Dean&Francis
ISSN 2959-409X

Application of Gene Expression Profiling Big Data in Anticancer Drug Discovery and Development

Zhihuan Fang[1]

Xiangchen Li[2,]

Zixuan Zhou[2]

[1]Hangzhou Entel Foreign Language School, Hangzhou, China
[2] St. George's School, Vancouver, BC, Canada
[3] The Affiliated International School of Shenzhen University, Shenzhen, China

*Corresponding author: harry.li26@stgeorges.bc.ca

Abstract:

As an important component of functional genomics, gene expression profiling plays an important role in many fields such as biology, medicine and drug development. Especially with the concept of precision medicine, integrating multi-omics data including gene expression profiles for personalized medicine is a future trend. From the background of anticancer drug development, the big data analysis method based on cellular perturbation expression profiling is introduced, and its application in anticancer covenant development is summarized. The findings suggest that continuous improvement of these methods will drive the discovery of novel biomarkers that can predict drug efficacy and toxicity, thereby accelerating the drug development process and improving patient prognosis. In the future, gene expression profiling is expected to be widely used in clinical practice, changing the landscape of cancer management and ushering in a new era of individualized cancer treatment that will achieve superior efficacy with fewer side effects than ever before.

Keywords: Gene expression; big data; anticancer drugs.

1. Introduction

Gene expression profiling big data, as the cornerstone of functional genomics, plays a key role in biology, medicine and drug discovery. The development of high-throughput technologies, such as gene sequencing and transcriptomics, enables us to deeply analyze the gene expression of cells or tissues under specific conditions, providing massive data for understanding biological mechanisms, disease causes, and drug effects. The rise of precision medicine has led to the fusion of gene expression profiling and multi-omics data to promote personalized medicine. In the field of anti-cancer drug development, gene expression profiling big data has helped identify tumor-related gene changes, reveal tumor heterogeneity, and lay the foundation for targeted therapy development. However, the massive nature and complexity of the data have become a challenge, and there is still a research gap on how to efficiently mine the guidance information and translate it into clinical applications.

This study focuses on the potential of gene expression profiling big data in anticancer drug discovery and development, and accelerates new drug discov-

ISSN 2959-409X

ery and development through innovative analysis methods and algorithm optimization to accurately predict drug targets, assess efficacy and toxicity. Meanwhile, personalized medication models are constructed to support clinical precision medicine. This study aims to fill the research gaps and promote scientific progress with significant clinical application value and social significance, aiming to provide safer and more effective treatment options for cancer patients.

2. Development Trends and Goals of Precision Medicine in China

Precision medicine is booming in China, with a clear trend towards personalization, efficiency and intelligence [1]. Big data, as a core driving force, has shown great potential especially in anti-cancer drug discovery and development. Through the application of high-throughput technologies such as second-generation sequencing, transcriptomics, and whole genomics, scientists are able to dig deep into gene expression profiles and accurately identify tumor-specific markers and drug targets. These technologies have not only accelerated the process of new drug discovery, but also facilitated the personalized prediction of drug efficacy and toxicity, making treatment more precise and effective [2]. China is actively promoting the construction of a precision medicine system, aiming at realizing early diagnosis, precise treatment and prognosis assessment of cancer and other complex diseases through interdisciplinary cooperation and technological innovation, so as to improve the survival rate and quality of life of patients. In the future, with the deep integration of big data and artificial intelligence technology, precision medicine will usher in a broader development prospect in China, contributing Chinese wisdom and strength to the global healthcare cause.

The organization‘s long-term plan will be expanded to encompass a comprehensive coverage of all human diseases, with an in-depth assessment of disease risk factors. analysis of pathological mechanisms, and prediction of optimal therapies for various diseases. Under the influence of the U.S. “precision medicine” program, China is actively committed to the accelerated development of the field of precision medicine. 2016 saw the launch of a precision medicine program in China that surpasses similar programs in the U.S., with an estimated 60 billion yuan spent on the program by 2030. 2016 saw the Ministry of Science and Technology (MOST) of the People‘s Republic of China (PRC) launch a new precision medicine program in March of that year, with the aim of developing the program in China [3]. The MOST launched an official report on China‘s precision medicine sector in March of that year. The ultimate goal of the program, specifically titled “Precision Medicine Research,” is stated as follows.

(1) Select common and some rare diseases with high prevalence rates in China as the starting point for research, and establish a specialized cohort of health and major diseases with millions of samples through collaborative research.

(2) Construct a platform for sharing biomedical big data, aiming to overcome the technical difficulties of histology in clinical application and big data analysis, which will be used to discover and validate disease-related biomarkers, targets, and agents, and to develop a set of experimental and analytical technology systems to support early warning, accurate diagnosis, effective treatment, and precise evaluation of therapeutic efficacy of diseases.

(3) The organization will prioritize the needs of clinical application and establish a comprehensive precise diagnosis and treatment plan, prediction and early warning, early screening, typing and classification, individualized treatment, efficacy and safety prediction and monitoring of major diseases; furthermore, The organization will develop a classified application technology platform designed to support the entire spectrum of precision medicine applications. including biomedical big data reference consulting, analyzing and judging, rapid calculation and precise decision-making.

(4) Commit to building a demonstration, application and promotion system of precision medicine clinical programs for typical diseases of the Chinese population, and promote the inclusion of a series of precision therapeutic drugs and molecular testing technology products in the national health insurance catalog, which aims to significantly improve the national health level, reduce ineffective and excessive medical treatment, avoid harmful treatments, control the rapid expansion of medical costs, and provide scientific and technological support for the development of precision medicine to become a brand-new growth point for economic and social development [4]. It will also provide scientific and technological support for the development of precision medicine, making it a brand new growth point for economic and social development [5].

3. Cellular Perturbation Expression Profiling Big Data Analyzer

The gene expression dataset obtained by measuring genome-wide transcript levels of cells at multiple time points under specific perturbations, such as gene repres-

sion, gene overexpression, or different concentrations of small molecule compounds, is known as a perturbed gene expression profile of a cell. In 2006, Lamb et al. were the first to perform a genome-wide expression profiling of 54 human cancer cell lines in the presence of 1309 drugs. expression profiles and created an association mapping dataset accordingly. Using the CMap platform, the association network between genes, diseases and drugs was constructed by comparatively analyzing the gene expression profiles of tissue cells under different physiological, pathological and drug-activated conditions, which was then successfully applied to drug repositioning and mining of new drug action mechanisms [6]. The method consists of 3 parts: exploring the depth of the database in the field of gene expression profiling, and applying pattern matching algorithms in order to reveal the hidden laws in biological information. A large number of expression profiles consisting of human cell lines are used as a reference database, and the query expression profile imprints are scored by a pattern matching algorithm in order to assess the degree and direction of enrichment [7]; the drugs are sorted according to their scoring of association with the target cell state, and the drugs at the top and the drugs at the bottom are respectively associated with the The drugs at the top and the drugs at the bottom are associated with the expression latent imprints of the target cell states, respectively.

Referring to expression profiling data in different biological states (e.g., disease models, drug action, etc.), including Molecular Stamping Database (MSigDB) and Gene Stamping Database (GeneSigDB), etc., these resources provide a rich set of reference expression imprints, which greatly facilitates biomedical and translational medicine research. However, the scale of the available data is limited, e.g., the CMap project covers only a small number of cell lines and expression profiles after drug action [8]. In response, the NIH-funded Broad Institute initiated the LINCS project, which aims to measure gene expression changes in multiple cell lines under different types of perturbations on a large scale via L1000 technology, significantly reducing costs and increasing data size. As of 2016, LINCS has made public more than a million expression profiling data covering multiple cancer cell lines and multiple perturbations (e.g., gene silencing, overexpression, small molecule compounds) [9]. In the face of this big data challenge, multiple factors need to be considered, and efficient algorithms (e.g., deep learning) and massively parallel computing techniques need to be used to process and analyze the data in order to mine valuable biological information and promote the in-depth development of biomedical research.

4. Application of Datamapping Big Data in Anticancer Drug Discovery and Development

In the drug discovery process, a new strategy that distinguishes itself from traditional approaches is the large-scale correlation screening based on the unique gene expression patterns of a specific disease phenotype, which is guided by association mapping, as opposed to high-throughput screening tools that rely solely on a single pharmacodynamic metric [10]. For the development of cancer therapeutics, a common approach involves analyzing biological samples from patients with multiple cancers to identify specific gene expression patterns, referred to as expression profiling imprints, and comparing and correlating these imprints with cellular gene expression profiles under the influence of a wide range of drugs in public databases to screen for potential drug candidates that show significant negative correlations. These drug candidates are then evaluated for their anticancer activity by laboratory and preclinical in vitro assays. In the past decade, the spectrum imprint comparison and association network analysis technology led by CMap has achieved remarkable results in the exploration of anticancer drugs, which constructs a bridge between basic science research and drug development, and is widely used in the field of systems biology.

By analyzing the expression profiling data on gene chips, a signature marker highly correlated with lung adenocarcinoma was discovered, which was then utilized to identify a significant negative correlation between drugs such as HSP90 inhibitors, PPAR antagonists, and PI3K inhibitors and the signature marker. Experimental validation showed that AAG was effective in inhibiting the proliferation of lung adenocarcinoma cells. Siavelis and colleagues integrated five expression profiling datasets related to Alzheimer‘s disease and analyzed the cellular expression profiling datasets involved in drug action using four different analytical tools, which resulted in the revelation of twenty-seven drugs that could potentially treat Alzheimer‘s disease. Further signaling pathways and protein interaction network analysis revealed that these drugs may exert their effects by affecting the epidermal growth factor receptor-related pathway. The use of LINCS cell expression profiling data not only enables the prediction of anticancer drugs through the expression profiles of small molecule compounds, but also integrates gene silencing and overexpression data to explore possible anticancer drug targets, identify potential targets of action of

ISSN 2959-409X

small molecule compounds, and shape the gene regulatory networks of specific biological processes and identify key genes therein. As shown in Table 1.

Table 1. Summary of CMap and LINCS for anticancer drug development studies

Illnesses	Veterinary drug	Key findings
Liver cancer	Cisplatin (cisplatin), Sorafenib (sorafenib), 5-fluorouracil (5-fluorouracil)	3 Drugs Increase Histone Acetylation and Reverse the Expression Profile of Patients with Low Survival to Patients with High Survival
Glioblastoma	Pyrvinium	CD133 targeting inhibits glioblastoma self-renewal & proliferation in ex vivo studies, marking tumor stem cells.
Colorectal cancer	Chlorpromazine	Chlorpromazine, an antipsychotic, halts growth & induces apoptosis in p53-mutant colon cancer cells, modulating SIRT1, a key deacetylase in gene regulation, metabolism, & cancer.
Acute lymphoblastic leukemia	Retinoids	Retinoid upregulates ikzf1, a zinc-finger transcription factor crucial for lymphocyte differentiation. This inhibits bcr-abl1 leukemia cell growth and boosts dasatinib efficacy.
Colorectal cancer	Irinotecan (Irinotecan) and Etoposide (Etoposide) among more than 10 drugs	Multiple colon cancer datasets formed expression profiles, screening >10 drugs, including known chemotherapeutics, for associations.
Diffuse large b-cell lymphoma	Doxycycline	The antibiotic doxycycline can affect several lymphoma-forming signaling pathways and inhibit tumor cell growth, with CSN5 as its potential drug target
Adenocarcinoma of the lungs	Bezafibrate	Benzafibrate inhibits lung adenocarcinoma cell growth by targeting CDK2 (cell cycle protein kinase)
Lung cancer	Forskolin with 6 PGE-2 analogs	CRTC1 activators are associated with LKB1-deficient lung cancer, and COX-2 inhibitors suppress LKB1-deficient lung adenocarcinoma cell growth

5. Conclusion

In summary, gene expression profiling, as a key technological approach in anticancer drug discovery and development, runs through all aspects of early discovery, late-stage research and development, and clinical drug guidance. With the leap forward in sequencing technology, the cost of transcriptome sequencing has been significantly reduced, making it possible to measure gene expression profiling data on a large scale under a variety of conditions. This study not only helps to fill the gaps in the current research field and promote the scientific progress of anticancer drug research and development, but also has important clinical application value and social significance, which is expected to bring safer and more effective therapeutic options for cancer patients. Combined with the latest gene expression profiling methods, especially the data expansion in the second phase of the LINCS project, it opens up a broader prospect for drug discovery. Entering the multi-omics era, the fusion analysis of gene expression profiling and other histological data has become a trend, and combined with artificial intelligence technologies such as the IBM Watson system, it has been accurately serving the medical practice, providing personalized and optimal treatment plans by comparing the patient‘s histological data with the knowledge base of the disease. Despite the challenges of small sample size and feature redundancy, gene expression profiling will become more and more important in anticancer drug discovery and development as technology continues to innovate, contributing more valuable information to clinical medical research.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Guan Bo, Shan Weimin, Li Ming. Screening of potential therapeutic agents for bladder uroepithelial cancer based on gene

expression profiling data[J]. Journal of Modern Genitourinary Oncology, 2023, 15(4): 203-207.

[2] Li Xiaotian, Wu Qifei, Zhang Dong. Screening of lung adenocarcinoma prognosis-associated genes based on a comprehensive gene expression database and the Cancer Genome Atlas database[J]. Cancer Progress, 2024, 22(7): 745-753.

[3] Bow W.H., Sun H.Q., Shan Z.M., et al. Exploration of key genes and pathways in childhood pneumonia based on gene expression profiling[J]. Journal of Hebei Medical University, 2024, 45(5): 501-507.

[4] Han Xiaoling, Wang Shengchun, Han Prettyying, et al. Integration of multiple gene expression profiling data to identify key genes associated with osteosarcoma development and progression[J]. Journal of Guangdong Medical University, 2023, 41(6): 613-618.

[5] Lin Yaru, Yang Xi, Li Jinsong, et al. Analysis of SSBP1 gene expression and prognosis in hepatocellular carcinoma based on TCGA database[J]. Journal of Ningxia Medical University, 2024, 46(2): 146-152.

[6] Xiao Jingchuan, Zhang Yingai. Bioinformatics analysis of hepatocellular carcinoma genes based on expression profiling microarray data[J]. Journal of Central South University (Medical Edition), 2020, 45(9): 1053-1060.

[7] Chen Chao, Chen Tianxiang, Liu Qianwei, et al. Study of Hub genes in hepatocellular carcinoma analyzed and identified based on weighted gene co-expression network and cancer genome atlas clinical data[J]. Chinese Family Medicine, 2024, 27(32): 4050-4059.

[8] Yang Ning, Zhang Zhiqiang, Huang Feihu, et al. Modeling and evolutionary characterization of scientific data citation networks--an example of gene expression dataset[J]. Modern Intelligence, 2024, 44(5): 45-57.

[9] Chen Feng, Li Huainan, Zhang Xiaoyun, et al. Analysis of differential gene expression and immune cell infiltration in gout gene expression profile[J]. Chinese Journal of Immunology, 2024, 40(3): 592-598.

[10] Li Kunpeng, Wang Zepeng, Zhou Yu, et al. Research progress on the application of artificial intelligence in tumor gene expression data[J]. Chinese Journal of Medical Physics, 2024, 41(3): 389-396.