistent and dominant predictor of therapeutic variation. These findings are in line with the emerging consensus that single-layer genomic analyses are too simplistic to capture the complexity of drug response, and that integrative multi-omics approaches are needed to advance precision oncology (Feng et al., 2021; Nguyen et al., 2021).

At the international level, the regression model found statistically significant relationships between the molecular features and the drug response, but the explanatory power was low ($R^2 = 0.001$). This result raises a significant methodological issue: by combining heterogeneous drugs and cancer types, there is a risk of hiding biologically relevant relationships. In large-scale pharmacogenomic studies, similar observations have been reported where global models are often not able to capture subtle molecular effects due to underlying heterogeneity (Pich et al., 2022). As a result, the shift in the model to drug-specific models in this study made a significant contribution to the interpretability and the identification of stronger associations, which supports the need to use stratified analytical models in pharmacogenomics.

Drug-wise analyses revealed 95 compounds that had significant associations on their model with the model performance improving (R^2 up to -0.036). It is important to note that gene expression showed strong and significant correlations with a variety of drugs, and in the most successful models, e.g. LY2109761 and XAV939. This implies that the transcriptional activity is vital in regulating the drug sensitivity and drug resistance, probably by acting on the drug targets and the drug resistance signalling pathways. Such findings can be compared to prior studies that indicated that the profile of gene expression is a strong predictor of a drug response and can perform predictive modeling as well as the genomic alteration alone (Nguyen et al., 2021; Pak et al., 2023).

CNA showed inconsistent yet biologically relevant effects with significant associations observed in the selected drugs like AZ960 and RO-3306. The negative correlation coefficients of CNA in these models are indicative of a potential role in changing the drug sensitivity, possibly, due to gene dose effects or as a result of the interference with oncogenic pathways. This is consistent with previous findings that structural genomic changes can be key determinants affecting therapeutic outcomes by changing the expression or function of key drug targets (Schlueter and Schoenhuth, 2025). Nonetheless, the fact that CNA effects are less consistent across drugs only highlights its context-specific nature and the necessity to interpret CNA effects in a drug-target-specific manner.

DNA methylation showed a more modest and selective impact on drug response with substantial effects on drugs like Methotrexate and Tamoxifen. Methylation is an epigenetic process, which does not alter the DNA sequence, but regulates the expression of genes, thus, contributing to the phenotypic variation in drug response. The reported effects are in line with the studies demonstrating that epigenetic alterations can mediate drug sensitivity via the modulation of transcriptional programs and cellular pathways (LaBarre et al., 2019; Yildiz, 2018). Nevertheless, the relatively small effect size in comparison to gene expression may indicate that methylation may be a secondary or modulating factor and not a primary one.

The multi-omics data used in this study gives a more detailed insight into the pharmacogenomic variability. The prevailing idea that drug response is mediated by complex molecular interactions as opposed to individual biomarker is supported by the dominance of the expression of the genes coupled with the complementary roles of CNA and methylation. This is in line with the recent advances in the field of computational pharmacogenomics where multi-omics integration and feature selection were shown to significantly improve the predictive accuracy (Nguyen et al., 2021; Nguyen et al., 2021; Schlüter and Schoenhuth, 2025). In addition, it has been found that network-based and integrative strategies have an increased ability to identify clinically relevant biomarkers to use in personalised therapy (Pak et al., 2023).

Translational: These findings have significant implications to personalised medicine. The discovery of drug-specific molecular determinants has the potential to inform biomarker-based treatment plans and provide a chance to more precisely stratify patients and optimise treatment outcomes. The relatively low R^2 values seen even in drug specific models is however indicative of the inherent complexity of drug response and suggests that other factors even include tumor microenvironment, protein-level interactions, and clinical variables, may also play critical roles. This drawback is in line with larger problems with precision oncology where converting molecular understanding into clinical practice continues to be a major obstacle (Pich et al., 2022).

There are a number of constraints of this study that must be noted. First, although cancer cell lines are used, which provides a controlled experimental system, this may not in its entirety recapitulate the complexity of in vivo tumor biology. Two, binary encoding of molecular features (Y/N) can simplify underlying biological variability, which can decrease the sensitivity of the analysis. Third, there are no SNP-level pharmacogenomic data, which restrict the extent of genetic interpretation. Further research must include more detailed molecular data such as whole-genome sequencing and proteomics and clinical validation with patient-derived datasets.

With these shortcomings, this research paper offers strong evidence to validate the use of multi-omics pharmacogenomics to predict drug response. The findings add to the existing literature that has been supporting the idea of integrative and personalised oncology. Future studies need to aim to combine multi-omics data with more advanced machine learning models and clinical data to achieve a greater predictive performance and to be able to translate the pharmacogenomic insights into clinical practice.

5. CONCLUSION

The paper presents a multi-faceted pharmacogenomic analysis of drug response using large-scale GDSC data and proves that molecular determinants of drug efficacy are highly context-dependent and can be best understood based on drug-specific modeling. Although the global analysis demonstrated statistically significant, yet weak, relationships, stratified drug-wise models revealed meaningful relationships, with gene expression emerging to be the most consistent and dominant predictor of drug resistance. Context-specific responses were observed with copy number alterations and DNA methylation, which leads to drug sensitivity or resistance depending on the compound. The findings suggest that it is highly important to integrate multi-omics data in order to capture the complexity of pharmacogenomic interactions, and not to rely on single-layer genomic features. The paper also highlights the shortcomings of aggregated studies and justifies the use of more precision-oriented methodologies in the study of pharmacogenomics. In the context of translational research, the discovery of drug-specific molecular biomarkers has a profound implication in the field of personalized oncology, especially in informing the process of therapeutic selection and enhancing the outcome of treatment. Nonetheless, the small predictive value found suggests that more biological layers and clinical variables are needed to fully account the variability in drug responses. The overall impression of this work is that the integration of multi-omics and drug-specific modeling are valuable tools in advancing precision medicine and optimization of anticancer therapy.

REFERENCES

1. Ahangari, N., Doosti, M., Ghayour Mobarhan, M., Sahebkar, A., Ferns, G. A., & Pasdar, A. (2020). Personalised medicine in hypercholesterolaemia: the role of pharmacogenetics in statin therapy. *Annals of medicine*, *52*(8), 462-470.
2. Asghar, S., Virk, M. K. S., Taskeen, B., Fatima, M., & Wandala, A. (2025). The role of pharmacogenomics in Advancing Personalised Medicine: Implications for Drug Efficacy Safety and clinical practice Integration. *Indus Journal of Bioscience Research*, *3*(5), 637-645.
3. Bass, C., Hayward, A., Troczka, B. J., Haas, J., & Nauen, R. (2024). The molecular determinants of pesticide sensitivity in bee pollinators. *Science of the Total Environment*, *915*, 170174.
4. Carr, D. F., Turner, R. M., & Pirmohamed, M. (2021). Pharmacogenomics of anticancer drugs: Personalising the choice and dose to manage drug response. *British Journal of Clinical Pharmacology*, *87*(2), 237-255.
5. Feng, F., Shen, B., Mou, X., Li, Y., & Li, H. (2021). Large-scale pharmacogenomic studies and drug response prediction for personalized cancer medicine. *Journal of Genetics and Genomics*, *48*(7), 540-551.
6. LaBarre, B. A., Goncearenco, A., Petrykowska, H. M., Jaratlerdsiri, W., Bornman, M. R., Hayes, V. M., & Elnitski, L. (2019). MethylToSNP: identifying SNPs in Illumina DNA methylation array data. *Epigenetics & Chromatin*, *12*(1), 79.
7. Laes, J. F., Aftimos, P., Barthelemy, P., Bellmunt, J., Berchem, G., Camps, C., ... & Sadik, H. (2018). The clinical impact of using complex molecular profiling strategies in routine oncology practice. *Oncotarget*, *9*(29), 20282.
8. Lee, J. K., Liu, Z., Sa, J. K., Shin, S., Wang, J., Bordyuh, M., ... & Nam, D. H. (2018). Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. *Nature genetics*, *50*(10), 1399-1411.
9. Luo, F. (2019). A systematic evaluation of copy number alterations detection methods on real SNP array and deep sequencing data. *BMC bioinformatics*, *20*(Suppl 25), 692.
10. Merrill, N. M., Lachacz, E. J., Vandecan, N. M., Ulintz, P. J., Bao, L., Lloyd, J. P., ... & Soellner, M. B. (2020). Molecular determinants of drug response in TNBC cell lines. *Breast cancer research and treatment*, *179*(2), 337-347.
11. Nguyen, G. T., Vu, H. D., & Le, D. H. (2021). Integrating molecular graph data of drugs and multiple-omic data of cell lines for drug response prediction. *IEEE/ACM transactions on computational biology and bioinformatics*, *19*(2), 710-717.
12. Nguyen, L. C., Naulaerts, S., Bruna, A., Ghislat, G., & Ballester, P. J. (2021). Predicting cancer drug response in vivo by learning an optimal feature selection of tumour molecular profiles. *Biomedicines*, *9*(10), 1319.
13. Noshay, J. M., & Springer, N. M. (2021). Stories that can't be told by SNPs; DNA methylation variation in plant populations. *Current Opinion in Plant Biology*, *61*, 101989.
14. Pak, M., Lee, S., Sung, I., Koo, B., & Kim, S. (2023). Improved drug response prediction by drug target data integration via network-based profiling. *Briefings in Bioinformatics*, *24*(2), bbad034.
15. Pich, O., Bailey, C., Watkins, T. B., Zaccaria, S., Jamal-Hanjani, M., & Swanton, C. (2022). The translational challenges of precision oncology. *Cancer Cell*, *40*(5), 458-478.
16. Saeed, M. E., Mahmoud, N., Sugimoto, Y., Efferth, T., & Abdel-Aziz, H. (2018). Molecular determinants of sensitivity or resistance of cancer cells toward sanguinarine. *Frontiers in Pharmacology*, *9*, 136.
17. Salvatore, J. E., Han, S., Farris, S. P., Mignogna, K. M., Miles, M. F., & Agrawal, A. (2019). Beyond genome-wide significance: integrative approaches to the interpretation and extension of GWAS findings for alcohol use disorder. *Addiction Biology*, *24*(2), 275-289.
18. Schlüter, J., & Schönhuth, A. (2025). Genetic features for drug responses in cancer—Investigating an ensemble-feature-selection approach. *Computers in Biology and Medicine*, *196*, 110572.
19. White, C., Scott, R., Paul, C. L., & Ackland, S. P. (2022). Pharmacogenomics in the era of personalised medicine. *The Medical Journal of Australia*, *217*(10), 510.
20. Yildiz, G. (2018). Integrated multi-omics data analysis identifying novel drug sensitivity-associated molecular targets of hepatocellular carcinoma cells. *Oncology Letters*, *16*(1), 113-122.