Computational Molecular Biology 2025, Vol.15, No.3, 131-140
http://bioscipublisher.com/index.php/cmb

Research Insight Open Access

Genomic Biomarker Discovery for Drug Sensitivity Using Omics Data

Jiayi Wu, Keyan Fang
Traditional Chinese Medicine Research Center, Cuixi Academy of Biotechnology, Zhuji, 311800, China
Corresponding author: keyan.fang@cuixi.org

Computational Molecular Biology, 2025, Vol.15, No.3 doi: 10.5376/cmb.2025.15.0013
Received: 18 Mar., 2025
Accepted: 29 Apr., 2025
Published: 19 May, 2025

Preferred citation for this article:
Wu J.Y., and Fang K.Y., 2025, Genomic biomarker discovery for drug sensitivity using omics data, Computational Molecular Biology, 15(3): 131-140 (doi: 10.5376/cmb.2025.15.0013)

Abstract Drug sensitivity refers to the differences in the degree of response of different individuals or cells to drugs. Revealing its molecular mechanism is crucial for achieving individualized and precise treatment. However, the average efficacy rate of the anti-cancer drugs approved by the FDA among patients is only about 40%, indicating that the traditional "one-size-fits-all" treatment model is difficult to meet the diverse needs of patients. The development of omics technology has made it possible to conduct a global analysis of biomarkers related to drug responses. By integrating multi-level data such as genomics, transcriptomics, and proteomics, genomic markers closely related to drug sensitivity can be systematically screened out, thereby predicting patients' responses to specific drugs and guiding clinical medication. This study starts from the basic concepts and molecular mechanisms of drug sensitivity, reviews the application of omics data in drug response research, methods and algorithms for genomic marker screening, as well as common data resources, and conducts a case analysis of multi-omics marker screening taking the anti-cancer drug EGFR inhibitor as an example, discussing the current challenges and limitations. Finally, the development direction of precise drug response prediction driven by artificial intelligence is prospected. This study aims to provide a reference for mining drug sensitivity biomarkers using omics data, promoting precision medicine and new drug development.

Keywords Drug sensitivity; Genomic markers; Omics integration; Individualized treatment; Bioinformatics

1 Introduction

The same anti-cancer drug, when applied to different people, often yields vastly different results. Some people show remarkable therapeutic effects, while others have almost no reaction. The statistics are quite striking - even for drugs approved by the FDA, the average effective rate is only around 40%, which means that more than half of the patients may have suffered in vain. It's no wonder. Tumors, as a disease, are inherently complex, involving genetic differences, tissue heterogeneity, metabolic status... Either of them can affect the efficacy of the medicine (Creighton et al., 2013). Drug sensitivity, in essence, refers to whether a drug can exert its expected effect on a specific person or cell. The problem is that such differences cannot be discernment based on experience; answers have to be sought at the molecular level (Shaffer, 2022). If the key molecules that determine the drug response can really be identified, doctors can tailor the prescription to the patient's condition before prescribing drugs, which not only improves the therapeutic effect but also reduces the patient's suffering. This is precisely the core issue that precision medicine aims to address.

The development speed of omics technology in recent years has almost kept people at a loss. In the past, when studying drug responses, it was often done by checking each gene one by one. Nowadays, it is possible to observe the changes of tens of thousands of genes at once, and even the expression patterns can be analyzed simultaneously. Once the field of vision is broadened, there will naturally be more things to discover. Many new markers have been unearthed under this kind of "panoramic scanning" (Guang et al., 2018). The more data there is, the more complex the problem becomes. However, this has instead given rise to various algorithms - machine learning and network analysis, all of which have come into play to dig out connections and find patterns from the vast amount of data (Jung et al., 2021). Genome, transcriptome, proteome, metabolome... When these levels of information are put together, we can see how drugs truly work and even identify the root causes behind drug resistance (Sun et al., 2021). It can be said that current drug sensitivity research is almost inseparable from the

support of omics technology. It is like a searchlight, helping researchers find those key points in the vast amount of molecular information.

This article mainly aims to discuss how to use multi-omics data to identify genomic markers of drug sensitivity, and also incidentally sort out the common research ideas and progress at present. Overall, it is divided into several parts. First, let's clarify what drug sensitivity is and what the molecular mechanisms behind it are roughly like. Then, let's talk about the role of genomic markers in predicting drug responses and how pharmacogenomics has developed in recent years. Next, we will turn to the transcriptome, proteome, and metabolome levels to see what value they can provide when studying drug responses and how multi-omics data can be integrated and analyzed. The last part does not offer a summary but rather a outlook, discussing the potential of artificial intelligence in predicting drug responses in the future, how omics and clinical data can be better combined, and the inspirations these studies can bring to precision medicine and new drug development.

2 Theoretical Basis of Drug Sensitivity and Genomic Markers

2.1 Definition and molecular mechanism overview of drug sensitivity

Drug sensitivity, in essence, is about how strongly an organism or cell reacts to a certain drug concentration. Some people calculate it using the half-maximal inhibitory concentration (IC50), while others look at the area of the dose-response curve (AUC). Clinically, it is more intuitive, such as how much the tumor has shrunk and how long it has been since the patient relapsed. There are too many links that can affect it, from how the drug enters the cell to whether it can smoothly act on the target site after entering, all of which may go wrong. For instance, if the transporter or metabolic enzyme is inefficient, the drug cannot get in. Once the target gene mutates, the efficacy of the drug may be immediately reduced. What's even more troublesome is that tumor cells can "start anew", such as opening up backup signaling pathways or repairing damaged DNA more quickly. Some simply pump the drug out (like the exudation protein P-gp), or simply change their "personality" and undergo epithelial-mesenchymal transition (EMT), becoming insensitive to several drugs. Overall, drug sensitivity is the result of the combined action of multiple genes and pathways. To fully understand it, it is necessary to rely on the integrated analysis of multiple omics.

2.2 The role of genomic markers in drug response prediction

Some patients respond particularly well to medication, while others show almost no reaction. Such differences often need to be answered in genes. Genomic markers, in essence, are genetic characteristics that can predict a person's response to a certain drug, such as gene mutations, copy number changes, different expression levels, or single nucleotide polymorphisms. If doctors know this information before treatment, they can roughly determine who is more likely to be sensitive to the drug and who may be resistant to it, thus avoiding detours (Wang et al., 2008). There are many such cases in clinical practice - for instance, in patients with non-small cell lung cancer, those with EGFR-sensitive mutations usually respond well to tyrosine kinase inhibitors (gefitinib, erlotinib). However, if KRAS mutations occur in the tumor, such drugs are basically of no use (Martin et al., 2013). Genetic mutations not only help doctors select the right drugs but also make the development of new drugs more precise. Nowadays, many clinical trials screen patients in advance, selecting those who are more likely to respond (also known as the enrichment strategy), which leads to a higher success rate. Overall, identifying and verifying these genomic markers is the most crucial step for the implementation of precision medicine.

2.3 Development and application of pharmacogenomics

The discipline of pharmacogenomics was actually originally intended to figure out why the same drug has such different effects on different people. Later, research became increasingly detailed, evolving from initially focusing on individual genes to later being able to search for clues throughout the entire genome. Over the past two decades or so, scientists have discovered many genetic markers related to drug efficacy or side effects, and some of them have truly been incorporated into the hospital medication process. For instance, when it comes to the antithrombotic drug clopidogrel, before prescriming the medication, doctors will check the patient's CYP2C19 genotype. For those with weak metabolic capacity, they either adjust the dosage or change the medication (Figure 1) (Angulo-Aguado et al., 2021). There are more and more such examples, which also provide doctors with

clearer references and no longer rely solely on experience. Nowadays, many clinical guidelines include genetic testing, which can be regarded as an important step in individualized medication (Lee et al., 2022). Of course, reality is not so smooth either. The genetic differences among different groups of people, the unified standards for detection technologies, and the acceptance of genetic testing by hospitals are all still under exploration. Despite this, the emergence of pharmacogenomics has indeed transformed precision medicine from a concept into an operational reality.

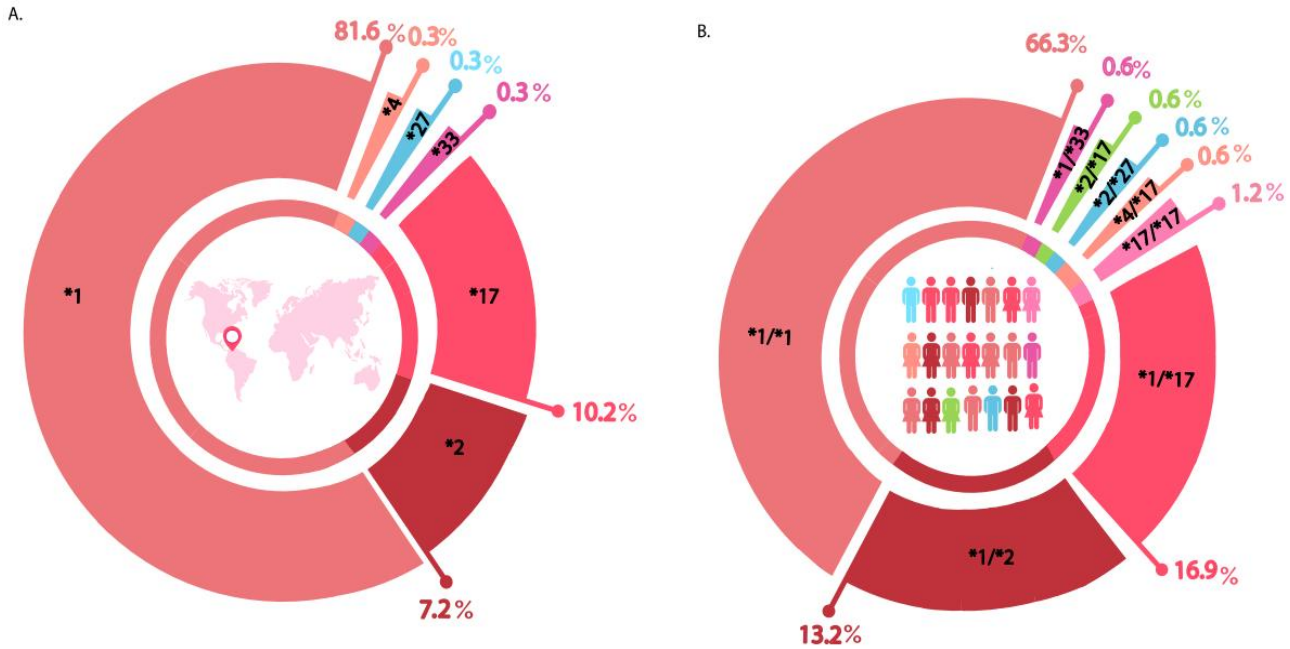

Figure 1 Allele and genotype frequencies. Data for allele and genotype frequencies of CYP2C19 are illustrated. (A) Allele frequency and (B) genotype frequency (Adopted from Angulo-Aguado et al., 2021)

3 Application of Omics Data in Drug Sensitivity Studies

3.1 Transcriptomics and analysis of drug response gene expression profiles

When studying drug sensitivity, people often start with the transcriptome. Because changes in gene expression often reveal the attitude of cells towards drugs. The difference is that some cell lines "surrender obediently" upon encountering drugs, while others are almost unaffected. Researchers will compare the gene expression profiles of these two types of samples to see which ones are up-regulated and which ones are down-regulated. For instance, the drug-resistant group often activates the efflux pump or anti-apoptotic pathways, while the sensitive group instead activates the genes related to the drug target at a higher level. Such differentially expressed genes were later often used as potential markers to predict whether new samples would respond to drugs (Talwar and Carter, 2020). In addition to identifying differences, transcriptome data can also be used for model training - by combining the expression data of different cell lines with IC50 values and using regression or classification algorithms to calculate which genes best reflect drug efficacy (Mannheimer et al., 2016). Although it sounds complicated, the core idea is actually very simple: let the gene expression map tell us the story of the drug. These analysis results can not only help explain the mechanism but also provide clues for subsequent experiments and clinical decisions.

3.2 The role of proteomics and metabolomics in the study of drug action mechanisms

When studying the mechanism of drugs, merely looking at genetic and transcriptional information is far from enough. What really works is often at the protein and metabolic levels. Proteomics can directly observe which proteins in cells are mobilized and which are modified, and the changes in pathways under different drug treatments are clear at a glance. Sometimes, drug-resistant cells quietly activate key proteins in backup signaling pathways or turn on the repair system more aggressively to evade drug attacks - all of which can be found as clues in the proteome (Figure 2) (Fortuin and Soares, 2022). In contrast, the transcriptome is merely a "plan", while the

proteome is more like a "construction site", where one can actually see what is happening. Looking at metabolomics, it focuses on the changes of various small molecule products within cells, reflecting the energy and material flow of cells. Once the sugar, amino acid and fat metabolism of tumor cells deviates, it often indicates the signs of drug resistance. By comparing the metabolic profiles of sensitive and drug-resistant samples, researchers can identify the abnormally activated pathways (Shajahan-Haq et al., 2015). Perhaps by intervening in these pathways, the drug efficacy can be "salvaged". By combining the proteome and metabolome, the full picture of drug effects can be depicted at the functional level.

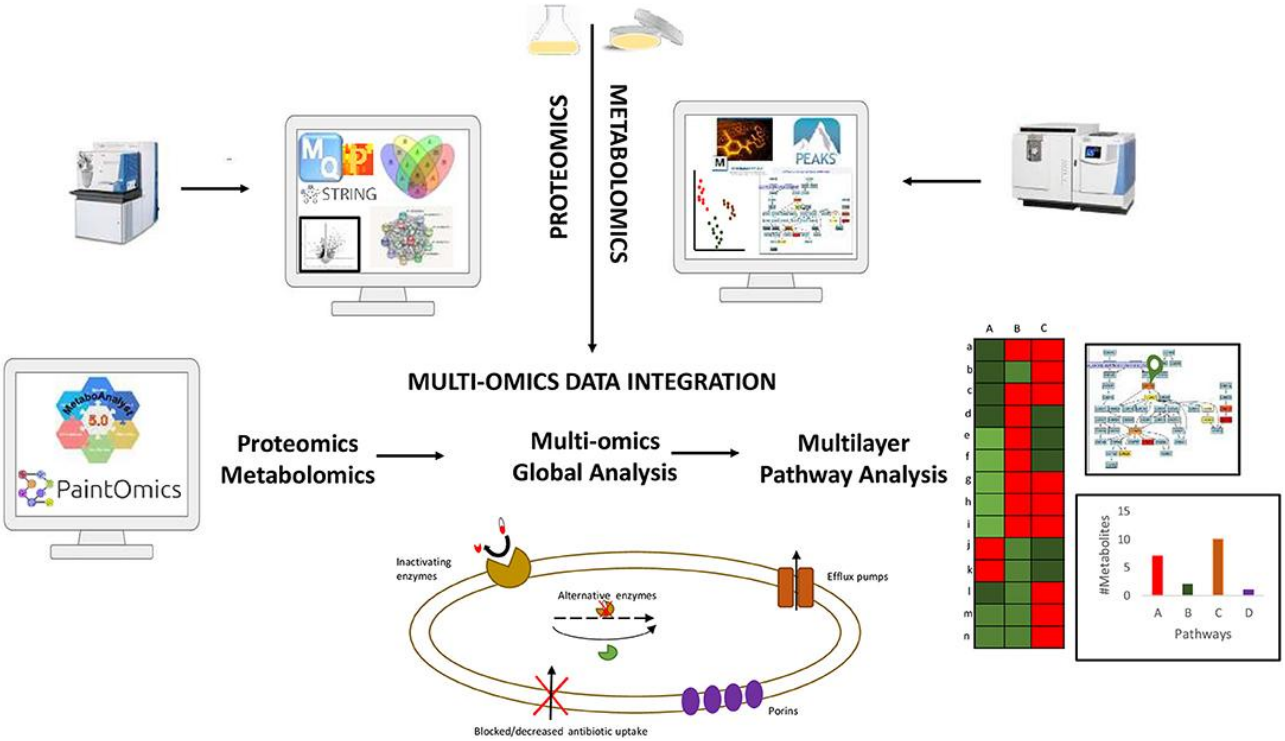

Figure 2 Summary of data integration workflow combining proteomics and metabolomics data for a comprehensive understanding of the biochemical alterations of pathogenic drug resistant bacteria (Adopted from Fortuin and Soares, 2022)

3.3 Multi-omics data integration strategy and bioinformatics analysis framework

When it comes to studying drug sensitivity, relying on a single type of data is often insufficient. There are various theories at the levels of genes, transcription, proteins, and metabolism, but when taken together, the picture becomes complete. There are several approaches to multi-omics integration. Some people prefer to "get started early", standardizing different data first and then mixing them into a high-dimensional matrix and throwing it all into the model at once (Liu and Mei, 2023). Some people are more cautious. They analyze each one separately first and then piece together the results at the end. No matter which path it is, it cannot do without the support of a complete set of bioinformatics tools. From data cleaning to feature screening, and then to modeling and verification, every step must be meticulous. Old methods like principal component analysis, co-clustering, and network analysis are still effective and can uncover commonalities and complementary information among different omics. There is a more systematic approach - mapping multi-omics data onto the same molecular network, allowing genes, proteins, and metabolites to "match" in pathways. This makes it easier to identify functional modules related to drug responses (Oh et al., 2020). Finally, pathway enrichment is used for verification, which also makes it convenient to explain exactly what these markers mean.

4 Methods and Algorithms for Screening Genomic Markers

4.1 Single-omics feature selection and statistical analysis methods

When studying drug sensitivity, many people initially start with single-omics data for the sake of intuitiveness. A common practice is to first separate the drug sensitivity and resistance samples into two groups to see which molecular features are significantly different on both sides - the old methods such as t-test and chi-square test are

still effective. Sometimes, the correlation between gene expression levels or mutation frequencies and the IC50 of drugs is also calculated to identify possible markers. In terms of models, linear regression is the most commonly used one. Although it is simple, the results are clear and easy to explain. To prevent the model from being distorted due to "learning too much", researchers usually add regularization penalties to make the results more stable. Overall, although these basic methods are simple, they are important starting points for screening markers (Nguyen et al., 2016).

4.2 Multi-omics integrated analysis based on machine learning and artificial intelligence

In recent years, the advancement of algorithms has made machine learning and artificial intelligence almost standard in multi-omics analysis. Researchers are no longer content with the set of linear relationships; instead, they are more eager to capture those complex and elusive nonlinear patterns. Deep learning has its advantages. Data from different omics can be processed separately and then "converged" at the middle layer to extract more comprehensive features (Tan et al., 2020). Some people prefer ensemble learning, which combines models trained for different omics, and the results are often more stable (Yang et al., 2022). One study did it this way: by integrating gene expression, mutation and copy number data into a Stacking model, the accuracy of predicting drug sensitivity directly increased. Of course, such methods are not omnipotent. However, on the whole, multi-omics models are indeed more reliable than single-omics models and have greater reference value for clinical decision-making.

4.3 Application of biological networks and systems biology methods in marker recognition

When it comes to studying drug responses, looking at just a few genes is often insufficient. Systems biology places more emphasis on the "sense of the whole". It attempts to weave the relationships among molecules such as genes, proteins, and drugs into a web, and identify the truly crucial nodes or modules from the network structure. Some people use protein-protein interaction networks, while others conduct gene co-expression analysis (Zogopoulos et al., 2022). Their approaches are different, but their goals are similar - to identify those factors that are at the "hub" position in the network and may dominate drug responses. Sometimes, the results also need to be combined with pathway enrichment analysis, placing candidate genes in known biological pathways for comparison. This can make their functional roles clearer and also make the model's explanation more intuitive (Zhang et al., 2018). Overall, this type of network method enables people to understand the underlying logic of drug effects from a more systematic perspective.

5 Data Sources and Database Resources

5.1 Comparison and application of drug sensitivity datasets (such as CCLE, GDSC, NCI-60)

Most people studying drug sensitivity cannot avoid several commonly used databases. The most famous one is probably CCLE, which collected approximately a thousand cancer cell lines. It not only contains genomic information but also multi-omics and drug response data. The GDSC developed by the Sanger Institute in the UK is also frequently mentioned. Currently, it has collected nearly 700 cell lines and tested their responses to over 100 types of anti-cancer drugs. It is regarded as one of the largest drug sensitivity databases. Its focus lies in analyzing drug efficacy data together with information such as gene mutations and copy number changes, from which molecular clues that may affect drug responses are unearthed (Reinhold et al., 2015). A little further back, NCI-60 can be regarded as a "senior". Although it only contains sixty cell lines, it covers over a thousand compounds and remains an important reference for the study of traditional chemotherapy drugs to this day (Takamatsu and Matsumura, 2023).

5.2 Integration and utilization of public omics databases (such as TCGA, GEO)

Most people who conduct drug response research cannot do without those large cancer databases. TCGA is a representative among them. It not only contains the genomic and transcriptomic data of the patient's tumor, but also the corresponding clinical information, which can be used to analyze which molecular features may affect the therapeutic effect. Some researchers will match these patient data with the drug sensitivity information of cell lines to infer which drugs the patients might be more sensitive to. For instance, the R package pRRophetic combines the gene expression profile of GDSC with the tumor data of TCGA, and uses the ridge regression model

to predict the IC50 value of the drug, helping doctors make a rough judgment before treatment (Clayton et al., 2020). Of course, TCGA is not the only source. Public databases like GEO also store a large amount of transcriptome data, which can be used not only to validate new markers but also to train predictive models, and are equally convenient to use (Wang et al., 2023).

5.3 Issues regarding data standardization and repeatability

When the data of multi-source omics are combined, problems also arise. Due to different experimental conditions and the mixture of batch effects, the analysis results often do not match. Even in different drug sensitivity databases, the test results of the same drug can vary greatly, and this situation is not uncommon. To make the data more reliable, we must first start from the source - the experimental process, data processing, and standardization. Each link should be as unified as possible. Even if the result looks very good, it still needs to be tried again in an independent queue to verify whether it can be reproduced; otherwise, it is hard to say how reliable it is (Moossavi et al., 2020). Nowadays, many teams are making efforts in this regard, unifying the analysis process and promoting data sharing, hoping to make drug sensitivity studies more stable and comparable (Sinke et al., 2021).

6 Case Study

6.1 Case selection: anti-cancer drug sensitivity analysis based on GDSC and TCGA

Here, we take EGFR-targeted therapy for lung cancer as an example. Let's start with the GDSC database, pick out the genomic markers related to EGFR-TKI drug sensitivity, and then see if these markers "make sense" in real patients. The specific approach is to take the screened candidate genes into the TCGA lung adenocarcinoma data for verification to see if there is a consistent trend between them and clinical efficacy (Huang et al., 2018; Cheng et al., 2020). In this way, not only can clues be found at the cell line level, but also the reliability of the results can be tested with patient data.

6.2 Multi-omics feature screening process and identification of key markers

After analysis, the result is quite interesting. Sensitive mutations of EGFR only occur in those cell lines that respond well to drugs, while the drug-resistant batch often carries driver mutations such as KRAS (Ohashi et al., 2012). In addition, drug-resistant cells also exhibit distinct EMT characteristics, with different forms and expression patterns. Piecing together these multi-omics clues reveals a more complete picture - different mutation backgrounds and transcriptional states seem to be jointly shaping the differences in cells' responses to EGFR-Tkis (Yamaguchi et al., 2012).

6.3 Verification and clinical correlation analysis: taking EGFR mutation and TKI response as an example

Clinical data also confirm this point. Lung cancer patients with EGFR mutations often have a much more obvious response after using EGFR-TKI than those with wild-type EGFR-TKI, and the difference in therapeutic effect is obvious at a glance (Mitsudomi et al., 2006; Li et al., 2010). For this reason, nowadays, doctors basically conduct EGFR gene testing before formulating treatment plans for patients with advanced lung adenocarcinoma. This step has become a routine operation. Without clarifying the mutation situation, it is often difficult to even choose the right medicine.

7 Challenges and Limitations

7.1 Heterogeneity and high-dimensional feature issues of multi-omics data

Multi-omics data may seem rich in information, but there are also many problems. The noise brought by different platforms and batches often overshadows the truly meaningful signals, resulting in a chaotic analysis (Liu and Park, 2024). Moreover, when there are many features and few samples, if the model is not careful, it will "learn off course" and overfit may occur (Hu et al., 2022). At such times, the data must be sorted out first - standardization, feature screening, and dimensionality reduction. None of these steps can be omitted. Only by suppressing the redundancy and noise can the subsequent models remain stable.

7.2 Obstacles to model interpretability and clinical translability

No matter how accurate many models are, doctors are still reluctant to believe them directly. The main reason is that it's too complicated to understand and explain clearly. Nowadays, some research is seeking ways to make

models "speak human language", such as using simpler algorithms or adding interpretation tools like SHAP, allowing people to see exactly how much effect each feature plays (Zeng et al., 2024). Even so, to truly enter clinical practice, several hurdles still need to be overcome - the model must be verified in clinical trials, and the detection of markers cannot be too complicated (Baek et al., 2024). Otherwise, even the smartest model may only remain at the research stage.

7.3 Data sharing, privacy protection and ethical issues

The genomic and clinical data of patients are indeed a "gold mine" for scientific research, but problems also arise - privacy, ethics, and regulation - none of which can be avoided. Regulations in all countries are very strict, and it is almost impossible to freely use these data. Some people have proposed controlled sharing, while others have attempted federated learning, hoping to strike a balance between "opening" and "preventing" (Calvino et al., 2024; Chandrashekar et al., 2024). What researchers can do is to advance their work as much as possible under the premise of adhering to the ethical bottom line, making the data valuable while also ensuring that patients' rights and interests are not infringed upon.

8 Future Outlook and Conclusions

Looking ahead, the weight of artificial intelligence in drug response prediction will only become increasingly significant. The more data there is and the stronger the computing power, the more detailed and accurate the model will be trained. However, nowadays people no longer merely pursue "accuracy", but also want models to explain "why" clearly. Thus, the integration of interpretability and biological knowledge has become a new direction. On the other hand, multimodal fusion is also on the rise - not only looking at genetic or transcriptional information, but also taking into account the patient's clinical indicators and imaging data to piece together a more complete "digital patient". This integration approach might enable AI to truly understand individual differences, thereby being closer to reality when predicting drug responses. When these technologies mature, perhaps precise medication will no longer be the goal but will become a habit.

In the future medical scene, it is likely that it will not be as simple as just looking at medical records and images. The patient information in the hands of doctors may also carry a complete set of genomic sequencing results, transcriptional profiling data, and even real-time monitored molecular changes. As hospitals accumulate more and more such comprehensive data, the boundary between clinical practice and omics will become blurred. Imagine a system that can simultaneously understand molecular-level features and, in combination with the patient's clinical condition, judge drug responses. This "clinical-omics integration" decision-making approach is much more precise than the past empirical speculation. The emergence of liquid biopsy has also made dynamic adjustment treatment a reality. In the future, multi-center collaboration and data sharing may make these systems smarter and smarter as they are used. At that time, doctors did not rely solely on intuition but had a complete set of intelligent tools behind them to support them.

The most fundamental goal of precision medicine is to ensure that the right medicine is used for the right people. Biomarker screening serves precisely this purpose - it can inform doctors in advance who is more likely to benefit and who is at greater risk. EGFR mutations in lung cancer are a typical example, transforming targeted therapy from an attempt into a precise strike. For pharmaceutical companies, such markers are equally crucial. New drug trials no longer involve blind submissions but first identify potentially sensitive patient groups to increase success rates and shorten approval cycles. The drug resistance mechanisms revealed by multi-omics studies are also constantly providing new directions for new drugs. For instance, osimertinib was developed to address the drug resistance of EGFR T790M mutations. In the future, biomarkers will no longer be regarded as "add-ons" in new drug development, but will be incorporated into the design from the very beginning. When drugs and biomarkers appear in pairs, the development efficiency will naturally be higher.

Acknowledgments

We would like to express my heartfelt thanks to all the teachers who have provided guidance for this study.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Angulo-Aguado M., Panche K., Tamayo-Agudelo C., Ruiz-Torres D., Sambracos-Parrado S., Niño-Orrego M., Páez N., Piñeros-Hernandez L., Castillo-León L., Pardo-Oviedo J., Abaunza K., Laissue P., Contreras N., Calderón-Ospina C., and Fonseca-Mendoza D., 2021, A pharmacogenetic study of *CYP2C19* in acute coronary syndrome patients of Colombian origin reveals new polymorphisms potentially related to clopidogrel therapy, Journal of Personalized Medicine, 11(5): 400.
https://doi.org/10.3390/jpm11050400

Baek S., Jeong Y., Kim Y., Kim J., Kim J., Kim E., Lim J., Kim J., Kim Z., Kim K., and Chung M., 2024, Development and validation of a robust and interpretable early triaging support system for patients hospitalized with COVID-19: predictive algorithm modeling and interpretation study, Journal of Medical Internet Research, 26: e52134.
https://doi.org/10.2196/52134

Calvino G., Peconi C., Strafella C., Trastulli G., Megalizzi D., Andreucci S., Cascella R., Caltagirone C., Zampatti S., and Giardina E., 2024, Federated learning: breaking down barriers in global genomic research, Genes, 15(12): 1650.
https://doi.org/10.3390/genes15121650

Chandrashekar R., Dasari K., Mishra M., Nijhawan G., Kalra R., and Abood B., 2024, Distributed bioengineering data analysis preserving privacy with federated learning, In: 2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0, IEEE, pp.1-6.
https://doi.org/10.1109/OTCON60325.2024.10687990

Cheng C., Zhao Y., Schaafsma E., Weng Y., and Amos C., 2020, An EGFR signature predicts cell line and patient sensitivity to multiple tyrosine kinase inhibitors, International Journal of Cancer, 147(9): 2621-2633.
https://doi.org/10.1002/ijc.33053

Clayton E., Pujol T., McDonald J., and Qiu P., 2020, Leveraging TCGA gene expression data to build predictive models for cancer drug response, BMC Bioinformatics, 21(Suppl 14): 364.
https://doi.org/10.1186/s12859-020-03690-4

Creighton C., 2013, Widespread molecular patterns associated with drug sensitivity in breast cancer cell lines, with implications for human tumors, PLoS ONE, 8(12): e71158.
https://doi.org/10.1371/journal.pone.0071158

Fortuin S., and Soares N., 2022, The integration of proteomics and metabolomics data paving the way for a better understanding of the mechanisms underlying microbial acquired drug resistance, Frontiers in Medicine, 9: 849838.
https://doi.org/10.3389/fmed.2022.849838

Guang M., McCann A., Bianchi G., Zhang L., Dowling P., Bazou D., O'Gorman P., and Anderson K., 2018, Overcoming multiple myeloma drug resistance in the era of cancer 'omics', Leukemia & Lymphoma, 59(3): 542-561.
https://doi.org/10.1080/10428194.2017.1337115

Hu Y., Zhao L., Li Z., Dong X., Xu T., and Zhao Y., 2022, Classifying the multi-omics data of gastric cancer using a deep feature selection method, Expert Systems with Applications, 200: 116813.
https://doi.org/10.1016/j.eswa.2022.116813

Huang C., Huang Y., Shen S., Liu C., Yang C., and Kuo C., 2018, Factors predicting the efficacy of osimertinib in T790M-mutated metastatic lung adenocarcinoma, Respirology, 23(S2): 20.
https://doi.org/10.1111/resp.13419_43

Jung H., Sung Y., and Kim H., 2021, Omics and computational modeling approaches for the effective treatment of drug-resistant cancer cells, Frontiers in Genetics, 12: 742902.
https://doi.org/10.3389/fgene.2021.742902

Lee C., Luzum J., Sangkuhl K., Gammal R., Sabatine M., Stein C., Kisor D., Limdi N., Lee Y., Scott S., Hulot J., Roden D., Gaedigk A., Caudle K., Klein T., Johnson J., and Shuldiner A., 2022, Clinical Pharmacogenetics Implementation Consortium guideline for CYP2C19 genotype and clopidogrel therapy: 2022 update, Clinical Pharmacology & Therapeutics, 112(5): 959-967.
https://doi.org/10.1002/cpt.2526

Li J., Chen Z., Sun D., Xu M., Zhang X., and Xu J., 2010, The quantitative analysis of epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer (NSCLC) and its correlation with response rate of EGFR tyrosine kinase inhibitors (TKI), Clinical Cancer Research, 16(19_Supplement): PR10.
https://doi.org/10.1158/DIAG-10-PR10

Liu X., and Mei X., 2023, Prediction of drug sensitivity based on multi-omics data using deep learning and similarity network fusion approaches, Frontiers in Bioengineering and Biotechnology, 11: 1156372.
https://doi.org/10.3389/fbioe.2023.1156372

Liu Z., and Park T., 2024, DMOIT: denoised multi-omics integration approach based on transformer multi-head self-attention mechanism, Frontiers in Genetics, 15: 1488683.
https://doi.org/10.3389/fgene.2024.1488683

Mannheimer J., Fowles J., Shaumberg K., Duval D., Prasad A., and Gustafson D., 2016, Predicting drug sensitivity based on gene array data for cytotoxic chemotherapeutic agents, Cancer Research, 76(14_Supplement): 1522-1522.
https://doi.org/10.1158/1538-7445.AM2016-1522

Martin P., Leighl N., Tsao M., and Shepherd F., 2013, KRAS mutations as prognostic and predictive markers in non-small cell lung cancer, Journal of Thoracic Oncology, 8(5): 530-542.
https://doi.org/10.1097/JTO.0b013e318283d958

Mitsudomi T., Kosaka T., and Yatabe Y., 2006, Biological and clinical implications of EGFR mutations in lung cancer, International Journal of Clinical Oncology, 11(3): 190-198.
https://doi.org/10.1007/s10147-006-0583-4

Moossavi S., Fehr K., Moraes T., Khafipour E., and Azad M., 2020, Repeatability and reproducibility assessment in a large-scale population-based microbiota study: case study on human milk microbiota, Microbiome, 9(1): 41.
https://doi.org/10.1101/2020.04.20.052035

Nguyen L., Dang C., and Ballester P., 2016, Systematic assessment of multi-gene predictors of pan-cancer cell line sensitivity to drugs exploiting gene expression data, F1000Research, 5: ISCB Comm J-2927.
https://doi.org/10.12688/f1000research.10529.2

Oh M., Park S., Lee S., Lee D., Lim S., Jeong D., Jo K., Jung I., and Kim S., 2020, DRIM: a web-based system for investigating drug response at the molecular level by condition-specific multi-omics data integration, Frontiers in Genetics, 11: 564792.
https://doi.org/10.3389/fgene.2020.564792

Ohashi K., Sequist L., Arcila M., Morán T., Chmielecki J., Lin Y., Pan Y., Wang L., De Stanchina E., Shien K., Aoe K., Toyooka S., Kiura K., Fernandez-Cuesta L., Fidias P., Yang J., Miller V., Riely G., Kris M., Engelman J., Vnencak-Jones C., Dias-Santagata D., Ladanyi M., and Pao W., 2012, Lung cancers with acquired resistance to EGFR inhibitors occasionally harbor *BRAF* gene mutations but lack mutations in *KRAS*, *NRAS*, or *MEK1*, Proceedings of the National Academy of Sciences, 109(31): E2127-E2133.
https://doi.org/10.1073/pnas.1203530109

Reinhold W., Varma S., Rajapakse V., Luna A., Sousa F., Kohn K., and Pommier Y., 2015, Using drug response data to identify molecular effectors, and molecular omic data to identify candidate drugs in cancer, Human Genetics, 134(1): 3-11.
https://doi.org/10.1007/s00439-014-1482-9

Shaffer S., 2022, Rare cell variability and drug resistance in cancer, Cancer Research, 82(10_Supplement): IA018.
https://doi.org/10.1158/1538-7445.evodyn22-ia018

Shajahan-Haq A., Cheema M., and Clarke R., 2015, Application of metabolomics in drug resistant breast cancer research, Metabolites, 5(1): 100-118.
https://doi.org/10.3390/metabo5010100

Sinke L., Cats D., and Heijmans B., 2021, Omixer: multivariate and reproducible sample randomization to proactively counter batch effects in omics studies, Bioinformatics, 37(18): 3051-3052.
https://doi.org/10.1093/bioinformatics/btab159

Sun Z., Liu Y., Ouyang Q., Liu Z., and Liu Y., 2021, Research progress of omics technology in the field of tumor resistance: from single-omics to multi-omics combination application, Journal of Central South University (Medical Sciences), 46(6): 620-627.
https://doi.org/10.11817/j.issn.1672-7347.2021.200561

Takamatsu S., and Matsumura N., 2023, Homologous recombination deficiency in the CCLE cell lines correlates with in vitro resistance to platinum agents and PARP inhibitors, Cancer Research, 83(7_Supplement): 2773-2773.
https://doi.org/10.1158/1538-7445.AM2023-2773

Talwar J., and Carter H., 2020, Abstract 2099: Assessing cancer drug response prediction from gene expression, Clinical Research, 80(16_Supplement): 2099-2099.
https://doi.org/10.1158/1538-7445.AM2020-2099

Tan X., Yu Y., Duan K., Zhang J., Sun P., and Sun H., 2020, Current advances and limitations of deep learning in anticancer drug sensitivity prediction, Current Topics in Medicinal Chemistry, 20(21): 1858-1867.
https://doi.org/10.2174/1568026620666200710101307

Wang G., Liu X., Liu H., Zhang X., Shao Y., and Jia X., 2023, A novel necroptosis related gene signature and regulatory network for overall survival prediction in lung adenocarcinoma, Scientific Reports, 13(1): 15345.
https://doi.org/10.1038/s41598-023-41998-2

Wang Z., Wu Y., Zhang G., Zhou Q., Xu C., and Guo A., 2008, EGFR/KRAS mutations and gefitinib therapy in Chinese NSCLC patients, Oncology Research and Treatment, 31(4): 174-178.
https://doi.org/10.1159/000116736

Yamaguchi F., Kugawa S., Tateno H., Kokubu F., and Fukuchi K., 2012, Analysis of EGFR, KRAS and P53 mutations in lung cancer using cells in the curette lavage fluid obtained by bronchoscopy, Lung Cancer, 78(3): 201-206.
https://doi.org/10.1016/j.lungcan.2012.08.014

Yang C., Liu Z., Dai P., Zhang Y., Huang P., Lin Y., and Xie L., 2022, Predicting tumor drug sensitivity with multi-omics data, Chinese Journal of Biotechnology, 38(6): 2201-2212.
https://doi.org/10.13345/j.cjb.210676

Zeng X., 2024, Enhancing the interpretability of SHAP values using large language models, arXiv, 2409: 00079.
https://doi.org/10.48550/arXiv.2409.00079

Zhang S., 2018, Comparisons of gene coexpression network modules in breast cancer and ovarian cancer, BMC Systems Biology, 12(Suppl 1): 8.
https://doi.org/10.1186/s12918-018-0530-9

Zogopoulos V., Saxami G., Malatras A., Papadopoulos K., Tsotra I., Iconomidou V., and Michalopoulos I., 2022, Approaches in gene coexpression analysis in eukaryotes, Biology, 11(7): 1019.
https://doi.org/10.3390/biology11071019