Seminars in Immunopathology (2023) 45:61–69
https://doi.org/10.1007/s00281-022-00981-1

REVIEW

Single-cell technologies uncover intra-tumor heterogeneity in childhood cancers

Yu-Chen Lo[1] · Yuxuan Liu[1] · Marte Kammersgaard[1] · Abhishek Koladiya[1] · Timothy J. Keyes[1,2] · Kara L. Davis[1]

Received: 10 August 2022 / Accepted: 11 December 2022 / Published online: 10 January 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract
Childhood cancer is the second leading cause of death in children aged 1 to 14. Although survival rates have vastly improved over the past 40 years, cancer resistance and relapse remain a significant challenge. Advances in single-cell technologies enable dissection of tumors to unprecedented resolution. This facilitates unraveling the heterogeneity of childhood cancers to identify cell subtypes that are prone to treatment resistance. The rapid accumulation of single-cell data from different modalities necessitates the development of novel computational approaches for processing, visualizing, and analyzing single-cell data. Here, we review single-cell approaches utilized or under development in the context of childhood cancers. We review computational methods for analyzing single-cell data and discuss best practices for their application. Finally, we review the impact of several studies of childhood tumors analyzed with these approaches and future directions to implement single-cell studies into translational cancer research in pediatric oncology.

Keywords Childhood cancer · Single-cell technology · Machine learning · Mass cytometry · Cancer resistance · Pediatrics

Introduction

Cancer is the second leading cause of death in children aged 1 to 14 [1]; thus, there remains an urgent and ongoing need to study these dreadful diseases of childhood. Of the many subtypes of childhood cancer, leukemia is the most common, followed by tumors of the central nervous system (CNS). Altogether, 10,500 children are diagnosed with cancer each year in the USA [1]. Although 5-year event-free survival rates for childhood leukemia and solid tumors have markedly improved, treatment-refractory disease and relapse remain a significant clinical challenge. Nearly all tumors in children will respond to initial therapies and go into remission, but recurrence is significantly more difficult to treat and cure. For instance, childhood B-cell acute lymphoblastic leukemia (B-ALL), the most common cancer in children, is highly curable, and over 80% of children will maintain a durable remission and go on to be long-term survivors [2]. However, 15% of these patients will relapse, and for these children, only 50% will maintain a durable remission. Genetic studies in pediatric leukemia identified recurrent genetic alterations that are prognostic to treatment outcomes [3]. These studies suggest a high degree of heterogeneity between subclones within each tumor and highlight the importance of understanding tumor heterogeneity in the diagnosis and treatment of childhood cancers [4, 5]. The need to understand the genetic, phenotypic, and functional heterogeneity of each tumor is critical to learning routes to resistance and identifying therapeutic targets.

The development of single-cell technologies has enabled better understanding of tumor heterogeneity and has revealed potential mechanisms of relapse and treatment resistance in cancers, including childhood cancer [6]. Individual tumor cells can be evaluated for genetic, epigenetic, transcriptomic, proteomic, and metabolomic features (Fig. 1). Further, many of these approaches can be combined for multi-omics measurements on individual

Yu-Chen Lo and Yuxuan Liu contributed equally.

This article is a contribution to the special issue on: Single-cell and spatial multi-omics in clinical outcomes studies - Guest Editor: Brice Gaudillière

✉ Kara L. Davis
kardavis@stanford.edu

[1] Department of Pediatrics, Division of Hematology, Oncology, Stem Cell Transplant and Regenerative Medicine, Stanford University, Stanford, CA, USA

[2] Medical Scientist Training Program, Stanford University School of Medicine, Stanford, CA, USA

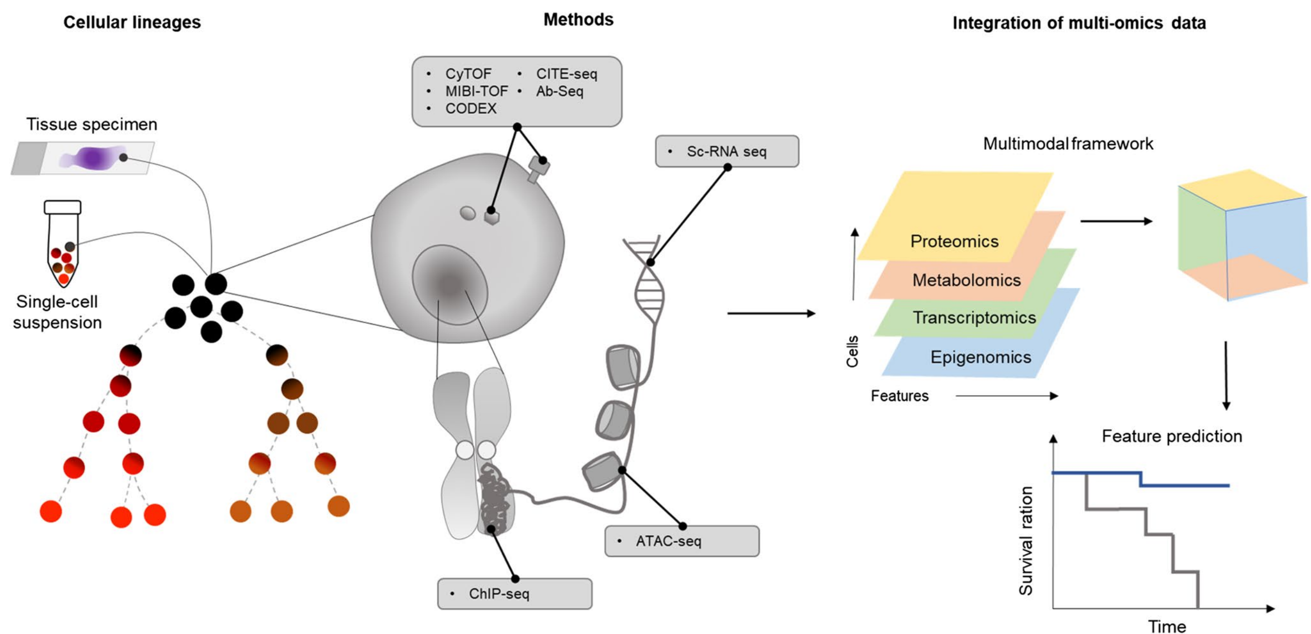

Fig. 1 Overview of single-cell technology. Various single-cell methods have been proposed to understand cellular heterogeneity and lineage identification. These methods are applicable to intact tissue specimens and dissociated cells. These methods can be divided into two categories: (1) measures only one layer of omics, e.g., CyTOF (cytometry by time-of-flight) measures proteins and sc-RNA seq measures mRNA; (2) more than one omics measurement, e.g., Ab-seq, CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) both proteins and mRNA from one cell simultaneously. Finally, data integration algorithms can be applied to these multi-omics datasets in order to predict cellular features correlating with health and disease conditions

cells. Different modalities confer different degrees of resolution, advantages, and limitations that must be appreciated in order to appropriately evaluate the resulting data. At the same time, single-cell data generates high-dimensional signals in large volumes that demands new computational approaches for data analysis, management, and integration.

Despite these challenges, the new paradigm of understanding cancer heterogeneity at the single-cell level has broad implications for all types of childhood cancer. These approaches can inform how future childhood cancers should be diagnosed, managed, and treated. Here, we will review common single-cell technologies, their applications, their limitations, and examples of their applications to childhood cancer. We will discuss approaches to high-dimensional single-cell data quality control and analysis with suggestions for best practices. Finally, we will discuss current applications of single-cell approaches to childhood cancers and suggest future directions to guide the development of this emerging field.

Single-cell platforms in childhood cancer

High-dimensional cytometry: CyTOF (cytometry by time-of-flight) and spectral flow cytometry

The development of single-cell proteomics technologies has enabled the identification of cell surface markers or intracellular proteins amenable for therapeutic intervention or diagnostic applications. CyTOF, or cytometry by time-of-flight, is now an established approach that can measure over 40 protein targets simultaneously to resolve cell identity and function [7, 8]. By conjugating antibodies to stable metal isotopes rather than to fluorophores, CyTOF enables the measurement of an increased number of cellular proteins with minimal overlap. CyTOF has been broadly applied to single-cell phenotyping of suspension cells including peripheral blood immune cells and leukemia, but methods have also been developed to apply this approach to the study of solid tumors [9]. More recently, the development of spectral flow cytometry (SFC) which measures the entire fluorescence spectrum of fluorophores across multiple lasers has demonstrated its ability to detect over 40 proteins per cells [9, 10]. SFC also has been applied to single-cell phenotyping of immune and leukemia cells [10]. A comparative study of CyTOF and SFC suggested that both methods

yield highly comparable results using manual and high-dimensional data analysis appraoches [11]. CyTOF has a relatively higher throughput and is less expensive compared to other single-cell proteomic techniques. Limitations of CyTOF include slower acquisition time compared to flow cytometry, the need for clean sample preparation, and that the cells cannot be preserved after analysis [12]. Our group demonstrated the use of CyTOF in a cohort of bone marrow (BM) samples from children with B-cell lymphoblastic leukemia (B-ALL) obtained at the time of diagnosis using CyTOF. The developmental state of each leukemia cell was determined by matching the protein expression to the most similar healthy B-cell population from control BM resulting in a developmental identifier for each leukemia cell. Using a machine learning model, we identified a subset of pre-B cells with activation of preBCR signaling molecules to be highly predictive of future relapse. Our model was 32% more accurate to predict relapse risk than the clinical gold standard predictor—minimal residual disease (MRD) risk determination, indicating the identified signatures from specific subpopulations improved the relapse predictive power [11, 13]. Although this approach has not yet been translated clinically, work is ongoing to do so.

Single-cell tissue imaging: MIBI-TOF, IMC, CODEX

Several spatial proteomics technologies, including multiplexed ion beam imaging by time-of-flight (MIBI-TOF), imaging mass cytometry (IMC), and co-detection by indexing (CODEX), have been recently developed to allow a high number of protein targets to be simultaneously imaged on a formalin-fixed, paraffin-embedded (FFPE) tissue specimens [14–19]. These tissue imaging methods enable single-cell analysis of protein expression and deliver spatial information to allow tumor cells to be analyzed in their environment.

Both MIBI-TOF and imaging mass cytometry (IMC) are related to immunohistochemistry (IHC) and utilize metal-conjugated antibodies to detect simultaneous expression of up to 40 proteins in intact tissues [14, 15, 18, 19]. MIBI-TOF requires specialized equipment yet is comparatively quick to run and has higher sensitivity and resolution (~ 260 nm vs ~ 1000 nm) than imaging mass cytometry (IMC), though higher resolution can result in longer acquisition time [12]. MIBI was introduced by Angelo and colleagues in 2014, who demonstrated its utility to potentially detect 100 targets on breast tumor samples, later on identifying varying tumor-immune compositions across patient samples correlated with overall survival [14, 18]. MIBI has since been applied to the study of brain tissue, the maternal–fetal interface, tuberculosis granulomas, and others [18, 20, 21]. Our lab is using MIBI-TOF to understand divergent cellular states in neuroblastoma, the most common non-CNS solid tumor of childhood. By optimizing a 40-antibody panel to capture adrenergic or mesenchymal identity, infiltrating immune cells, tissue architecture, and known neuroblastoma proteins, we hope to demonstrate how divergent cell states relate to each other, to immune cells, and to clinical outcomes [22].

IMC has been applied to profile sarcoma cell lines and patient-derived xenograft models of sarcomas, including osteosarcoma, a common sarcoma of childhood. Further work utilized IMC to analyze circulating tumor cells (CTC) in peripheral blood from patients with metastatic sarcoma and metastatic prostate cancer identifying dozens of protein biomarkers on these CTCs, demonstrating the capability to measure rare cell populations [23, 24].

CODEX is a fluorescence microscopy-based method that uses oligonucleotide-conjugated antibodies and specialized fluorescent probes to detect up to 60 markers in one tissue Sect. [17]. Although CODEX has yet to be applied to pediatric tumors, it has demonstrated applicability to broad cancer and tissue types [25]. Furthermore, the availability of fluorescent microscopes, the automation of staining and slide-scanning procedures, the compatibility with microfluidics devices, and the use of hybridization techniques shared with other multi-omic approaches make the method promising for spatial multi-omics [16, 26].

Current challenges to utilizing these single-cell spatial proteomics imaging methods are the high cost of equipment, lack of widespread accessibility, and the potentially lengthy and laborious process of panel development and validation. Further, differences in data extraction and post-processing between research groups challenge the consistency and comparability of results [27]. Despite these challenges, obtaining intact single-cell information from intact tumor cells and their surrounding microenvironment presents an exciting opportunity to learn more about solid tumors of childhood.

Single-cell genomic sequencing

Several single-cell genomic sequencing approaches can be used to interrogate genomic heterogeneity within a given tumor and to determine the clonal structures of tumors. Gawad et al. performed single-cell-targeted DNA sequencing in pediatric B-ALL patient samples to identify co-dominant clones present in most patients and late mutation events not required for tumor development [28]. Despite the promise of single-cell genomic sequencing, its use is sparse owing to the newness of the technology. Currently, single-cell RNA sequencing (scRNA-seq) is one of the most commonly used single-cell techniques for understanding intratumor heterogeneity in childhood cancer studies having been employed in studies of ALL, juvenile myelomonocytic leukemia, neuroblastoma, medulloblastoma, cerebellar tumors, and ependymoma [29–34]. An advantage of scRNA-seq over the single-cell proteomics approaches is that thousands of RNA transcripts are simultaneously measured to identify

both transient and stable cell states in cell lineage development and differentiation. Transcriptomic information from single cells can help identify dysregulated oncogenic pathways and therapeutic targets. Medulloblastoma (MB), a childhood brain tumor, comprises four subgroups (*WNT*, *SHH*, group 3, and group 4) based on genomic studies at the bulk level. Each subgroup was characterized by different genomic landscapes, patient demographics, and clinical phenotypes, suggesting that MB subgroups may arise from different cell types in precise developmental trajectories. To investigate the genotype-to-cell-type association, Hovestadt and colleagues applied single-cell transcriptomics in 25 medulloblastoma samples including all four subgroups. Integrating bulk-level whole genome and exome sequencing data with scRNA-seq data, they classified each cell into malignant or non-malignant subsets. Leveraging scRNA-seq data for mouse cerebellar development spanning 13 embryonic and early postnatal time points, they correlated each MB cell to normal developmental hierarchies. The single-cell approaches enabled the identification of a fraction of intermediate tumors characterized by various proportions of both undifferentiated and more differentiated neuron cell populations, which was hampered by prior bulk analysis [35].

In neuroblastoma tumors, Jansky et al. performed scRNA-seq on tumor samples along with developing human adrenal glands at various stages of embryonic and fetal development. After defining the normal differentiation map of adrenal tissue, they identified the tumor progenitor cells by projecting each tumor cell onto these developmental trajectories. Malignant cells were defined with lacking expression of markers for normal infiltrating cell types and genetic alterations at the single-cell level. Molecular features of single neuroblastoma cells were identified by comparing their closest matching normal cell types. With the single-cell data in hand, they used deconvolution of bulk RNA-seq data from 498 neuroblastoma samples and identified cellular compositions associated with clinical outcome. High proportions of mature neuroblasts and low proportions of cycling neuroblasts indicated poor outcomes, while the conventional prognostic marker, MYCN alone, was not risk-predictive [36].

To better understand transcriptional regulation, epigenomic profiles can add additional information to complement scRNA-seq data, including assay for transposase-accessible chromatin using sequencing (ATAC-seq) to measure the chromatin accessibility and ChIP-seq. Corces et al. developed a fast ATAC-seq protocol, optimized for human blood cells that allow for rapid high-quality measurements to profile the chromatin landscapes in primary blood cells [37]. Ranzoni et al. utilize scRNA-seq and scATAC-seq to assess the transcriptional identity and chromatin accessibility in several immunophenotypic blood cell populations from human fetal bone marrow (FBM) and liver (FL). They identified three oligopotent progenitor populations differentiated from hematopoietic stem cell/multi-potent progenitor cells (HSC/MPPs). They show that transcriptional profiles alone do not reveal significant lineage priming in stem cells, while this is apparent when one examines the chromatin. Their work refined the sorting strategy for HSC/MPPs and provided a useful framework for future study in human hematopoiesis.

Single-cell multimodal profiling

High-throughput single-cell sequencing can be used to study single-cell level transcriptomics but does not provide phenotypic information. Recently, single-cell multimodal methods like CITE-seq and Ab-Seq integrate cellular protein and transcriptome measurements into a single-cell readout [38–40]. Using oligonucleotide-conjugated antibodies which contain a barcode for antibody identification and a handle for PCR amplification, these techniques simultaneously quantify RNA and surface protein abundance at a single-cell level for any population of interest [41]. Recently, Bai et al. applied CITE-seq to analyze the chimeric antigen receptor (CAR) T-cells in infusion products from 12 B-ALL patients including 5 with long-term remission (CR), 5 with CD19-positive relapse (RL), and 2 non-responders (NR) to identify the determinants of CD19-positive relapse. The infusion products were analyzed upon the CD19-specific stimulation by CD19 + antigen presenting cells (APCs) along with other stimulation conditions (TCR-mediated stimulation, non-target APC stimulation, and unstimulated controls). Focusing on the CD19-specific stimulated CAR + T-cells only, distinct cytokine co-expression modules were identified to reveal the heterogeneous immune functions of different CAR T-cell subsets. Immune pathway analysis showed that T helper 2 (T_H2) function-related pathways, genes, and upstream regulators are significantly downregulated in CAR T-cells upon CAR-specific stimulation from RL compared to CR patients, whereas other major immune programs remained unchanged [42]. Further validation in infusion products from 49 B-ALL patients confirmed the deficiency of T_H2 function in RL patients. To study the cellular compositions of memory T-cells, phenotypic proteomic profiles based on CITE-seq data revealed lower abundance of early memory T-cells in RL compared to CR patients. Integrating these major features that distinguish RL from CR patients, they developed a machine learning model to achieve a sensitivity of ~ 70% to predict long-term remission. This study fully showcased the potential of utilizing multi-omics techniques in improving the predictive power for relapse. In addition, other omics techniques that measure genome, transcriptome, and epigenome simultaneously have been developed [43], but have not applied in pediatric tumors yet. We anticipate that the continued maturation of these multi-omic techniques will impact the field of pediatric oncology.

Single-cell metabolic profiling

Metabolomics is one emerging branch of "omics" analysis to explore the biochemical processes of cancer cells to produce and consume energy. Using mass spectrometry, single-cell metabolomics measures metabolites in single cells in different states, better characterizes metabolites in specific cell types, and potentially identifies treatment targets. Single-cell metabolite extraction is the first critical step, and the technique to accomplish this has been optimized over the past decades. Mizuno et al. is the first group to establish live single-cell video-mass spectrometry to introduce the contents of rat basophil leukemia cells into a quadrupole-time-of-flight mass spectrometry (Q-TOF–MS) using a nano-electrospray ionization (nano-ESI) tip [44]. Based on this work, Pan et al. subsequently developed a single probe that can be inserted into single cells to extract intracellular compounds [45]. Multiple cellular metabolites in a single cell can be analyzed in real time by coupling the probe with a mass spectrometer.

Besides mass spectrometry, antibody-based and imaging techniques are other emerging tools to study cellular metabolism in single cells. The Connolly group utilized an antibody-based method to measure cellular metabolic profiles of human peripheral blood mononuclear cells and compared abundances of metabolic transporters before and after glycolytic inhibition [46]. Arguello et al. developed a flow cytometry-based method (SCENITH) to monitor ATP synthesis using puromycin as a reporter. Different metabolic perturbations enabled the study of metabolic responses and revealed variable metabolic profiles in multiple cell types [47]. Sharick et al. developed optical metabolic imaging (OMI), a novel, high-resolution fluorescence microscopy technique that quantifies the optical redox ratio, defined as the ratio of the fluorescence intensity of NAD(P)H to that of FAD, and reflects the redox state of the cell. In their work, the OMI was able to distinguish cell subpopulations with diverse dynamic responses to treatment in living organoids and was evaluated as a predictor of clinical treatment response [48]. To our best knowledge, these techniques have not been applied in childhood cancers, however provide great opportunities to study the single-cell metabolite flux, and characterize the metabolic profiles in different cell subpopulations.

Thus, established and emerging approaches support unprecedented resolution into many facets of tumor biology in primary cells. With these approaches, proteomic, signaling, metabolic, transcriptomic, and epigenomic features can be evaluated in individual cells and provide granular information regarding multiple levels of regulation or dysregulation to provide an integrated view of tumor cell heterogeneity.

Approach to data analysis

General pipelines for single-cell data analysis and visualization

Rapid innovations in the statistical and algorithmic tools used to analyze single-cell data present a unique challenge for building stable, well-tested software pipelines for single-cell analysis. Furthermore, an important practical consideration for single-cell data analysis is that single-cell data are collected and analyzed by researchers with a diverse range of experience in statistics, data science, and computer programming. This means that there is rarely a "one-size-fits-all" strategy for developing single-cell data analysis pipelines for particular tasks and building general use toolkits for the single-cell data processing remains an area of active development in the bioinformatics community (Fig. 2).

As new analytical methods are actively developed, open-source platforms that can be used at the command line in data science-friendly languages like R, Python, or Julia provide the most flexibility for customizing single-cell data analysis pipelines. For scRNA-seq data, the two most widely used open-source platforms are Scater (for quality-control and preprocessing) and Seurat (for a large variety of downstream analyses), both of which are open-source R packages distributed through the Bioconductor project [41, 49, 50]. In addition, significant efforts have been dedicated to the development of the "scverse," a Python-based consortium of tools for single-cell data analysis that includes *scanpy* for gene expression analysis, *squidpy* for spatial single-cell data analysis, and *scvi-tools* for training machine learning models at the single-cell level [51–53]. Similarly, integrated open-source packages for analyzing high-dimensional cytometry data (i.e., mass cytometry)—such as *premessa*, *CATALYST*, *cytofkit*, and *tidytof*—have been developed to provide interoperability between many data analyses and visualization tools commonly used by the cytometry community [54, 55]. Likewise, the *ARK* Python library is a toolkit developed for the integrated analysis of data collected using MIBI-TOF and other high-dimensional imaging modalities [56]. For researchers without extensive programming experience, many open-source libraries also implement graphical user interfaces (GUIs) that can facilitate manual analysis of single-cell data using either local servers or cloud computing resources [57]. Commercial platforms like Cytobank, Astrolabe, Omiq, and others exist under a subscription model to allow users to conduct exploratory data analysis and visualization interactively, often with support for state-of-the-art tools [58–60].

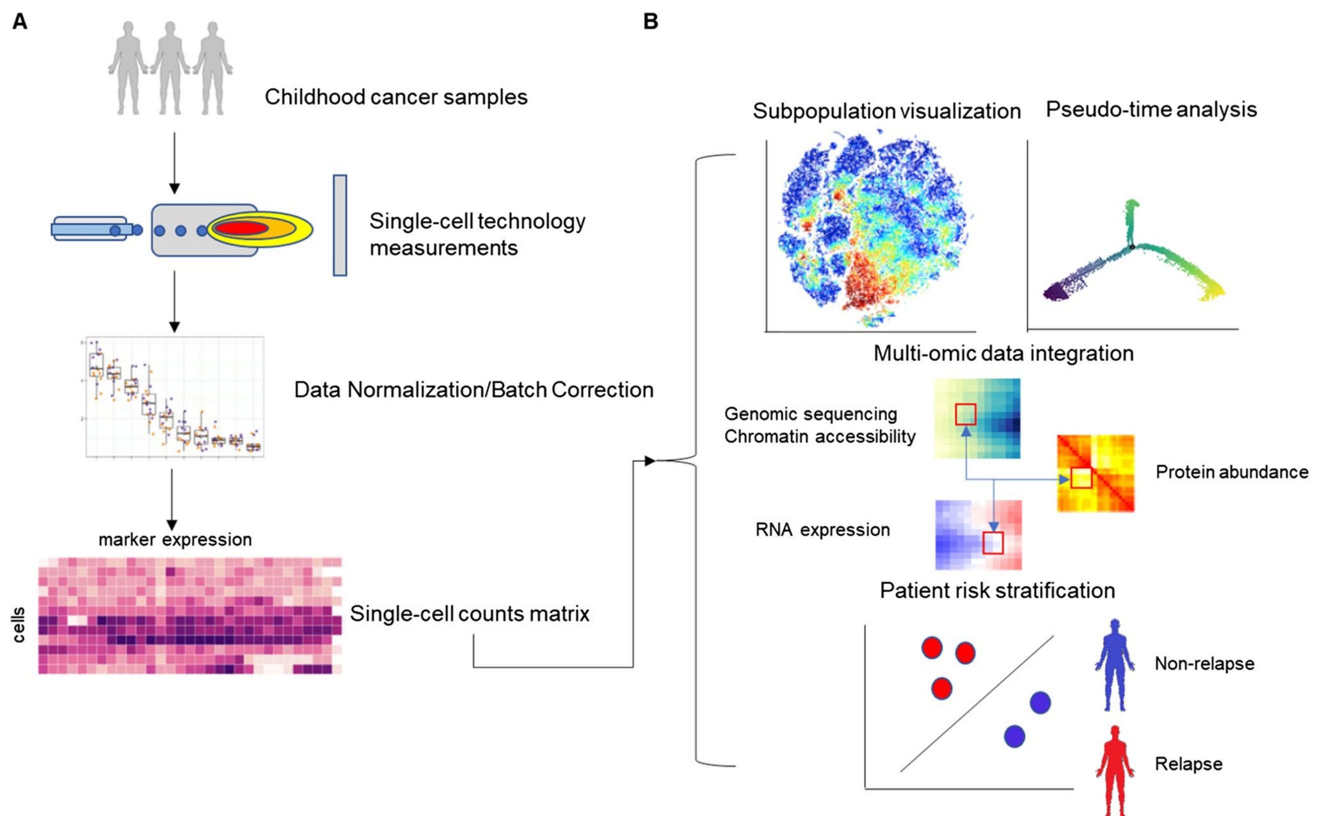

Fig. 2 Example workflow for single-cell analysis in childhood cancer samples. **A** Samples were quantified using the single-cell proteomic approach. The retrieved data were normalized and batch-corrected to identify relevant biological signals. This data preprocessing step results in a single-cell count matrix where each row represents a cell and each column represents a measured parameter. **B** Single-cell data analysis can be performed to identify subpopulations using dimensionality reduction and clustering approaches unsupervised clustering algorithms to predict expression evolution using pseudo time analysis, combined with genomic or transcriptomic data for multi-omics integration or used develop models for patient risk stratification or clinical prediction

Data normalization and processing of single-cell data

One significant challenge for analyzing single-cell data is technical variation irrelevant to true biological signals also known as batch effects. For comparative analyses across patient cohorts, technical variabilities need to be removed or minimized prior to analysis. Lo et al. recently developed an approach to batch correction of mass cytometry datasets enabling cross-cohort comparisons of 989 pediatric leukemia patient files. The approach, CytofIn, will allow more cross-cohort comparison of public datasets as more mass cytometry data is generated and available for childhood cancer patients [61]. Similarly, batch normalization of scRNA-seq data can be accomplished computationally using programs such as Seurat [62]. In these approaches, the mutual nearest neighbor (MNN) of cells between datasets is first identified, and the program moves the neighboring cells toward a common centroid through an iterative process. Batch correction for scATAC-seq data can be accomplished using algorithms such as Harmony [63]. One emerging trend is the simultaneous integration and batch effect reduction of single-cell data using matrix factorization or deep neural networks [64]. A comparison and benchmark study of the scRNA-seq normalization approach using common datasets has been reported in a recent study, where Harmony, linked inference of genomic experimental relationships (LIGER), and Seurat 3 were identified as the recommended methods for batch effect correction efficacy while preserving cell type purity [65].

Multi-omic integration

Despite rapid advance in multi-omics technologies, clinical samples where a cell is simultaneously measured on multiple omic layer remain sparse. In this case, computational approaches can be used to align the most similar cell states or extract latent space shared between multiple omic layers. Although clustering algorithms are a core part of the single-cell mono-omics analysis to identify cell types and their

functional states, the integration of single-cell data from different platforms remains challenging and requires a careful selection of statistical and mathematical frameworks. A plethora of analysis approaches has been proposed to address this. Broadly, the multi-omics data analysis approach can be divided into two major categories. Given two omics layers, the first approach examines the linear relationship between each observation of them to achieve maximum correlation. Canonical correlation analysis (CCA) is an example of these approaches [66]. The second approach utilizes matrix factorization to find a set of hidden factors within each omics layer that captures biological and technical variability. These hidden factors are subsequently utilized for data imputation and sample subgroup identification. The multi-omics factor analysis (MOFA) and LIGER utilize the matrix factorization approach for data integrations [67, 68]. Multiple omics layers can also be simultaneously analyzed by integrating sample similarities combined with dimension reduction techniques. For sample similarities, mainly spectral clustering approaches and graph fusion algorithms are utilized. For example, the recent version of Seurat utilizes a weighted-nearest neighbor (WNN) graph-based integration for cluster analysis [41]. The aim of combined dimensionality reduction techniques is to define a lower-dimensional representation of the single-cell multi-omics data by projecting them into a common latent space. The emergence of single-cell multi-omics technologies and data analysis tools has provided novel insight into the tumor microenvironment (TME) of pediatric cancers. By utilizing scRNA-seq and cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), Witkowski et al. described a role for non-classical monocytes in B-ALL outcomes [69] finding that the presence of non-classical monocytes in the BM at diagnosis is predictive of survival in pediatric and adult B-ALL patients. One challenge of the current computational approach for multi-omics data integration is the potential of information loss from dimensional reduction. Due to high multiplicity in precisely defining a cell state, errors can be generated using the cell alignment approach. Another challenge is the deconvolution of bulk data which consists of the majority of the clinical data in the public domain. To tackle this challenge, several computational techniques for single-cell deconvolution such as CIBERSORT and others have recently been proposed [70, 71].

The promise and challenge of single-cell techniques for clinical translation

The emerging use of single-cell technologies in the context of childhood cancers demonstrates the potential clinical impact yet highlights the barriers to wide adoption. For instance, minimal residual disease (MRD) is a widely trusted prognostic indicator of relapse risk in children with leukemia [72, 73]. Minimal residual disease is defined as cells that persist after chemotherapy at early time points. In the setting of pediatric B-ALL, it has been used to guide therapy intensity and justify bone marrow transplantation [72, 73]. Current techniques detect MRD but do not inform on the features of these resistant cells. Given the challenge to capture low-abundance residual cells after treatment, single-cell sequencing techniques provide great opportunities to better understand MRD cells. To analyze the dynamic changes in B-ALL under treatment, Zhang and colleagues studied the molecular signatures from longitudinal samples obtained at the diagnostic, MRD, and relapse stages from 4 patients. Using scRNA-seq from healthy BM cells, they identified the developmental states of each leukemic cell based on its transcriptomic signature. Performing single-cell B-cell receptor (scBCR-seq) sequencing distinguished leukemic from non-leukemic cells based on BCR clonality in the CD19 + populations. Pathway enrichment analysis demonstrated upregulation of genes involved in the hypoxia pathway in MRD cells from two patients. Further experiments demonstrated inhibition of the hypoxia pathway by a HIF1α inhibitor sensitized cell line and PDX cells to chemotherapy [74]. This study reveals both the promise and challenge of the use of single-cell technologies for clinical practice. Analyzing MRD cells themselves reveals new targets for therapy, and single-cell approaches are required to approach these rare and critical MRD cells. However, the cost and complexity limit the wide application of these approaches, thus restricting wide adoption into clinical care. Data analysis approaches will need to be standardized and accessible to enable these techniques to be integrated into more standard clinical care. Further, greater numbers of patient samples will need to be examined to understand the shared and unique features revealed by single-cell approaches.

Conclusions

Single-cell technologies are proving to be invaluable for dissecting clonal heterogeneity and diversity in childhood cancers. With continued application, we are learning more about the molecular basis of disease and uncovering novel strategies to address cancer resistance. In many published approaches, including our own, single-cell data has been modeled based on the normal tissue of origin of the tumor itself. This highlights the developmental nature of childhood tumors and helps organize the highly heterogeneous nature of this data. Already, the application of CyTOF and scRNA-seq has provided insight into the molecular mechanisms of tumor resistance in both pediatric leukemia and solid tumors in growing studies. We anticipate that large-scale integration

of identified signatures from multi-dimensional data will further improve diagnostic and relapse predictive power. Similarly, the generation of single-cell data will also necessitate the development of novel computational approaches to integrate datasets and maximize the knowledge generated from these studies. Taken together, the evolution and application of single-cell analyses to childhood cancer will be transformative in our understanding of tumor biology and in improving patient outcomes for children with cancer.

Acknowledgements We thank all members of the Davis lab for helpful discussions.

Funding KLD is the Anne T. and Robert M. Bass Endowed Faculty Scholar in Pediatric Cancer and Blood Diseases. This work is supported by Stanford Maternal and Child Health Research Institute, NCI U54 CA232568, NCI R01 CA251858, NCI R21 CA234529, NCI R01 CA251858-01A1S1, Mark Foundation Aspire Award, The Andrew McDonough B Positive Foundation, W81XWH-19-PRCRP-CDA Department of Defense Young Investigator Award.

Declarations

Competing interests The authors declare no competing interests.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A (2021) Cancer statistics, 2021. CA Cancer J Clin 71:7–33
2. Neaga A et al (2021) Why do children with acute lymphoblastic leukemia fare better than adults?. Cancers (Basel) 13
3. Lee SHR, Li Z, Tai ST, Oh BLZ, Yeoh AEJ (2021) Genetic alterations in childhood acute lymphoblastic leukemia: interactions with clinical features and treatment response. Cancers (Basel) 13
4. Lee SHR, Li Z, Tai ST, Oh BLZ, Yeoh AEJ (2021) Genetic alterations in childhood acute lymphoblastic leukemia: interactions with clinical features and treatment response. Cancers (Basel) 13:4068
5. Radtke I et al (2009) Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia. Proc Natl Acad Sci U S A 106:12944–12949
6. Aynaud MM et al (2020) Transcriptional programs define intratumoral heterogeneity of Ewing sarcoma at single-cell resolution. Cell Rep 30:1767-1779.e1766
7. Bandura DR et al (2009) Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. Anal Chem 81:6813–6822
8. Ornatsky O et al (2010) Highly multiparametric analysis by mass cytometry. J Immunol Methods 361:1–20
9. Leelatian N et al (2017) Single cell analysis of human tissues and solid tumors with mass cytometry. Cytometry B Clin Cytom 92:68–78
10. Jaimes MC et al (2021) Full spectrum flow cytometry and mass cytometry: a 32-marker panel comparison. Cytometry A
11. Jager A, Sarno J, Davis KL (2021) Mass cytometry of hematopoietic cells. Methods Mol Biol 2185:65–76
12. Spitzer MH, Nolan GP (2016) Mass cytometry: single cells, many features. Cell 165:780–791
13. Good Z et al (2018) Single-cell developmental classification of B cell precursor acute lymphoblastic leukemia at diagnosis reveals predictors of relapse. Nat Med 24:474–483
14. Angelo M et al (2014) Multiplexed ion beam imaging of human breast tumors. Nat Med 20:436–442
15. Giesen C et al (2014) Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat Methods 11:417–422
16. Black S et al (2021) CODEX multiplexed tissue imaging with DNA-conjugated antibodies. Nat Protoc 16:3802–3835
17. Goltsev Y et al (2018) Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. Cell 174:968-981.e915
18. Keren L et al (2018) A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. Cell 174:1373-1387.e1319
19. Keren L et al (2019) MIBI-TOF: a multiplexed imaging platform relates cellular phenotypes and tissue structure. Sci Adv 5:eaax5851
20. Liu, C.C. et al. Reproducible, high-dimensional imaging in archival human tissue by multiplexed ion beam imaging by time-of-flight (MIBI-TOF). Laboratory Investigation (2022).
21. Liu CC et al (2022) Multiplexed ion beam imaging: insights into pathobiology. Annu Rev Pathol 17:403–423
22. Kammersgaard MB et al (2020) Abstract PO-041: Multiplexed ion beam imaging to describe tumor-immune microenvironment and tumor heterogeneity in neuroblastoma. Cancer Res 80:PO-041-PO-041
23. Batth IS et al (2020) Rare osteosarcoma cell subpopulation protein array and profiling using imaging mass cytometry and bioinformatics analysis. BMC Cancer 20:715
24. Gerdtsson E et al (2018) Multiplex protein detection on circulating tumor cells from liquid biopsies using imaging mass cytometry. Converg Sci Phys Oncol 4
25. Black S et al (2021) CODEX multiplexed tissue imaging with DNA-conjugated antibodies. Nat Protoc 16:3802–3835
26. Bosisio FM et al (2022) Next-generation pathology using multiplexed immunohistochemistry: mapping tissue architecture at single-cell level. Front Oncol 12:918900
27. Cesano A, Marincola FM, Thurin M (2020) Status of immune oncology: challenges and opportunities. Methods Mol Biol 2055:3–21
28. Gawad C, Koh W, Quake SR (2014) Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. Proc Natl Acad Sci 111:17947–17952
29. Mehtonen J et al (2020) Single cell characterization of B-lymphoid differentiation and leukemic cell states during chemotherapy in ETV6-RUNX1-positive pediatric leukemia identifies drug-targetable transcription factor activities. Genome Med 12:99
30. Caron M et al (2020) Single-cell analysis of childhood leukemia reveals a link between developmental states and ribosomal protein expression as a source of intra-individual heterogeneity. Sci Rep 10:8079
31. Louka E et al (2021) Heterogeneous disease-propagating stem cells in juvenile myelomonocytic leukemia. J Exp Med 218
32. Hovestadt V et al (2019) Resolving medulloblastoma cellular architecture by single-cell genomics. Nature 572:74–79
33. Vladoiu MC et al (2019) Childhood cerebellar tumours mirror conserved fetal transcriptional programs. Nature 572:67–73
34. Gillen AE et al (2020) Single-cell RNA sequencing of childhood ependymoma reveals neoplastic cell subpopulations that impact molecular classification and etiology. Cell Rep 32:108023
35. Hovestadt V et al (2019) Resolving medulloblastoma cellular architecture by single-cell genomics. Nature 572:74–79
36. Jansky S et al (2021) Single-cell transcriptomic analyses provide insights into the developmental origins of neuroblastoma. Nat Genet 53:683–693
37. Corces MR et al (2017) An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14:959–962

38. Shahi P, Kim SC, Haliburton JR, Gartner ZJ, Abate AR (2017) Abseq: ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. Sci Rep 7:44447
39. Stoeckius M et al (2017) Simultaneous epitope and transcriptome measurement in single cells. Nat Methods 14:865–868
40. Mimitou EP et al (2019) Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat Methods 16:409–412
41. Hao Y et al (2021) Integrated analysis of multimodal single-cell data. Cell 184:3573-3587.e3529
42. Bai Z et al (2022) Single-cell antigen-specific landscape of CAR T infusion product identifies determinants of CD19-positive relapse in patients with ALL. Sci Adv 8:2820
43. Lee J, Hyeon DY, Hwang D (2020) Single-cell multiomics: technologies and data analysis methods. Exp Mol Med 52:1428–1442
44. Mizuno H, Tsuyama N, Date S, Harada T, Masujima T (2008) Live single-cell metabolomics of tryptophan and histidine metabolites in a rat basophil leukemia cell. Anal Sci 24:1525–1527
45. Pan N, Rao W, Yang Z (2020) Single-probe mass spectrometry analysis of metabolites in single cells. Methods Mol Biol 2064:61–71
46. Ahl PJ et al (2020) Met-Flow, a strategy for single-cell metabolic analysis highlights dynamic changes in immune subpopulations. Commun Biol 3:305
47. Arguello RJ et al (2020) SCENITH: a flow cytometry-based method to functionally profile energy metabolism with single-cell resolution. Cell Metab 1063–1075:e1067
48. Sharick JT et al (2020) Metabolic heterogeneity in patient tumor-derived organoids by primary site and drug treatment. Front Oncol 10:553
49. McCarthy DJ, Campbell KR, Lun AT, Wills QF (2017) Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics 33:1179–1186
50. Satija R, Farrell JA, Gennert D, Schier AF, Regev A (2015) Spatial reconstruction of single-cell gene expression data. Nat Biotechnol 33:495–502
51. Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 19:15
52. Palla G et al (2022) Squidpy: a scalable framework for spatial omics analysis. Nat Methods 19:171–178
53. Gayoso A et al (2022) A Python library for probabilistic analysis of single-cell omics data. Nat Biotechnol 40:163–166
54. Nowicka M et al (2017) CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets. Res 6:748
55. Chen H et al (2016) Cytofkit: a bioconductor package for an integrated mass cytometry data analysis pipeline. PLoS Comput Biol 12:e1005112
56. Greenwald NF et al (2022) Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nat Biotechnol 40:555–565
57. Luecken MD, Theis FJ (2019) Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol 15:e8746
58. Kotecha N, Krutzik PO, Irish JM (2021) Web-based analysis and publication of flow cytometry experiments. Current protocols in cytometry Chapter 10, Unit10.17-Unit10.17
59. Amir E-AD et al (2019) Development of a comprehensive antibody staining database using a standardized analytics pipeline. Front Immunol 10:1315–1315
60. Belkina AC et al (2019) Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. Nat Commun 10:5415
61. Lo YC et al (2022) CytofIn enables integrated analysis of public mass cytometry datasets using generalized anchors. Nat Commun 13:934
62. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36:411–420
63. Baek S, Lee I (2020) Single-cell ATAC sequencing analysis: from data preprocessing to hypothesis generation. Comput Struct Biotechnol J 18:1429–1439
64. Kopp W, Akalin A, Ohler U (2022) Simultaneous dimensionality reduction and integration for single-cell ATAC-seq data using deep learning. Nature Machine Intelligence 4:162–168
65. Tran HTN et al (2020) A benchmark of batch-effect correction methods for single-cell RNA sequencing data. Genome Biol 21:12
66. Hardoon DR, Shawe-Taylor J (2011) Sparse canonical correlation analysis. Mach Learn 83:331–353
67. Argelaguet R et al (2018) Multi-omics factor analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol 14:e8124
68. Welch JD et al (2019) Single-cell multi-omic integration compares and contrasts features of brain cell identity. Cell 177:1873-1887.e1817
69. Witkowski MT et al (2020) Extensive remodeling of the immune microenvironment in b cell acute lymphoblastic leukemia. Cancer Cell 37:867-882.e812
70. Avila Cobos F, Alquicira-Hernandez J, Powell JE, Mestdagh P, De Preter K (2020) Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat Commun 11:5650
71. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA (2018) Profiling tumor infiltrating immune cells with CIBERSORT. Methods Mol Biol 1711:243–259
72. Campana D (2010) Minimal residual disease in acute lymphoblastic leukemia. Hematology Am Soc Hematol Educ Program 2010:7–12
73. van der Velden VH, Boeckx N, van Wering ER, van Dongen JJ (2004) Detection of minimal residual disease in acute leukemia. J Biol Regul Homeost Agents 18:146–154
74. Zhang Y et al (2022) Elucidating minimal residual disease of paediatric B-cell acute lymphoblastic leukaemia by single-cell analysis. Nat Cell Biol 24:242–252

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.