Integrating Single-cell RNA Sequencing and Chinese Medicine Dialectics for Personalized Disease Treatment: A Survey

Abstract

This survey paper explores the interdisciplinary integration of single-cell RNA sequencing (scRNA-seq) with Chinese medicine dialectics, aiming to advance personalized disease treatment by combining modern genomic technologies with traditional medical philosophies. The paper is structured to provide foundational knowledge on scRNA-seq, its role in genomics, and the principles of Chinese medicine, followed by an examination of their integration for creating personalized treatment strategies. Key advancements in scRNA-seq technologies, such as the Single-Cell Latent Diffusion (SCLD) method, enhance the understanding of cellular heterogeneity, crucial for unraveling complex disease mechanisms. Advanced computational techniques, including scFusionTTT and Poisson mash models, are highlighted for their superior performance in multi-omics data analysis. The integration of AI and machine learning further enhances predictive capabilities, with frameworks like MLE ensuring data privacy while improving precision medicine outcomes. Ethical and privacy concerns are addressed, emphasizing the need for robust data protection measures and decentralized trust mechanisms. Case studies demonstrate the successful application of integrated approaches in conditions like Helicobacter pylori infection and novel coronavirus pneumonia, underscoring the potential of combining genomic insights with traditional Chinese medicine principles. Future directions focus on improving data quality, model interpretability, and ethical frameworks to fully realize the potential of this interdisciplinary approach, promising more precise, effective, and individualized healthcare solutions.

1 Introduction

1.1 Structure of the Survey

This survey is structured to facilitate an interdisciplinary exploration of integrating single-cell RNA sequencing with Chinese medicine dialectics for personalized disease treatment. It commences with an **Introduction** that underscores the importance of merging traditional Chinese medicine with modern genomic technologies. The subsequent section, **Background and Core Concepts**, provides essential knowledge on single-cell RNA sequencing, genomics, and Chinese medicine dialectics, elucidating their relevance to personalized medicine. The segment titled **Single-cell RNA Sequencing and Its Role in Genomics** addresses technological advancements, challenges, and the implications of scRNA-seq in elucidating disease mechanisms, supported by computational techniques such as those from nferX, accessible to both specialists and non-specialists [1]. Following this, **Chinese Medicine Dialectics in Disease Treatment** examines the principles and applications of Chinese medicine in specific diseases. The section **Integration of Genomic Technologies and Traditional Chinese Medicine** explores the synthesis of genomic data with Chinese medicine to develop personalized treatment strategies, backed by case studies. The penultimate section, **Personalized Medicine: Strategies and Applications**, discusses strategies for developing personalized medicine, emphasizing
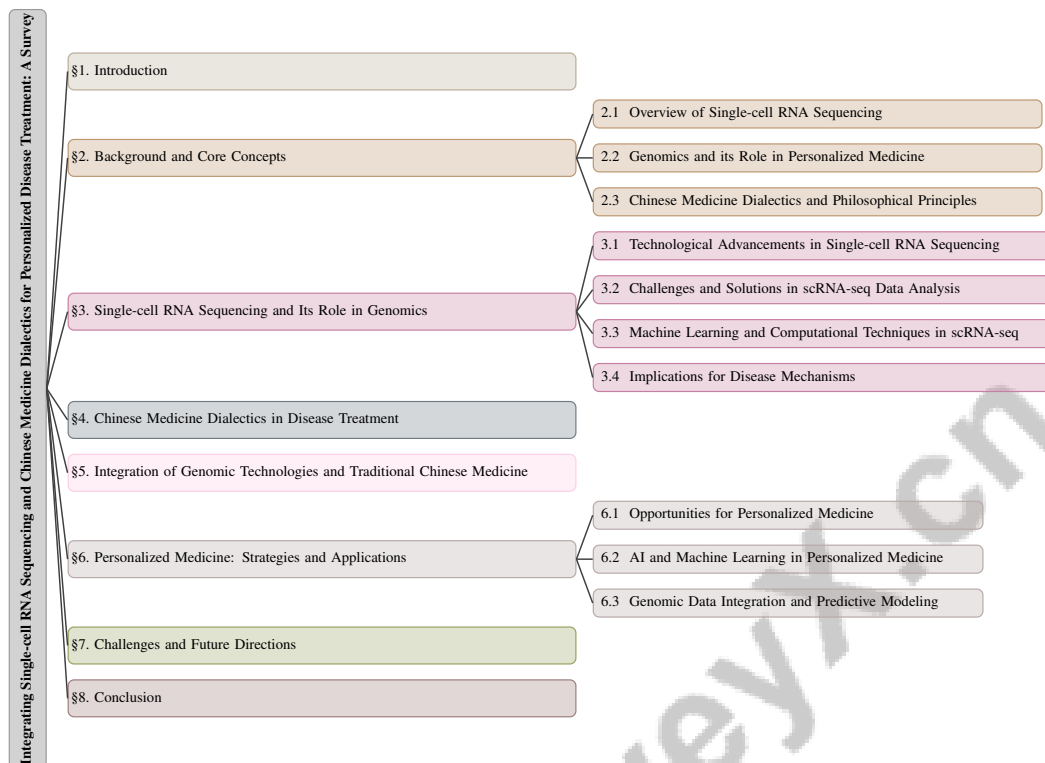
Figure 1: chapter structure

the roles of AI and machine learning. The survey concludes with **Challenges and Future Directions**, identifying key obstacles and innovative approaches for future research in this interdisciplinary field.

1.2 Interdisciplinary Approaches in Personalized Medicine

Advancements in personalized medicine require a robust interdisciplinary framework that integrates diverse scientific disciplines to address the complexities of human health. The Human Genome Project (HGP) exemplifies a pivotal initiative in molecular biology, mapping DNA sequences and identifying human genes, thereby establishing a foundational platform for personalized medical treatments [2]. However, the limitations of genomic precision medicine in effectively treating cancer patients highlight the necessity for functional methodologies that complement genetic data to enhance patient outcomes [3].

Integrating genetic information with single-cell RNA sequencing data is vital for understanding the impact of genetic variants on gene expression, thereby refining personalized disease treatment strategies [4]. The iASiS project exemplifies the need for comprehensive access to diverse data sources to generate actionable insights for personalized medicine policy-making, emphasizing interdisciplinary collaboration [5]. Furthermore, the integration of quantum computing in medicine indicates the transformative potential of quantum algorithms in addressing knowledge gaps and advancing computational medicine [6].

Balancing the benefits of Whole Genome Sequencing (WGS) with the ethical and privacy challenges it presents necessitates collaboration among computer scientists, biomedical researchers, and legal experts. Explainable artificial intelligence (xAI) methods enhance the interpretability of deep learning models in multi-omics research, particularly in clinical settings, exemplifying the intersection of AI and biomedical research [7]. Interdisciplinary collaboration is also crucial for advancing interpretable machine learning in genomics, significantly impacting the personalization of medical interventions [8].

Moreover, addressing the intersection of personalized medicine and health equity involves overcoming cost barriers and ethical considerations, ensuring that advancements in personalized medicine are accessible to diverse populations [9]. The misalignment between societal definitions of race and

research categorizations further underscores the need for interdisciplinary approaches to ensure equitable and effective personalized medicine [10]. Additionally, the implications of kidney tonifying traditional Chinese medicine (KTTCM) for osteoporosis prevention and treatment reflect the growing demand for alternative or complementary treatment options in an aging population [11]. Collectively, these interdisciplinary efforts are essential for the evolution of personalized medicine, enabling the integration of diverse data sources and methodologies to tailor treatments to individual patient profiles.

1.3 Significance of Integrating Traditional and Modern Technologies

The integration of traditional Chinese medicine (TCM) with modern genomic technologies represents a promising pathway for advancing personalized medicine. This synthesis combines TCM's holistic and individualized approach with the precision of genomic insights. Network pharmacology has emerged as a valuable tool, elucidating complex interactions within TCM formulations and their biological impacts, thereby enhancing TCM-based interventions [12]. For example, applying TCM to treat novel coronavirus pneumonia (NCP) highlights its potential for personalized treatment based on individual syndromes, aligning with precision medicine principles [13].

Moreover, the combination of TCM with modern medical approaches has shown promise in treating conditions such as Helicobacter pylori infection, underscoring the benefits of a hybrid methodology that merges traditional wisdom with contemporary scientific practices [14]. The Human Genome Project illustrates the transformative impact of integrating scientific innovation with societal needs, creating a foundation for personalized medicine that can be further enriched by TCM's diagnostic and therapeutic frameworks [2].

Dynamic diagnosis and treatment schemes (DDTS) in TCM, particularly optimized for chronic diseases like diabetes, demonstrate TCM's adaptability and efficacy when informed by genomic data, enabling more effective sequential treatment strategies [15]. Furthermore, integrating multi-platform molecular and genomic data enhances our understanding and treatment of complex diseases such as cancer, showcasing the complementary strengths of TCM and modern genomics [16].

Lastly, developing machine learning frameworks that incorporate encryption techniques to safeguard sensitive genomic data highlights the importance of protecting patient information while leveraging both traditional and modern technologies for improved health outcomes [17]. Collectively, these interdisciplinary efforts emphasize the rationale for integrating TCM with genomic technologies, promising more personalized, effective, and secure healthcare solutions.The following sections are organized as shown in Figure 1.

2 Background and Core Concepts

2.1 Overview of Single-cell RNA Sequencing

Single-cell RNA sequencing (scRNA-seq) is pivotal in genomics, offering detailed insights into transcriptomic variations at the cellular level. This technology enhances our understanding of cellular heterogeneity by enabling comprehensive gene expression analyses across thousands of individual cells in a single experiment [18]. scRNA-seq's ability to provide an unbiased snapshot of cellular diversity within tissues is crucial for deciphering complex biological systems and disease mechanisms.

Advancements in methodologies, including whole transcriptome amplification, cell isolation techniques, and barcoding strategies, have significantly improved scRNA-seq's utility, facilitating detailed profiling of cellular states and dynamics [19]. However, the high dimensionality and sparsity of scRNA-seq data present challenges for accurate cell type classification, necessitating robust computational methods [20]. The rapid evolution of scRNA-seq methods has created a demand for benchmarks to assess their performance, particularly in biological and medical research contexts [21].

In disease research, scRNA-seq has been instrumental in elucidating intercellular heterogeneity and providing insights into the molecular underpinnings of various conditions, including cardiovascular diseases [22]. Bayesian models, such as the gamma-negative binomial model, have effectively characterized molecular heterogeneity and elucidated underlying cellular processes [23]. Innovative computational approaches, including topological data analysis (TDA) and algorithms like Mapper, have advanced the analysis of high-dimensional genomic data, offering novel perspectives on tran-

scriptional variations in diseases such as lung adenocarcinoma [24]. These advancements underscore scRNA-seq's critical role in modern genomics, enhancing our understanding of cellular function and disease mechanisms, and paving the way for targeted therapeutic interventions.

2.2 Genomics and its Role in Personalized Medicine

Genomics plays a transformative role in personalized medicine by providing comprehensive insights into the genetic basis of diseases, enabling the tailoring of therapeutic strategies to individual genetic profiles. Integrating clinical and genomic datasets is essential for enhancing disease prediction models, as demonstrated by the limitations of relying solely on one data type [25]. This integration facilitates more precise disease predictions and treatment strategies.

Efficient allocation of sequencing resources is crucial, with optimal distribution between the number of cells sequenced and sequencing depth per cell being vital for accurate gene property estimation [26]. Genome-wide association studies (GWAS), integrated with functional genomics techniques like SNP enrichment and colocalization, provide valuable insights into gene-disease associations, enhancing our understanding of complex genetic architectures [27]. In oncology, genomics significantly contributes to identifying robust gene biomarkers from high-dimensional datasets, critical for precision medicine [28].

Computational models of single-cell genomic, transcriptomic, and proteomic data further highlight genomics' pivotal role in personalizing cancer treatment [29]. Machine learning frameworks, such as the MLE framework, exemplify how genomic data can enhance predictive modeling in precision medicine [17]. The challenge of analyzing RNA-seq data to identify distinct tumor subgroups and their gene expression profiles compared to healthy subjects underscores the need for sophisticated computational approaches [24]. Additionally, operationalizing complex causal responses from indirect interventions presents significant challenges, necessitating advanced methodologies for effective outcome predictions [30].

These advancements in genomics highlight its fundamental role in the evolution of personalized medicine. By integrating diverse genomic datasets and leveraging advanced computational techniques, genomics significantly enhances personalized healthcare solutions, tailoring treatments to the specific clinical and genetic characteristics of patients. This approach not only improves therapy effectiveness and precision but also raises important considerations regarding patient privacy and data security, necessitating robust frameworks for safe genomic data sharing among researchers [31, 32].

2.3 Chinese Medicine Dialectics and Philosophical Principles

Chinese medicine dialectics is founded on a philosophical framework that perceives the human body as an integrated whole, emphasizing balance and harmony between internal and external environments. This holistic perspective is central to Traditional Chinese Medicine (TCM), which employs multi-component and multi-targeted strategies to restore balance and promote health, particularly relevant in personalized medicine where individualized treatment strategies are crucial [15]. TCM's historical effectiveness in treating infectious diseases, exemplified by its application in novel coronavirus pneumonia, highlights its potential to address modern health challenges by classifying syndromes based on symptom patterns rather than isolated disease markers [33].

Integrating TCM principles with modern genomic technologies presents a promising avenue for enhancing disease treatment strategies. Establishing coherent mappings between distinct biological measurement domains, such as genomic and proteomic data, is essential for accurately representing relationships between corresponding samples [34]. Advanced computational techniques, including transformer-based models that leverage imaging and genomic data for predictions without invasive procedures, further support this integration [1].

Modeling and interpreting scRNA-seq data, characterized by high dimensionality and sparsity, necessitates sophisticated computational approaches. The challenge of identifying distinct cell types in scRNA-seq data is compounded by zero-inflated counts obscuring accurate clustering, underscoring the need for robust methodologies. Effective computational strategies are also required to predict disease states using scRNA-seq data, particularly for complex diseases like Multiple Sclerosis, which lack a definitive diagnostic test [35]. Linking GWAS findings, especially variants in non-coding regions, to biological functions poses significant challenges in understanding gene interactions that

4

contribute to complex traits. The integration of diverse genomic datasets from different studies can lead to inaccuracies due to demographic differences and calibration artifacts [36].

Combining the philosophical principles of TCM with genomic insights, supported by network pharmacology and advanced artificial intelligence techniques, creates a comprehensive framework for tailoring personalized treatment strategies. This approach addresses unique biological, environmental, and lifestyle factors, enhancing the efficacy of disease management and therapeutic interventions [37, 12]. By considering both biological and energetic profiles, this interdisciplinary approach has the potential to improve therapeutic efficacy and patient outcomes.

3 Single-cell RNA Sequencing and Its Role in Genomics

Category	Feature	Method
Technological Advancements in Single-cell RNA Sequencing	Dimensionality Reduction	SCLD[38]
	Resource Efficiency	OSBA[26]
Challenges and Solutions in scRNA-seq Data Analysis	Data Integration and Enhancement	ADV-MTL[36], DNNG[39]
	Pattern and Relationship Identification	PM[30]
	Dimensionality and Noise Management	LPS[28]
Machine Learning and Computational Techniques in scRNA-seq	Graph and Regularization Techniques	RZiMM-scRNA[40], SFAG[41]
	Data Integration Approaches	scFusionTTT[42], Cox-nnet[43], MLE[17]
	Automation and Scalability	CP[44], CA[45]
Implications for Disease Mechanisms	Data Integration and Analysis	GAT-MSP[35], RCI-NB[46], hGNB[23], CCP[47], EPC[20]
	Personalized Medicine	LSCV-MCEA[48]
	Tumor Classification	Mapper-RNA[24]

Table 1: This table provides a comprehensive summary of recent methodologies and computational techniques in single-cell RNA sequencing (scRNA-seq). It categorizes advancements into technological innovations, challenges and their solutions, machine learning applications, and implications for understanding disease mechanisms, highlighting specific methods that enhance scRNA-seq data analysis and interpretation.

Single-cell RNA sequencing (scRNA-seq) is a transformative technology in genomics, enabling detailed exploration of cellular diversity and gene expression. This section highlights recent advancements in scRNA-seq methodologies, emphasizing innovations that enhance data quality and interpretability, which are crucial for addressing challenges in scRNA-seq data analysis. Table 1 presents a detailed summary of the key methodologies and computational techniques that have been developed to advance single-cell RNA sequencing (scRNA-seq), addressing various challenges and applications in genomics. Additionally, Table 2 provides a comparative overview of key methodologies in single-cell RNA sequencing, elucidating their unique features and contributions to advancing genomics research. **??** illustrates the hierarchical organization of key advancements, challenges, and applications of scRNA-seq in genomics. It categorizes technological innovations, challenges with corresponding solutions, machine learning techniques, and implications for understanding disease mechanisms. Each category is further divided into specific methodologies and computational techniques that enhance scRNA-seq data analysis, offering insights into cellular diversity and personalized therapeutic strategies. This visual representation not only complements the textual discussion but also provides a structured overview that aids in comprehending the multifaceted nature of scRNA-seq advancements.

3.1 Technological Advancements in Single-cell RNA Sequencing

Recent innovations in scRNA-seq have significantly improved the analysis of cellular heterogeneity and gene expression dynamics. The Single-Cell Latent Diffusion (SCLD) methodology generates high-quality scRNA-seq samples, capturing diverse cellular subpopulations effectively [38]. The Hierarchical Gamma-Negative Binomial (hGNB) model, a Bayesian hierarchical approach, addresses high over-dispersion and covariate effects in scRNA-seq data without normalization [23].

Optimal Sequencing Budget Allocation (OSBA) optimizes resource allocation between cell numbers and sequencing depth, reducing gene expression estimation errors [26]. This balance is crucial for studies targeting rare cell populations [22]. Computational techniques like the correlated clustering and projection (CCP) method create lower-dimensional representations through gene clustering, enhancing data analysis [47]. The EnProCell method exemplifies this trend by using PCA and

5

MDA for cell type classification [20]. The LP-Stability method identifies robust gene signatures, outperforming existing clustering and classification techniques [28].

The Mapper algorithm visually represents RNA-seq data, aiding in identifying tumor subgroups and their gene expression characteristics [24]. Various scRNA-seq methods, such as Drop-seq and Smart-seq2, have been evaluated for their effectiveness in high-dimensional cluster analysis [21]. These advancements underscore the rapid evolution of scRNA-seq, facilitating precise analyses of cellular expression profiles and contributing to personalized therapeutic strategies, as highlighted by the omnigenic model [49].

3.2 Challenges and Solutions in scRNA-seq Data Analysis

Analyzing scRNA-seq data involves challenges due to high dimensionality, sparsity, and intrinsic noise. Technical noise from single-cell isolation and low capture efficiency complicates accurate gene expression assessment [18]. Dimensionality reduction methods like PCA, t-SNE, and UMAP are limited by their reliance on matrix diagonalization, which can be computationally intensive and unstable with noisy data [47].

The complexity of genomic data complicates survival predictions in clinical contexts, as deriving insights from limited patient samples is challenging [50]. Traditional machine learning methods may fail to capture the intricacies of scRNA-seq data, leading to reduced predictive accuracy [39]. The curse of dimensionality and insufficient samples per tumor subtype hinder the identification of statistically significant causal relationships [28].

Existing methods struggle to integrate diverse data types while maintaining predictive power [43]. Non-specific mRNA amplification and the need for automated cell isolation add to the challenges of processing large cell numbers [19]. Additionally, ambiguity in causal relationships in high-dimensional settings complicates direct manipulation of underlying causes [30].

Innovative computational approaches are essential for overcoming these challenges. Multi-task learning and adversarial representation learning enhance predictive modeling by integrating heterogeneous cohorts [36]. Advanced imputation techniques that accurately identify technical zeros and leverage multi-omics integration are crucial for improving data reliability. Developing robust methods for inferring clonal dynamics and understanding tumor microenvironment impacts on cancer evolution is vital [29].

Improving scalability and privacy through novel algorithms will facilitate large dataset handling. Establishing standardized protocols for data processing and sharing will promote collaboration and tool interoperability. Enhancing statistical rigor and developing frameworks for cell-to-cell communication analysis will significantly improve the interpretability and applicability of scRNA-seq data. Systematic evaluations of computational pipelines and experimental protocols, along with advanced machine learning techniques, can address challenges like data sparsity and dimensionality reduction. By optimizing normalization and library preparation, researchers can enhance biological signal detection, leading to more accurate insights into cellular behavior, gene function, and disease mechanisms [34, 51, 47, 52]. Addressing these obstacles will allow scRNA-seq to continue providing invaluable insights into cellular heterogeneity and disease mechanisms, fostering more precise therapeutic strategies.

3.3 Machine Learning and Computational Techniques in scRNA-seq

Machine learning and computational techniques are crucial for analyzing scRNA-seq data, offering innovative solutions to the complexities and high dimensionality of these datasets. Deep learning architectures significantly enhance the precision and efficiency of scRNA-seq data analysis [53].

Advanced methods, such as the Poisson multivariate adaptive shrinkage (Poisson mash) model, integrate raw count data modeling with effect-sharing pattern identification, improving accuracy [54]. The SFAG method, incorporating graph regularization into non-negative matrix factorization, enhances imputation accuracy by leveraging technical zero identification and cell relationships [41].

Frameworks like scFusionTTT, which utilize TTT layers alongside gene and protein order information, improve the integration of transcriptomics and proteomics data, offering a more comprehensive

6

analysis than traditional models [42]. LLM-driven multi-agent frameworks, inspired by large language models, automate scRNA-seq data analysis, enhancing scalability and efficiency [45].

Integrating multi-omics data through advanced models, including neural networks and pathway-based approaches, has improved survival predictions, highlighting the potential of these techniques in clinical outcomes [43]. Incorporating encrypted genomic data into machine learning frameworks ensures patient privacy while enabling predictive analytics, underscoring the importance of secure data handling in precision medicine [17].

Algorithms like CloudPred autonomously identify relevant cell subpopulations for disease phenotype prediction without prior annotations, showcasing the potential of machine learning in scRNA-seq analysis [44]. The introduction of benchmarks incorporating dropout rates and sensitivity measures provides a comprehensive evaluation framework, enhancing robustness in scRNA-seq analysis [21].

Recent advancements in machine learning and computational techniques are transforming our understanding of gene expression at the cellular level, enabling more precise and personalized therapeutic strategies. These innovations encompass sophisticated data processing methods, dimensionality reduction techniques, and clustering algorithms that identify distinct cell populations based on gene expression profiles. The application of deep learning has proven particularly beneficial in extracting meaningful features from complex, high-dimensional scRNA-seq data, enhancing our ability to analyze cellular states and phenotypes. By integrating computational innovations with biological insights, researchers are better equipped to tackle therapeutic development challenges, from target discovery to personalized medicine, paving the way for significant advancements in clinical applications [31, 55, 56].

(a) Performance Statistics of Various Machine Learning Models[57]

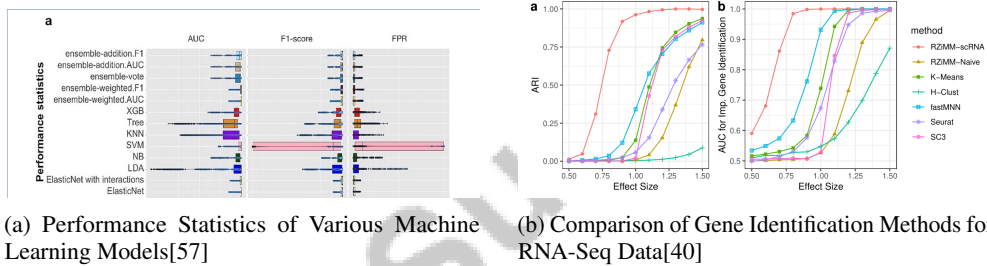(b) Comparison of Gene Identification Methods for RNA-Seq Data[40]

Figure 2: Examples of Machine Learning and Computational Techniques in scRNA-seq

As shown in Figure 2, scRNA-seq has emerged as a pivotal technique in genomics, offering unparalleled insights into cellular heterogeneity and the complexity of biological tissues. The integration of machine learning and computational techniques into scRNA-seq data analysis enhances the ability to decipher intricate patterns and derive meaningful biological inferences. The first figure compares various machine learning models, showcasing performance metrics such as AUC, F1-score, and FPR, evaluating models like ensemble-addition and ensemble-weighted alongside traditional approaches such as XGB, Tree, KNN, and SVM. The second figure focuses on the comparison of gene identification methods for RNA-Seq data, highlighting the Average Recall Index (ARI) and Area Under the Curve (AUC) across different effect sizes. Methods such as RZiMM-scRNA and RZiMM-naive are evaluated, demonstrating the critical role of computational techniques in enhancing gene identification accuracy. Together, these examples underscore the transformative impact of machine learning and computational methodologies in advancing our understanding of cellular genomics through scRNA-seq [57, 40].

3.4 Implications for Disease Mechanisms

scRNA-seq has revolutionized the understanding of disease mechanisms by providing detailed insights into cellular heterogeneity and gene expression dynamics. This technology enables the precise identification of rare cell populations, such as antigen-specific memory B cells and dormant cancer cells, clarifying their roles in disease progression. By integrating scRNA-seq with extensive biomedical literature analysis, researchers uncover unique molecular signatures and cellular behaviors within these populations, enhancing the understanding of complex biological processes, including mechanisms underlying diseases like COVID-19, and driving the discovery of new therapeutic targets

[34, 20, 18, 58, 1]. Profiling hundreds of thousands of cells through scRNA-seq underscores the significance of these advancements for future research.

The integration of scRNA-seq data with advanced computational models has deepened insights into cellular dynamics and disease states. For instance, the Graph Attention Network predicts disease states from single-cell data, achieving 92

Innovative frameworks like RCI-NB pinpoint patient-specific root causal contributions from scRNA-seq datasets, integrating novel sparse regression and goodness of fit testing procedures to circumvent Poisson measurement errors [46]. The Large-scale Constrained Variables based Multi-tasking Cooperative Evolutionary Algorithm (LSCV-MCEA) effectively identifies personalized drug targets, enhancing the understanding of disease pathways [48].

The Mapper-based framework has successfully identified distinct tumor subgroups and their gene expression profiles, providing critical insights into lung cancer development pathways [24]. The correlated clustering and projection (CCP) method improves dimensionality reduction without the computational drawbacks of matrix diagonalization, enhancing visualization and analysis of cell differentiation [47]. EnProCell has achieved an accuracy of 98.91

Moreover, PrescDRL optimizes herbal prescription planning in Traditional Chinese Medicine (TCM), demonstrating superior curative effects and performance in prescription recommendations compared to traditional methods [15]. Collectively, these advancements illustrate the transformative impact of scRNA-seq on studying disease mechanisms. By integrating biological insights with computational innovations, scRNA-seq provides a comprehensive framework for understanding complex diseases, ultimately facilitating the development of targeted therapeutic interventions.

Feature	Single-Cell Latent Diffusion	Hierarchical Gamma-Negative Binomial	Optimal Sequencing Budget Allocation
Data Quality	High-quality Samples	High Over-dispersion	Reduced Estimation Errors
Optimization Focus	Cellular Subpopulations	Covariate Effects	Resource Allocation
Analytical Approach	Diverse Capture	Bayesian Hierarchical	Budget Optimization

Table 2: This table provides a comparative analysis of three advanced methodologies in single-cell RNA sequencing. It highlights their distinct features, including data quality, optimization focus, and analytical approach, offering insights into their respective strengths and applications in genomics research.

4 Chinese Medicine Dialectics in Disease Treatment

4.1 Principles and Challenges of Chinese Medicine Dialectics

Chinese medicine dialectics perceives the human body as an interconnected system, prioritizing harmony between internal and external environments, which is fundamental to Traditional Chinese Medicine (TCM). TCM employs multi-component strategies to restore balance, exemplified by its effectiveness in treating osteoporosis through kidney tonifying traditional Chinese medicine (KTTCM), offering stable efficacy with fewer side effects compared to conventional treatments [11].

Nevertheless, the application of Chinese medicine dialectics presents challenges. The biological variation addressed by TCM, through individual patient consideration, complicates its implementation [59]. High dimensionality in genomic datasets further complicates integrating TCM principles, obscuring the identification of informative genes due to complexity and redundancy [50]. Precise biomarker identification and understanding their roles in diseases such as cancer are crucial for TCM's successful application, as demonstrated by functional integrative Bayesian analysis (fiBAG) methodologies [16].

Research often neglects the complexities of clonal dynamics and tumor microenvironment interactions, leading to incomplete understandings of cancer evolution [29]. The reliance on large datasets, not always accessible in healthcare, limits the applicability of studies within the TCM context [60]. Enhancing diagnostic accuracy by integrating genomic and imaging data aligns with TCM's holistic diagnostic approach [61].

Innovative computational methods, such as the LP-Stability method, outperform traditional clustering algorithms by yielding more accurate and biologically relevant gene signatures, advancing TCM dialectics in personalized medicine [28]. Addressing these challenges through comprehensive research

8

and innovative methodologies is essential for realizing the full potential of Chinese medicine dialectics in contemporary healthcare, enabling the integration of genomic insights to enhance personalized treatment strategies and improve patient outcomes.

4.2 Applications of Chinese Medicine Dialectics in Specific Diseases

The application of Chinese medicine dialectics in treating specific diseases demonstrates its capacity to offer personalized therapeutic strategies based on a holistic understanding of patient conditions. In colorectal cancer, integrating Chinese medicine dialectics with modern genomic insights has revealed significant metabolic behavior differences among cancer subtypes, emphasizing the need to tailor treatment strategies to these unique metabolic profiles to enhance therapeutic efficacy [62].

Despite promising applications, challenges such as biases in RNA-Seq library preparation, high sequencing costs, and data analysis complexities impede current research. These obstacles hinder the comprehensive integration of transcriptomic data within the Chinese medicine dialectical framework, limiting its full potential in disease treatment [63]. Overcoming these challenges through innovative methodologies and cost-effective technologies is crucial for advancing the application of Chinese medicine dialectics in personalized medicine.

By merging the philosophical principles of Chinese medicine with cutting-edge genomic technologies, there is a substantial opportunity to develop more effective treatment strategies for complex diseases. This interdisciplinary approach enhances our understanding of complex disease mechanisms and facilitates the creation of personalized healthcare solutions by integrating advanced AI-driven strategies with diverse fields such as genomics, proteomics, and clinical data analysis. Leveraging these technologies allows healthcare providers to customize treatment strategies based on individual genetic, environmental, and lifestyle factors, transforming disease management into a proactive, patient-centric model that enhances diagnostic accuracy and therapeutic effectiveness [37, 1, 60].

(a) AI-Driven Predictive Model for Disease Outcomes[37]

(b) The image shows a graph with two lines representing the expected reward given treatment, labeled R1 and R2, plotted against the treatment assigned.[64]
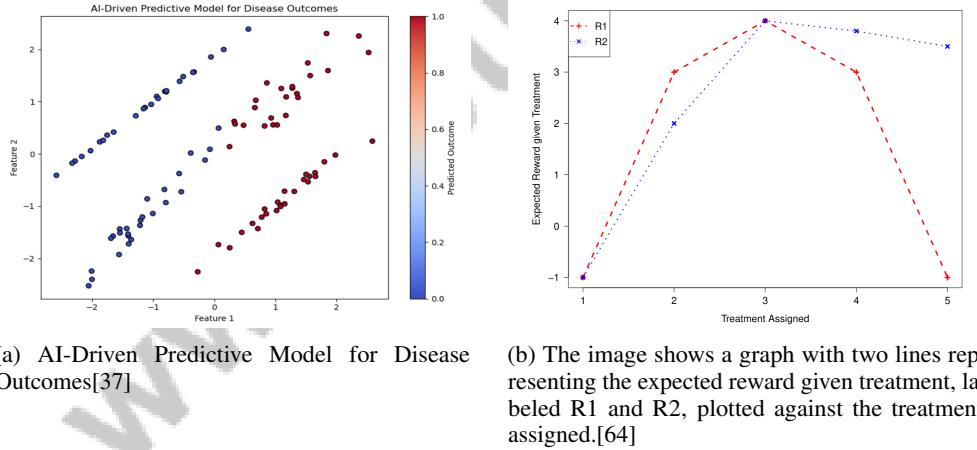
Figure 3: Examples of Applications of Chinese Medicine Dialectics in Specific Diseases

As shown in Figure 3, the exploration of Chinese Medicine Dialectics in treating specific diseases highlights the integration of traditional practices with modern technological advancements. The first example illustrates a scatter plot showcasing the relationship between two features and their predicted outcomes, where color-coded data points reveal varying levels of predicted outcomes. This visualization aids in identifying patterns and correlations crucial for tailoring individualized treatment strategies. The second example presents a graph depicting the expected reward given different treatments, with lines R1 and R2 indicating potential benefits of each treatment option. By plotting these expected rewards against assigned treatments, practitioners can make informed decisions regarding the most effective therapeutic approaches. Collectively, these figures underscore the potential of combining Chinese Medicine Dialectics with data-driven methodologies to optimize disease treatment and improve patient outcomes [37, 64].

5 Integration of Genomic Technologies and Traditional Chinese Medicine

5.1 Integration of Genomic Data with Chinese Medicine

Integrating genomic data with Traditional Chinese Medicine (TCM) facilitates personalized treatments by merging TCM's holistic principles with genomic precision. Advanced computational frameworks like LSCV-MCEA optimize this integration by balancing driver nodes with known drug-target data, improving treatment outcomes [48]. The MLE framework further exemplifies this potential by integrating genomic data with machine learning while ensuring patient privacy, aligning with modern ethical standards [17]. Instance-based learning approaches enhance disease prediction by combining clinical and genomic datasets, fostering strategies informed by both TCM and genomic insights [25].

Blockchain technology enhances genomic data privacy, ownership, and sharing, supporting secure integration with TCM [65]. By distinguishing between commercial and non-commercial applications, blockchain facilitates the ethical use of genomic data in personalized medicine. These integrative approaches underscore the transformative potential of merging genomic data with TCM, aiming to enhance therapeutic efficacy and patient outcomes. This interdisciplinary strategy leverages genomics, proteomics, and clinical data, shifting healthcare from reactive to proactive, patient-centric models [37, 1].

5.2 Challenges and Opportunities in Integrating Core Concepts

The integration of genomic technologies with TCM presents unique challenges and opportunities crucial for advancing personalized medicine. Privacy concerns and ethical issues regarding genomic data usage may deter participation in genomic research [66]. This is compounded by the inconsistent translation of genomic data into improved patient outcomes, highlighting the need for broader functional approaches beyond genomic information alone [3].

The complexity of gene interactions complicates integration, as many studies overlook peripheral genes, limiting genomic analyses' effectiveness [49]. Additionally, lack of standardization and interoperability across datasets hampers the development of comprehensive tools necessary for integrating diverse genomic and TCM data [67].

Despite these challenges, advanced computational methodologies offer significant opportunities for enhancing the integration of genomic technologies with TCM. The ADV-MTL approach, for example, improves predictive performance in cancer genomics by effectively integrating heterogeneous datasets and addressing data insufficiency, facilitating more personalized treatment strategies [36]. Innovative methods for estimating causal effects from interventions further demonstrate the potential for applying these insights across domains, paving the way for future research [30].

5.3 Case Studies and Applications

Several case studies illustrate the successful integration of genomic technologies with TCM, showcasing the potential of this interdisciplinary approach to enhance personalized treatments. A notable example is the application of TCM principles with genomic insights to address Helicobacter pylori infection, where a hybrid approach combining traditional wisdom with contemporary methodologies yielded promising results [14]. This case highlights the complementary strengths of TCM and genomics in developing effective therapeutic interventions.

Network pharmacology elucidates complex interactions within TCM formulations, identifying key bioactive compounds and their targets, thereby enhancing TCM-based interventions' precision [12]. This methodology optimizes treatment strategies for novel coronavirus pneumonia (NCP), aligning personalized treatment plans with precision medicine principles [13].

In chronic diseases like diabetes, dynamic diagnosis and treatment schemes (DDTS) informed by genomic insights have optimized herbal prescriptions and improved treatment outcomes [15]. This underscores TCM's adaptability and efficacy when informed by genomic data, facilitating more effective sequential treatment strategies.

Moreover, machine learning frameworks incorporating encryption techniques to protect sensitive genomic data highlight the importance of safeguarding patient information while utilizing both traditional and modern technologies for improved health outcomes [17]. These case studies collectively

demonstrate the transformative potential of integrating genomic technologies with TCM, offering promising avenues for developing personalized, effective, and secure healthcare solutions.

(a) Large language model and downstream tasks in transcriptomics[68]

(b) Deep Learning Model Training and Deployment Process[31]

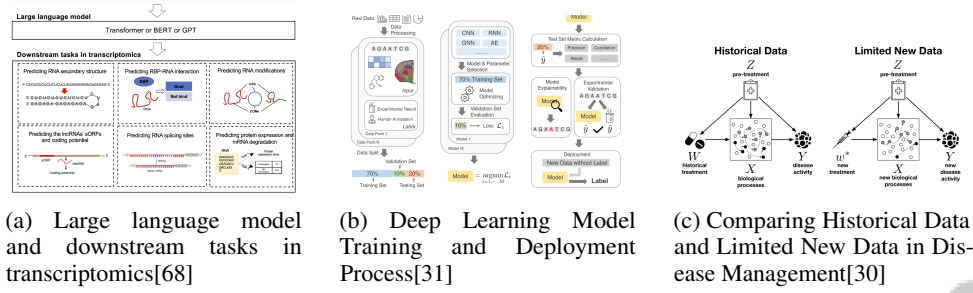(c) Comparing Historical Data and Limited New Data in Disease Management[30]

Figure 4: Examples of Case Studies and Applications

As illustrated in Figure 4, integrating genomic technologies with TCM represents a groundbreaking approach to enhancing healthcare outcomes through the fusion of ancient wisdom and modern science. This convergence is exemplified in various case studies and applications. The first example highlights large language models in transcriptomics, illustrating their capacity to predict complex RNA structures and interactions, offering insights into gene expression and protein synthesis. The second case study explores the deep learning model training and deployment process, showcasing how advanced computational models are optimized to interpret vast biological datasets, improving predictive accuracy in therapeutic applications. Lastly, the comparison of historical and limited new data in disease management underscores the potential of integrating traditional knowledge with cutting-edge genomic data to tailor personalized treatment strategies. These examples collectively demonstrate the transformative potential of combining genomic technologies with TCM, paving the way for more precise and effective healthcare solutions [68, 31, 30].

6 Personalized Medicine: Strategies and Applications

Personalized medicine aims to tailor healthcare interventions to individual patient characteristics, significantly enhancing treatment efficacy and outcomes. This section explores opportunities presented by integrating advanced technologies such as artificial intelligence (AI), machine learning (ML), and genomic data, illustrating how these innovations can revolutionize patient care.

6.1 Opportunities for Personalized Medicine

Personalized medicine offers transformative opportunities to revolutionize disease management by customizing therapeutic strategies to individual profiles. Integrating AI with precision medicine enhances treatment efficacy and minimizes adverse effects through comprehensive data analysis [37]. Machine learning algorithms, such as ElasticNet for small datasets and XGB for larger ones, are crucial for maximizing predictive accuracy [57]. Advanced clustering methods like RZiMM-scRNA improve biomarker detection and identify critical cell subgroups for targeted therapies in complex diseases [40].

Network pharmacology integrates traditional Chinese medicine (TCM) with modern scientific methods, enhancing treatment precision and enabling personalized interventions [12]. The effectiveness of TCM in treating novel coronavirus pneumonia (NCP) further illustrates its potential in personalized medicine [13]. Computational frameworks like LSCV-MCEA expand personalized therapy options by offering multiple candidate driver-node sets [48]. The GeoTyper pipeline enhances the accessibility of personalized medicine strategies in clinical settings [69].

Integrating multi-omics data improves predictive accuracy and biological interpretability, enhancing disease management strategies [43]. Pipelines like ASGARD demonstrate superior accuracy in providing personalized drug recommendations [70]. Combining clinical and genomic datasets enhances disease prediction models, achieving up to 73

Advancements in personalized medicine, including AI integration and digital twins, highlight its transformative potential by facilitating tailored strategies considering genetic, environmental, and

11

lifestyle factors. However, successful implementation requires addressing challenges related to cost, health equity, and robust data integration, particularly in low- and middle-income countries [71, 37, 3, 9]. By integrating cutting-edge technologies, personalized medicine promises tailored healthcare solutions that enhance patient outcomes and advance precision medicine.

6.2 AI and Machine Learning in Personalized Medicine

AI and ML revolutionize personalized medicine by analyzing complex datasets from genomic, clinical, and lifestyle information, facilitating accurate predictions of clinical outcomes and customization of therapeutic interventions. These technologies enhance treatment precision while addressing data privacy and security challenges, ensuring protection of sensitive health information [17, 37, 32, 31]. Deep learning techniques applied to scRNA-seq data improve data imputation and representation learning, essential for robust clustering and annotation of cellular data [72].

Incorporating uncertainty measures in AI models enhances trustworthiness and transparency, crucial for clinical decision-making [73]. Explainable AI (xAI) methods enhance interpretability, fostering trust in clinical contexts [7]. Interpretable machine learning (iML) approaches elucidate underlying mechanisms of ML models in genomic applications [8].

Different ML models, particularly deep learning approaches, vary in effectiveness for predicting gene expression and regulatory signals, emphasizing the importance of model selection for specific genomic applications [74]. Innovative frameworks like CellAgent leverage large language models to automate complex analysis tasks in scRNA-seq data [45]. CloudPred demonstrates superior performance in predicting clinical phenotypes from scRNA-seq data [44].

These advancements illustrate AI and ML's pivotal role in advancing personalized medicine. By integrating advanced AI technologies with comprehensive patient data, healthcare providers can develop highly tailored and effective treatment strategies. This multidisciplinary approach enhances diagnostic precision and treatment selections, significantly improving patient outcomes across complex diseases [37, 32, 3].

6.3 Genomic Data Integration and Predictive Modeling

Integrating genomic data with predictive modeling advances personalized treatment strategies by enhancing disease predictions and therapeutic interventions. Comprehensive frameworks that integrate technical solutions with ethical considerations are necessary due to the unique privacy risks associated with genomic data [75]. The iASiS benchmark exemplifies a robust framework for integrating diverse data sources, facilitating better decision-making [5].

Innovative methodologies like Metadata-guided Feature Disentanglement (MFD) enhance interpretability and accuracy by incorporating metadata into model training [76]. The Diet Networks method improves predictive performance by employing a parameter prediction network [77]. Combining genomic data integration with advanced models like MIGTrans enhances personalized treatment strategies in complex conditions [78].

Genomic data integration is crucial for developing personalized treatment strategies, as demonstrated by improved survival prediction accuracy in oncology through advanced gene selection methods [50]. Optimal sequencing depth determination achieves better estimation accuracy, particularly in shallow sequencing scenarios [26].

These integrative approaches underscore the transformative potential of combining genomic data with predictive modeling to develop personalized treatment strategies. Employing sophisticated computational techniques aims to significantly improve personalized medicine's accuracy and effectiveness, tailoring interventions to individual genetic profiles, medical histories, environments, and lifestyles. Integrating machine learning with genomic data enhances therapeutic discovery while addressing patient privacy challenges, ensuring sensitive health information protection [31, 32, 17].

7 Challenges and Future Directions

The integration of advanced technologies such as single-cell RNA sequencing (scRNA-seq) with traditional practices like Chinese medicine presents a complex landscape in personalized medicine.

This convergence enhances biological understanding but also introduces challenges, particularly in data quality and integration. This section explores critical issues underpinning effective personalized treatment strategies.

7.1 Data Quality and Integration Challenges

Integrating scRNA-seq data with Chinese medicine in personalized medicine faces challenges related to data quality and integration. Noise and sparsity in scRNA-seq data complicate accurate interpretation, with datasets often failing to represent cell type diversity or exhibiting high heterogeneity, undermining clustering and classification methods [20, 28]. Reliance on specific datasets limits generalizability across scRNA-seq applications, particularly when mapping in vitro signatures to in vivo counterparts, where heterogeneity and confounders obscure causal relationships [21].

Privacy concerns further hinder genomic data integration with electronic health records, restricting collaboration and machine learning efficacy. Computational complexity in scRNA-seq data analysis, particularly with noise or low-quality data, poses additional challenges. Parameter-sensitive techniques like Mapper graphical structures influence results' robustness, potentially leading to misleading interpretations if not optimized [77, 24, 79, 60].

Addressing these challenges requires sophisticated computational methodologies integrating diverse multi-omics data. Establishing cohesive frameworks promoting interoperability within the genomics community is crucial, leveraging advanced machine learning and deep learning techniques to analyze genomic complexities [39, 80, 33]. Overcoming these obstacles can enhance scRNA-seq integration with Chinese medicine, leading to precise personalized treatment strategies.

7.2 Model Interpretability and Validation

Model interpretability and validation are critical in personalized medicine, ensuring computational tools yield actionable insights for clinicians. The complexity of biological data can lead to models that, while powerful, are difficult to interpret, necessitating a balance between predictive accuracy and comprehensibility [1].

Validation requires rigorous testing to ensure reliability across diverse datasets and clinical scenarios. Robust frameworks addressing biases are essential, particularly in single-cell RNA sequencing, where annotation methods can be influenced by biased training data, especially for novel or rare cell types. Multi-model fusion strategies, like the LICT software package, improve annotation reliability by assessing discrepancies between model outputs and expert annotations. Integrative learning approaches accounting for heterogeneity across datasets can mitigate challenges posed by small sample sizes and high-dimensional data, enhancing weak biological signal detection [51, 81, 39, 82].

Innovative validation approaches, such as integrating multi-omics data and applying advanced cross-validation techniques, are crucial for enhancing predictive models' robustness and interpretability. Explainable AI (xAI) methods within these frameworks create transparent models that deepen molecular interaction understanding and improve patient outcome predictions amidst heterogeneous data [7, 82, 43, 80]. Developing xAI methods significantly enhances model interpretability, fostering trust in AI-driven solutions and advancing personalized medicine towards more transparent and reliable computational tools.

7.3 Ethical and Privacy Concerns

Integrating scRNA-seq with traditional Chinese medicine and genomic technologies raises significant ethical and privacy concerns that must be addressed for responsible advancement in personalized medicine. Handling sensitive genomic data poses risks of privacy breaches and genetic disclosure, necessitating stringent data protection measures [17].

AI applications in healthcare further complicate these concerns, requiring robust guidelines to govern AI use and ensure responsible patient data handling [37]. Collaborative efforts among researchers, ethicists, and policymakers are essential to address genomic privacy challenges and develop comprehensive frameworks safeguarding individual rights [66].

Blockchain technology offers promising solutions for enhancing data privacy and security in genomics, with its immutability and decentralization being key advantages. However, challenges related to

13

privacy protection and smart contract security must be addressed to fully leverage blockchain's potential in safeguarding genomic data [65]. Anonymized datasets in genomic research mitigate ethical concerns, ensuring individual identity protection while facilitating scientific insights [83].

Ethical and privacy considerations highlight the need for comprehensive strategies that harness personalized medicine's transformative potential while safeguarding individual rights. This includes addressing access disparities between high-income and low- and middle-income countries, ensuring robust data protection for sensitive genomic information, and implementing secure frameworks like blockchain technology for genomic data sharing [32, 9, 17, 65, 75]. Addressing these challenges ensures responsible advancement of scRNA-seq integration with traditional Chinese medicine and genomic technologies, making personalized medicine effective and ethically sound.

7.4 Innovative Approaches and Future Directions

Integrating single-cell RNA sequencing with Chinese medicine dialectics and genomics in personalized medicine offers numerous innovative approaches and future research directions to enhance understanding of complex biological systems. Improving single-cell sequencing methodologies, developing standardized data analysis protocols, and exploring genetic variation implications on drug response and disease susceptibility are key focus areas. Refining single-cell isolation techniques and integrating multi-omics approaches are crucial for advancing cellular function and disease state understanding [18].

Future research should enhance computational models to incorporate temporal dynamics, improve interpretability, and leverage prior biological knowledge for better causal inference in single-cell genomics. Addressing scalability and latent confounding factors unrelated to measurement error is vital for improving causal analysis. Developing algorithms capable of working with limited data and tuning existing models for specific applications remains critical [24].

Emerging trends suggest focusing on integrating single-cell data, enhancing statistical analysis methods, and utilizing gene-editing technologies to validate candidate genes identified through genome-wide association studies (GWAS) [27]. Exploring modular Bayesian approaches, improving scalability and robustness in larger datasets, applying advanced learning methods to diverse feature spaces, and investigating more cancer subtypes could significantly enhance predictive accuracy and treatment outcomes [21].

Future research should develop broader metrics and benchmarks encompassing various scRNA-seq methods and applications. This expansion will improve analysis robustness and enhance scRNA-seq applicability in diverse contexts, such as identifying rare cell populations, understanding gene regulatory networks, and tracking cell lineage trajectories. Systematically evaluating different experimental protocols and computational pipelines establishes best practices for accurate and reliable insights into cellular heterogeneity and complex biological systems [18, 52, 55, 51, 56]. Collectively, these innovative approaches and future research directions hold the potential to significantly advance personalized medicine. By leveraging cutting-edge technologies and methodologies, researchers can develop precise and effective treatment strategies tailored to individual patient profiles, ultimately improving healthcare outcomes.

8 Conclusion

The fusion of single-cell RNA sequencing with Chinese medicine dialectics represents a pivotal step in advancing personalized disease treatment by harmonizing modern genomic technologies with age-old medical paradigms. This integration facilitates an enhanced understanding of cellular diversity and gene expression, which is essential for unraveling intricate disease mechanisms and devising precise therapeutic interventions. Cutting-edge computational methodologies, such as scFusionTTT and Poisson mash models, significantly improve the analysis of multi-omics data, uncovering new perspectives on disease causatives and cellular interactions. The CCP method advances the accuracy of visualization techniques like UMAP and t-SNE, offering more robust tools for analyzing scRNA-seq data.

The infusion of artificial intelligence and machine learning into genomic analysis amplifies predictive precision, aligning with the holistic approaches inherent in traditional Chinese medicine. This synergy is evident in the personalized treatment frameworks for conditions like Helicobacter pylori

infection and novel coronavirus pneumonia, where genomic insights are seamlessly integrated with traditional principles. Ethical considerations, particularly concerning data privacy, remain a critical focus. Future research should prioritize the development of sophisticated cryptographic solutions and explore innovative blockchain and decentralized trust frameworks to secure genomic data sharing.

The collective advancements in integrating single-cell RNA sequencing with Chinese medicine dialectics highlight the transformative potential of this interdisciplinary approach. By merging the precision of genomic technologies with the holistic nature of traditional Chinese medicine, this collaborative effort promises to deliver more tailored, effective healthcare solutions, thereby enhancing patient outcomes and driving the evolution of personalized medicine.

References

[1] AJ Venkatakrishnan, Arjun Puranik, Akash Anand, David Zemmour, Xiang Yao, Xiaoying Wu, Ramakrishna Chilaka, Dariusz K. Murakowski, Kristopher Standish, Bharathwaj Raghunathan, Tyler Wagner, Enrique Garcia-Rivera, Hugo Solomon, Abhinav Garg, Rakesh Barve, Anuli Anyanwu-Ofili, Najat Khan, and Venky Soundararajan. Knowledge synthesis from 100 million biomedical documents augments the deep expression profiling of coronavirus receptors, 2020.

[2] Monika Gisler, Didier Sornette, and Ryan Woodard. Exuberant innovation: The human genome project, 2010.

[3] Anthony Letai. Functional precision cancer medicine—moving beyond pure genomics. *Nature medicine*, 23(9):1028–1035, 2017.

[4] Monique G. P. van der Wijst, Dylan H. de Vries, Hilde E. Groot, Gosia Trynka, Chung-Chau Hon, Martijn C. Nawijn, Youssef Idaghdour, Pim van der Harst, Chun J. Ye, Joseph Powell, Fabian J. Theis, Ahmed Mahfouz, Matthias Heinig, and Lude Franke. Single-cell eqtlgen consortium: a personalized understanding of disease, 2019.

[5] Anastasia Krithara, Fotis Aisopos, Vassiliki Rentoumi, Anastasios Nentidis, Konstantinos Bougatiotis, Maria-Esther Vidal, Ernestina Menasalvas, Alejandro Rodriguez-Gonzalez, Eleftherios G. Samaras, Peter Garrard, Maria Torrente, Mariano Provencio Pulla, Nikos Dimakopoulos, Rui Mauricio, Jordi Rambla De Argila, Gian Gaetano Tartaglia, and George Paliouras. iasis: Towards heterogeneous big data analysis for personalized medicine, 2024.

[6] Frederik F. Flöther. Early quantum computing applications on the path towards precision medicine, 2024.

[7] Ahmad Hussein, Mukesh Prasad, and Ali Braytee. Explainable ai methods for multi-omics analysis: A survey, 2024.

[8] David S. Watson. Interpretable machine learning for genomics, 2021.

[9] Kishi Kobe Yee Francisco, Andrane Estelle Carnicer Apuhin, Myles Joshua Toledo Tan, Mickael Cavanaugh Byers, Nicholle Mae Amor Tan Maravilla, Hezerul Abdul Karim, and Nouar AlDahoul. Can personalized medicine coexist with health equity? examining the cost barrier and ethical implications, 2024.

[10] Daphne O. Martschenko, Hannah Wand, Jennifer L. Young, and Genevieve L. Wojcik. Mind the gap: how multiracial individuals get left behind when we talk about race, ethnicity, and ancestry in genomic research, 2022.

[11] Yan Duan, Yu-Ting Su, Jie Ren, Qun Zhou, Min Tang, Juan Li, and Shun-Xiang Li. Kidney tonifying traditional chinese medicine: potential implications for the prevention and treatment of osteoporosis. *Frontiers in pharmacology*, 13:1063899, 2023.

[12] Ling Li, Lele Yang, Liuqing Yang, Chunrong He, Yuxin He, Liping Chen, Qin Dong, Huaiying Zhang, Shiyun Chen, and Peng Li. Network pharmacology: a bright guiding light on the way to explore the personalized precise medication of traditional chinese medicine. *Chinese medicine*, 18(1):146, 2023.

[13] DU Hong-Zhi, HOU Xiao-Ying, MIAO Yu-Huan, Bi-Sheng Huang, and LIU Da-Hui. Traditional chinese medicine: an effective treatment for 2019 novel coronavirus pneumonia (ncp). *Chinese journal of natural medicines*, 18(3):206–210, 2020.

[14] Ru-Jia Li, Yuan-Yuan Dai, Chun Qin, Gan-Rong Huang, Yan-Chun Qin, Yong-Yi Huang, Zan-Song Huang, Xian-Ke Luo, and Yan-Qiang Huang. Application of traditional chinese medicine in treatment of helicobacter pylori infection. *World Journal of Clinical Cases*, 9(35):10781, 2021.

[15] Kuo Yang, Zecong Yu, Xin Su, Xiong He, Ning Wang, Qiguang Zheng, Feidie Yu, Zhuang Liu, Tiancai Wen, and Xuezhong Zhou. A optimization framework for herbal prescription planning based on deep reinforcement learning, 2023.

[16] Rupam Bhattacharyya, Nicholas Henderson, and Veerabhadran Baladandayuthapani. Functional integrative bayesian analysis of high-dimensional multiplatform genomic data, 2022.

[17] William Briguglio, Parisa Moghaddam, Waleed A. Yousef, Issa Traore, and Mohammad Mamun. Machine learning in precision medicine to preserve privacy via encryption, 2021.

[18] Byungjin Hwang, Ji Hyun Lee, and Duhee Bang. Single-cell rna sequencing technologies and bioinformatics pipelines. *Experimental & molecular medicine*, 50(8):1–14, 2018.

[19] Valentine Svensson, Roser Vento-Tormo, and Sarah A Teichmann. Exponential scaling of single-cell rna-seq in the last decade, 2017.

[20] Muhammad Umar, Muhammad Asif, and Arif Mahmood. Lower-dimensional projections of cellular expression improves cell type classification from single-cell rna sequencing, 2024.

[21] Jiawei Long and Yu Xia. Cluster analysis of high-dimensional scrna sequencing data, 2019.

[22] David T Paik, Sangkyun Cho, Lei Tian, Howard Y Chang, and Joseph C Wu. Single-cell rna sequencing in cardiovascular development, disease and medicine. *Nature Reviews Cardiology*, 17(8):457–473, 2020.

[23] Siamak Zamani Dadaneh, Paul de Figueiredo, Sing-Hoi Sze, Mingyuan Zhou, and Xiaoning Qian. Bayesian gamma-negative binomial modeling of single-cell rna sequencing data, 2019.

[24] Erik J. Amézquita, Farzana Nasrin, Kathleen M. Storey, and Masato Yoshizawa. Genomics data analysis via spectral shape and topology, 2022.

[25] Moeez M. Subhani and Ashiq Anjum. Multiclass disease predictions based on integrated clinical and genomics datasets, 2020.

[26] Martin Jinye Zhang, Vasilis Ntranos, and David Tse. Determining sequencing depth in a single-cell rna-seq experiment. *Nature communications*, 11(1):774, 2020.

[27] Eddie Cano-Gamez and Gosia Trynka. From gwas to function: using functional genomics to identify the mechanisms underlying complex diseases. *Frontiers in genetics*, 11:424, 2020.

[28] Enzo Battistella, Maria Vakalopoulou, Roger Sun, Théo Estienne, Marvin Lerousseau, Sergey Nikolaev, Emilie Alvarez Andres, Alexandre Carré, Stéphane Niyoteka, Charlotte Robert, Nikos Paragios, and Eric Deutsch. Cancer gene profiling through unsupervised discovery, 2021.

[29] Allen W Zhang and Kieran R Campbell. Computational modelling in single-cell cancer genomics: methods and future directions, 2020.

[30] Limor Gultchin, David S. Watson, Matt J. Kusner, and Ricardo Silva. Operationalizing complex causes: A pragmatic view of mediation, 2021.

[31] Kexin Huang, Cao Xiao, Lucas M. Glass, Cathy W. Critchlow, Greg Gibson, and Jimeng Sun. Machine learning applications for therapeutic tasks with genomics data, 2021.

[32] Chloé-Agathe Azencott. Machine learning and genomics: precision medicine vs. patient privacy, 2018.

[33] Julia Zeiltinger, Sushmita Roy, Ferhat Ay, Anthony Mathelier, Alejandra Medina-Rivera, Shaun Mahony, Saurabh Sinha, and Jason Ernst. Perspective on recent developments and challenges in regulatory and systems genomics, 2024.

[34] Yongjian Yang. Deciphering cell systems: Machine learning perspectives and approaches for the analysis of single-cell data, 2024.

[35] Neal G. Ravindra, Arijit Sehanobish, Jenna L. Pappalardo, David A. Hafler, and David van Dijk. Disease state prediction from single-cell data using graph attention networks, 2020.

[36] Safoora Yousefi, Amirreza Shaban, Mohamed Amgad, Ramraj Chandradevan, and Lee A. D. Cooper. Learning clinical outcomes from heterogeneous genomic data sources, 2019.

17

[37] Ghizal fatima, Risala H. Allami, and Maitham G. Yousif. Integrative ai-driven strategies for advancing precision medicine in infectious diseases and beyond: A novel multidisciplinary approach, 2023.

[38] Yixuan Wang, Shuangyin Li, Shimin DI, and Lei Chen. Single-cell rna-seq synthesis with latent diffusion model, 2023.

[39] Gökcen Eraslan, Žiga Avsec, Julien Gagneur, and Fabian J Theis. Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics*, 20(7):389–403, 2019.

[40] Xinlei Mi, William Bekerman, Peter A. Sims, Peter D. Canoll, and Jianhua Hu. Rzimm-scrna: A regularized zero-inflated mixture model framework for single-cell rna-seq data, 2021.

[41] Linfeng Jiang and Yuan Zhu. Data smoothing filling method based on scrna-seq data zero-value identification, 2024.

[42] Dian Meng, Bohao Xing, Xinlei Huang, Yanran Liu, Yijun Zhou, Yongjun xiao, Zitong Yu, and Xubin Zheng. scfusionttt: Single-cell transcriptomics and proteomics fusion with test-time training layers, 2024.

[43] Lana X Garmire. Strategies to integrate multi-omics data for patient survival prediction, 2020.

[44] Bryan He, Matthew Thomson, Meena Subramaniam, Richard Perez, Chun Jimmie Ye, and James Zou. Cloudpred: Predicting patient phenotypes from single-cell rna-seq, 2021.

[45] Yihang Xiao, Jinyi Liu, Yan Zheng, Xiaohan Xie, Jianye Hao, Mingzhi Li, Ruitao Wang, Fei Ni, Yuxiao Li, Jintian Luo, Shaoqing Jiao, and Jiajie Peng. Cellagent: An llm-driven multi-agent framework for automated single-cell data analysis, 2024.

[46] Eric V. Strobl. Root causal inference from single cell rna sequencing with the negative binomial, 2023.

[47] Yuta Hozumi and Gu-Wei Wei. Analyzing scrna-seq data by ccp-assisted umap and t-sne, 2023.

[48] Jing Liang, Zhuo Hu, Zong-Wei Li, Kang-Jia Qiao, and Wei-Feng Guo. Multi-objective optimization based network control principles for identifying personalized drug targets with cancer, 2023.

[49] Marc Subirana-Granés, Jill Hoffman, Haoyu Zhang, Christina Akirtava, Sutanu Nandi, Kevin Fotso, and Milton Pividori. Genetic studies through the lens of gene networks, 2024.

[50] Akhila Krishna, Ravi Kant Gupta, Pranav Jeevan, and Amit Sethi. Advancing gene selection in oncology: A fusion of deep learning and sparsity for precision gene selection, 2024.

[51] Wenjin Ye, Yuanchen Ma, Junkai Xiang, Hongjie Liang, Tao Wang, Qiuling Xiang, Andy Peng Xiang, Wu Song, Weiqiang Li, and Weijun Huang. Objectively evaluating the reliability of cell type annotation using llm-based strategies, 2024.

[52] Beate Vieth, Swati Parekh, Christoph Ziegenhain, Wolfgang Enard, and Ines Hellmann. A systematic evaluation of single cell rna-seq analysis pipelines. *Nature communications*, 10(1):4667, 2019.

[53] Tianwei Yue, Yuanxin Wang, Longxiang Zhang, Chunming Gu, Haoru Xue, Wenping Wang, Qi Lyu, and Yujie Dun. Deep learning for genomics: A concise overview, 2023.

[54] Yusha Liu, Peter Carbonetto, Michihiro Takahama, Adam Gruenbaum, Dongyue Xie, Nicolas Chevrier, and Matthew Stephens. A flexible model for correlated count data, with application to multi-condition differential expression analyses of single-cell rna sequencing data, 2024.

[55] Megha Patel, Nimish Magre, Himanshi Motwani, and Nik Bear Brown. Advances in machine learning, statistical methods, and ai for single-cell rna annotation using raw count matrices in scrna-seq data, 2024.

[56] Matthew Brendel, Chang Su, Zilong Bai, Hao Zhang, Olivier Elemento, and Fei Wang. Application of deep learning on single-cell rna-sequencing data analysis: A review, 2022.

[57] Xiaowen Cao, Li Xing, Elham Majd, Hua He, Junhua Gu, and Xuekui Zhang. A systematic evaluation of methods for cell phenotype classification using single-cell rna sequencing data, 2021.

[58] Akira Nguyen, Weng Hua Khoo, Imogen Moran, Peter I Croucher, and Tri Giang Phan. Single cell rna sequencing of rare immune cell populations. *Frontiers in immunology*, 9:1553, 2018.

[59] Daniel Merl, Julia Ling-Yu Chen, Jen-Tsan Chi, and Mike West. An integrative analysis of cancer gene expression studies using bayesian latent factor modeling, 2010.

[60] Dinesh Samuel Sathia Raj, Vijayakumar V, Bharat Rawal, and Longzhi Yang. Analysis of big data technology for health care services, 2019.

[61] Jiaying Zhou, Mingzhou Jiang, Junde Wu, Jiayuan Zhu, Ziyue Wang, and Yueming Jin. Mgi: Multimodal contrastive pre-training of genomic and medical imaging, 2024.

[62] Davide Maspero, Claudio Isella, Marzia Di Filippo, Alex Graudenzi, Sara Erika Bellomo, Marco Antoniotti, Giancarlo Mauri, Enzo Medico, and Chiara Damiani. Metabolic enrichment through functional gene rules, 2017.

[63] Umar Ahmad, Buhari Ibrahim, Mustapha Mohammed, Ahmed Faris Aldoghachi, Mahmood Usman, Abdulbasit Haliru Yakubu, Abubakar Sadiq Tanko, Khadijat Abubakar Bobbo, Usman Adamu Garkuwa, Abdullahi Adamu Faggo, Sagir Mustapha, Mahmoud Al-Masaeed, Syahril Abdullah, Yong Yoke Keong, and Abhi Veerakumarasivam. Transcriptome profiling research in urothelial cell carcinoma, 2021.

[64] Jingxiang Chen, Haoda Fu, Xuanyao He, Michael R. Kosorok, and Yufeng Liu. Estimating individualized treatment rules for ordinal treatments, 2017.

[65] Mohammed Alghazwi, Fatih Turkmen, Joeri van der Velde, and Dimka Karastoyanova. Blockchain for genomics: A systematic literature review, 2022.

[66] Erman Ayday, Emiliano De Cristofaro, Jean-Pierre Hubaux, and Gene Tsudik. The chills and thrills of whole genome sequencing, 2015.

[67] Aaron Golden, S. George Djorgovski, and John M. Greally. Astrogenomics: big data, old problems, old solutions?, 2013.

[68] Jiajia Liu, Mengyuan Yang, Yankai Yu, Haixia Xu, Kang Li, and Xiaobo Zhou. Large language models in bioinformatics: applications and perspectives, 2024.

[69] Cecily Wolfe, Yayi Feng, David Chen, Edwin Purcell, Anne Talkington, Sepideh Dolatshahi, and Heman Shakeri. Geotyper: Automated pipeline from raw scrna-seq data to cell type identification, 2022.

[70] Bing He, Yao Xiao, Haodong Liang, Qianhui Huang, Yuheng Du, Yijun Li, David Garmire, Duxin Sun, and Lana X. Garmire. Asgard: A single-cell guided pipeline to aid repurposing of drugs, 2022.

[71] Manlio De Domenico, Luca Allegri, Guido Caldarelli, Valeria d'Andrea, Barbara Di Camillo, Luis M. Rocha, Jordan Rozum, Riccardo Sbarbati, and Francesco Zambelli. Challenges and opportunities for digital twins in precision medicine: a complex systems perspective, 2024.

[72] Yoshitaka Inoue. scvgae: A novel approach using zinb-based variational graph autoencoder for single-cell rna-seq imputation, 2024.

[73] Laura-Jayne Gardiner, Anna Paola Carrieri, Jenny Wilshaw, Stephen Checkley, Edward O Pyzer-Knapp, and Ritesh Krishna. Combining human cell line transcriptome analysis and bayesian inference to build trustworthy machine learning models for prediction of animal toxicity in drug development, 2019.

[74] Laurent Bréhélin. Advancing regulatory genomics with machine learning, 2023.

[75] Muhammad Naveed, Erman Ayday, Ellen W. Clayton, Jacques Fellay, Carl A. Gunter, Jean-Pierre Hubaux, Bradley A. Malin, and XiaoFeng Wang. Privacy in the genomic era, 2015.

19

[76] Alexander Rakowski, Remo Monti, Viktoriia Huryn, Marta Lemanczyk, Uwe Ohler, and Christoph Lippert. Metadata-guided feature disentanglement for functional genomics, 2024.

[77] Adriana Romero, Pierre Luc Carrier, Akram Erraqabi, Tristan Sylvain, Alex Auvolat, Etienne Dejoie, Marc-André Legault, Marie-Pierre Dubé, Julie G. Hussin, and Yoshua Bengio. Diet networks: Thin parameters for fat genomics, 2017.

[78] Nagur Shareef Shaik, Teja Krishna Cherukuri, Vince D. Calhoun, and Dong Hye Ye. Multimodal imaging genomics transformer: Attentive integration of imaging with genomic biomarkers for schizophrenia classification, 2024.

[79] Yuhan Chen and David B. Dunson. Modular bayes screening for high-dimensional predictors, 2017.

[80] Stefan Stanojevic, Yijun Li, and Lana X. Garmire. Computational methods for single-cell multi-omics integration and alignment, 2022.

[81] Elsa Lawrence, Adham El-Shazly, Srijit Seal, Chaitanya K Joshi, Pietro Liò, Shantanu Singh, Andreas Bender, Pietro Sormanni, and Matthew Greenig. Understanding biology in the age of artificial intelligence, 2024.

[82] Changgee Chang, Zongyu Dai, Jihwan Oh, and Qi Long. Integrative learning of structured high-dimensional data from multiple datasets, 2022.

[83] Aiman Farooq, Deepak Mishra, and Santanu Chaudhury. Translating imaging to genomics: Leveraging transformers for predictive modeling, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.