

MEDICAL UNIVERSITY OF GDAŃSK

FACULTY OF PHARMACY



Cell-type resolution data analysis using different transcriptomics approaches for improvement of current methods

Analizy z rozdzielczością na poziomie pojedynczych komórek z zastosowaniem różnych podejść transkryptomicznych w celu udoskonalenia istniejących metod

Elyas Mohammadi

Supervisor:

Dr hab. Jakub Mieczkowski

**PhD thesis performed at 3P-Medicine
Laboratory at the Medical University
of Gdańsk**

GDAŃSK 2023

ACKNOWLEDGEMENTS

I sincerely thank:

Promoter, Dr hab. Jakub Mieczkowski for substantive support, valuable comments and ideas provided during the preparation of this work.

Professor dr. Jan Dumanski, Professor dr. hab. Arkadiusz Piotrowski and all employees of the 3P-Medicine Laboratory, Medical University of Gdansk for their help and valuable advice and numerous, inspiring scientific discussions. Edyta Rychlicka-Buniowska for her support in writing the Polish abstract.

My lovely wife Samira and adorable son Ryan for their patience, huge support, faith in my abilities and for the fact that I could always rely on them.

This work was supported by the Foundation for Polish Science under the International Research Agendas Program financed from the Smart Growth Operational Program 2014–2020 (Grant Agreement No. MAB/2018/6).



TABLE OF CONTENTS

LIST OF ARTICLES INCLUDED IN DOCTORAL DISSERTATION	3
Paper I:	3
Paper II:	3
Paper III:	3
ABBREVIATIONS	4
ABSTRACT IN POLISH	6
ABSTRACT IN ENGLISH	8
SUMMARY	10
Introduction	10
General introduction	10
Immortalized cell lines	10
Primary cell sorting	15
Single cell RNA sequencing	17
Spatial transcriptomics	19
Challenges and solutions	21
PAPER I	21
PAPER II	22
PAPER III	23
Aims	25
Material and Methods	26
PAPER I: scRNA-seq and ST of postmortem brain samples	26
PAPER II: primary cell sorting of CML Tregs	27
PAPER III: developing pipeline for data extraction from LINCS L1000 database	29
Results	31
PAPER I	31
PAPER II	32
PAPER III	33
Conclusion	36
REFERENCES	37
PUBLICATIONS	41
Paper I	41
Paper II	56
Paper III	76

LIST OF ARTICLES INCLUDED IN DOCTORAL DISSERTATION

This thesis is based on the below listed papers, which are referred to in the text by their Roman numerals I - III.

Paper I:

- **Mohammadi E**, Chojnowska K, Bienkowski M, Kostecka A, Koczkowska M, Zmijewski MA, Jakalski M, Ingelsson M, Filipowicz N, Olszewski P, Davies H, Wierzbička JM, Hyman BT, Dumanski JP, Piotrowski A, Mieczkowski J: Size matters: the impact of nucleus size on results from spatial transcriptomics, *J Transl Med* 2023, 21:270; (<https://www.ncbi.nlm.nih.gov/pubmed/37081484>) **IF: 8.448, MNiSW: 100**

Paper II:

- Swatler J, Tuross-Korgul L, Brewinska-Olchowik M, De Biasi S, Dudka W, Le BV, Kominek A, Cyranowski S, Pilanc P, **Mohammadi E**, Cysewski D, Kozłowska E, Grabowska-Pyrzewicz W, Wojda U, Basak G, Mieczkowski J, Skorski T, Cossarizza A, Piwocka K: 4-1BBL-containing leukemic extracellular vesicles promote immunosuppressive effector regulatory T cells, *Blood Adv* 2022, 6:1879-1894; (<https://www.ncbi.nlm.nih.gov/pubmed/35130345>) **IF: 7.642, MNiSW: 40**

Paper III:

- **Mohammadi E**, Tahmoorespur M, Benfeitas R, Altay O, Javadmanesh A, Lam S, Mardinoglu A, Sekhavati MH: Improvement of the performance of anticancer peptides using a drug repositioning pipeline, *Biotechnol J* 2022, 17: e2100417; (<https://www.ncbi.nlm.nih.gov/pubmed/34657375>) **IF: 5.081, MNiSW: 100**

Total Impact Factor: 21.171

Total score of the Ministry of Science and Higher Education: 240

ABBREVIATIONS

Abbreviation	Full term
ACPs	Anti-Cancer Peptides
AML	Acute Myeloid Leukemia
CCR4	CC chemokine Receptor 4
Cdna	complementary DNA
ChIP	Chromatin immunoprecipitation
CMap	Connectivity Map
CML	Chronic Myeloid Leukemia
CNVs	Copy Number Variations
CS	Chondroitin Sulfate
CSDI	Consecutive Slices Data Integration
DEP	Dielectrophoresis
E2F4	E2F Transcription Factor 4
EGA	The European Genome-phenome Archive
EGR-1	Early Growth Response protein 1
EGR-3	Early Growth Response protein 3
ENTPD1	Ectonucleoside Triphosphate Diphosphohydrolase 1
EV	Extracellular Vesicle
FACS	Fluorescence-Activated Cell Sorting
FITC	Fluorescein Isothiocyanate
GEM	Gel Bead in Emulsion
GEO	Gene Expression Omnibus
GM	Grey Matter
GO	Gene Ontology
H&E	Hematoxylin and Eosin
HAVCR2	Hepatitis A Virus Cellular Receptor 2
HS	Heparan Sulfate
hTert	human Telomerase reverse transcriptase
IL1R1	Interleukin 1 Receptor Type 1
KE	Kendall tau correlation
LFC	Log Fold Changes

LINCS	Library of Integrated Network-based Cellular Signatures
MACS	Magnetic-Activated Cell Sorting
MERFISH	Multiplexed Error-Robust Fluorescence in situ Hybridization
mRNA	Messenger RNA
OCT	Optimal Cutting Temperature
ON	Orbitofrontal Neocortex
PE	Pearson correlation
scRNA-seq	Single cell RNA sequencing
SNPs	Single Nucleotide Polymorphisms
snRNA-seq	single nuclei RNA sequencing
SP	Spearman correlation
ST	Spatial Transcriptomics
STARmap	Spatially-resolved Transcript Amplicon Readout Mapping
TFBM	Transcription Factor Binding Motifs
TFDP1	Transcription Factor Dp-1
TFRC	Transferrin Receptor
TGFB1	Transforming Growth Factor Beta 1
TN	Temporal Neocortex
TNFRSF1B	TNF Receptor Superfamily Member 1B
TNFRSF8	TNF Receptor Superfamily Member 8
Treg	Regulatory T cells
UMIs	Unique molecular identifiers
WM	White Matter
ZBTB7A	Zinc Finger And BTB Domain Containing 7A

ABSTRACT IN POLISH

W celu pełnego poznania złożoności systemów biologicznych jakimi są tkanki konieczne jest zastosowanie czułych metod umożliwiających przeprowadzenie analiz z rozdzielczością na poziomie pojedynczych komórek. W ramach niniejszej rozprawy wykorzystane zostały różnorodne metody z zakresu analizy transkryptomicznej, takie jak: sekwencjonowanie RNA z pojedynczych komórek, sekwencjonowanie RNA z homogennej puli komórek pierwotnych wyizolowanych za pomocą aktywowanego fluorescencyjnie sortera komórek, transkryptomika przestrzenna i mikromacierze genowe. Metody te zostały zastosowane do analizy ludzkiego mózgu, zbadania funkcji limfocytów T regulatorowych u pacjentów z białaczką szpikową oraz analizy zestawu danych LINCS L1000, będącego wielkoskalowym kompendium profili ekspresji genów z różnych ludzkich linii komórkowych poddanych działaniu rozmaitych leków. Poprzez wprowadzenie odpowiednich zmian w podejściach analitycznych dostępnych obecnie dla wymienionych powyżej metod, podjęte zostało dążenie do określenia wewnętrznej heterogenności tkanek, rozszyfrowania zależności przestrzennych pomiędzy komórkami, a także identyfikacji procesów molekularnych leżących u podstaw różnorodnych zjawisk biologicznych, zarówno w tkankach zdrowych jak i w procesie chorobowym.

Celem niniejszej pracy było: 1) przeprowadzenie oceny przydatności wybranych technik laboratoryjnych do identyfikacji różnych typów komórek; 2) wykazanie przydatności integracji danych uzyskanych za pomocą różnych metod transkryptomicznych; 3) wykazanie przydatności nowych podejść analitycznych do udoskonalenia analiz transkryptomiki przestrzennej, poznania mechanizmów chorobowych oraz identyfikacji nowych potencjalnych strategii terapeutycznych.

Wyniki niniejszej pracy przyczyniły się do istotnego postępu na polu badań transkryptomicznych i wskazują jak ważną rolę w rozszyfrowaniu złożoności systemów biologicznych odgrywa zastosowanie zaawansowanych narzędzi obliczeniowych i metod statystycznych. Analiza profili ekspresji genów specyficznych dla różnych typów komórek pozwoliła na pełniejsze zrozumienie mechanizmów chorobowych i może przyczynić się do rozwoju medycyny precyzyjnej. W przypadku analizy ludzkiego mózgu zastosowanie metody CSDI pozwoliło na zidentyfikowanie kompletnych przestrzennych profili transkryptomicznych komórek. Analiza RNA w puli wysortowanych komórek T regulatorowych poddanych działaniu pęcherzyków zewnątrzkomórkowych wyizolowanych z osocza pacjentów z białaczką szpikową umożliwiła identyfikację mechanizmów immunosupresji, które mają krytyczne

znaczenie dla wyników leczenia i przeżycia pacjentów i mogą być w przyszłości potencjalnym celem nowych terapii. Analiza zestawu danych LINCS L1000 pozwoliła na identyfikację nowych potencjalnych peptydów przeciwnowotworowych, których aktywność zostanie przetestowana w warunkach laboratoryjnych.

Słowa kluczowe: bioinformatyka, metody analizy z rozdzielczością na poziomie komórki, transkryptomika przestrzenna, profilowanie pojedynczych komórek/jąder komórkowych, mikromacierze genowe, transkryptomika zbiorcza, sortowanie komórek pierwotnych, linie komórkowe

ABSTRACT IN ENGLISH

To resolve the complexity of data analysis for biological systems, sensitive methods with single cell resolution are required. This research endeavors to advance our understanding of transcriptomic profiling methods, encompassing single-cell RNA sequencing, spatial transcriptomics, bulk transcriptomics and gene microarray. By improving the current data analysis approaches for these transcriptomics methods, we seek to decipher intricate cellular heterogeneity, spatial relationships, and the molecular underpinnings of diverse biological phenomena.

Our study seeks to achieve three main objectives: firstly, to assess the utility of diverse wet lab techniques in obtaining distinct cell types; secondly, to demonstrate the feasibility of integrating and analyzing data from different techniques using transcriptomics methods; and finally, to showcase the effectiveness of these approaches in enhancing spatial analysis, unraveling disease mechanisms, and identifying potential therapeutic strategies.

In our study, we utilized scRNA-seq, spatial transcriptomics, primary cell sorting, and microarray profiling. We used these methods to investigate human brain tissue, myeloid leukemia cells, and a large-scale compendium of functional perturbations in cultured human cells, known as LINCS L1000 dataset, respectively. These methods were carefully selected and executed to provide comprehensive insights into gene expression profiles across distinct biological systems.

Our findings represent significant advancements in transcriptomic analysis. In our study of human brain tissue, the introduction of Consecutive Slices Data Integration (CSDI) rectified incomplete spatial transcriptomic profiles, yielding superior spot clustering and label transferring. The bulk transcriptomics analysis of primary sorted myeloid leukemia cells unveiled the impact of leukemic extracellular vesicles on regulatory T cells, unraveling critical immunosuppression mechanisms. Furthermore, our drug repositioning pipeline, applied to LINCS L1000 data, unveiled potential drugs to enhance the application of anticancer peptides.

This research underscores the pivotal role of advanced computational tools and statistical methodologies in unraveling the complexity of transcriptomics. By deciphering cell-type-specific gene expression patterns and their functional implications, we pave the way for a deeper understanding of intricate biological systems and diseases. Our work holds promise for advancing precision medicine and the development of novel therapeutic strategies.

Keywords: Bioinformatics, Cell-type resolution methods, Spatial transcriptomics, Single cell/nucelli gene profiling, Gene microarray, Bulk transcriptomics, Primary cell sorting, Cell lines

SUMMARY

Introduction

General introduction

Cells are the fundamental units of life, and can be classified based on their structural and functional characteristics. In multicellular organisms, cells come together to form tissues. The composition of tissues and their functions are regulated by the transcription of different genes. Gene transcription is the process by which the genetic information stored in DNA is converted into RNA molecules. Thus, understanding the expression patterns of these genes is crucial for unraveling the molecular mechanisms underlying normal and pathological states of tissues ¹. Previous gene profiling techniques, such as bulk transcriptomics, lacked single cell resolution. Bulk transcriptomics, which involves sequencing RNA from a population of cells, averages the gene expression levels across all cells in a tissue. Although combining bulk transcriptomics with methods such as primary cell sorting can enhance the resolution of cell gene profiling, using it with a pool of cells obtained from tissue covers the information about cellular heterogeneity and diversity. This can be particularly problematic when studying tissues with complex cellular compositions, as rare cell-types or subpopulations of cells may be overlooked or underrepresented in the data ². Hence, it is important to pursue new ways to decompose tissues to cell-types and further profile their expressions separately.

There are multiple approaches to distinguish the expression of particular cell-types such as working on immortalized cell lines, primary cell sorting, and dissociating the tissue to single cells and labeling them with gene barcodes. Immortalized cell lines are established and genetically modified or continuously dividing cultured cells that overcome natural cellular lifespan, providing an enduring resource for research purposes. Primary cells are cells derived directly from living organisms, in contrast to established cell lines that have been cultivated and maintained in the laboratory for prolonged periods. Single-cell RNA sequencing (scRNA-seq) method captures and sequences the RNA of individual cells, enabling the analysis of gene expression at a single-cell level, revolutionizing our understanding of cellular diversity and dynamics. These methods will be discussed in more details in the following sections.

Immortalized cell lines

An immortalized cell line refers to cells that has acquired the ability to continuously replicate and divide beyond the typical limitations of normal cells. Normal cells have a limited lifespan due to a process called cellular senescence and apoptosis, which is regulated by

mechanisms like telomere shortening and DNA damage responses. Senescence is the process by which cells irreversibly stop dividing and enter a state of permanent growth arrest without undergoing cell death. Immortalized cell lines have overcome these limitations and can divide indefinitely, resembling a more embryonic or stem cell-like state ³.

There are both normal and cancer-derived immortalized cell lines. Normal cell lines are derived from healthy tissues and are often used in research to study basic cellular processes. Creating an immortalized cell line from normal tissues involves introducing specific genetic alterations that bypass cellular senescence and apoptosis checkpoints. This is often achieved by introducing viral genes (such as EBV, SV40 T antigens and HPV-16 E6/7 gene) or human telomerase reverse transcriptase (hTert) and oncogenes (like c-Myc) into the cells ⁴. Another method is employing a lentiviral vector library encoding several of aforementioned genes (Figure 1A). These genes have the capacity to either inhibit the function of tumor suppressor genes or activate oncogenes, thus stimulating cell division and extending the cell's lifespan. Furthermore, primary cells may undergo mutations during the process of division and multiplication. Their spontaneous mutations that occur over multiple cell culture passages can ultimately result in cellular immortalization. This phenomenon of cell immortalization resulting from spontaneous mutations is commonly referred to as "spontaneous immortalization" ⁴⁻⁵ (Figure 1B). But this method is inefficient in making normal immortalized cell lines as these cells will transform into the tumor cells in most cases. There is a basic difference between tumor cells and normal immortalized cells. Tumor cells have many classical characteristics such as losing contact inhibition, low adhesive ability, and apoptosis inhibition, while immortalized normal cells remain their normal genotype and phenotype ⁴.

Cancer-derived immortalized cell lines are derived from cancerous tissues and maintain the characteristics of the original tumor, making them valuable tools for cancer research and drug testing. These cells already possess alterations that have allowed them to bypass apoptosis/senescence and divide in an uncontrolled manner. Researchers can culture and propagate these cells under controlled conditions to establish immortalized cell lines that represent the characteristics of the original cancer ⁶.

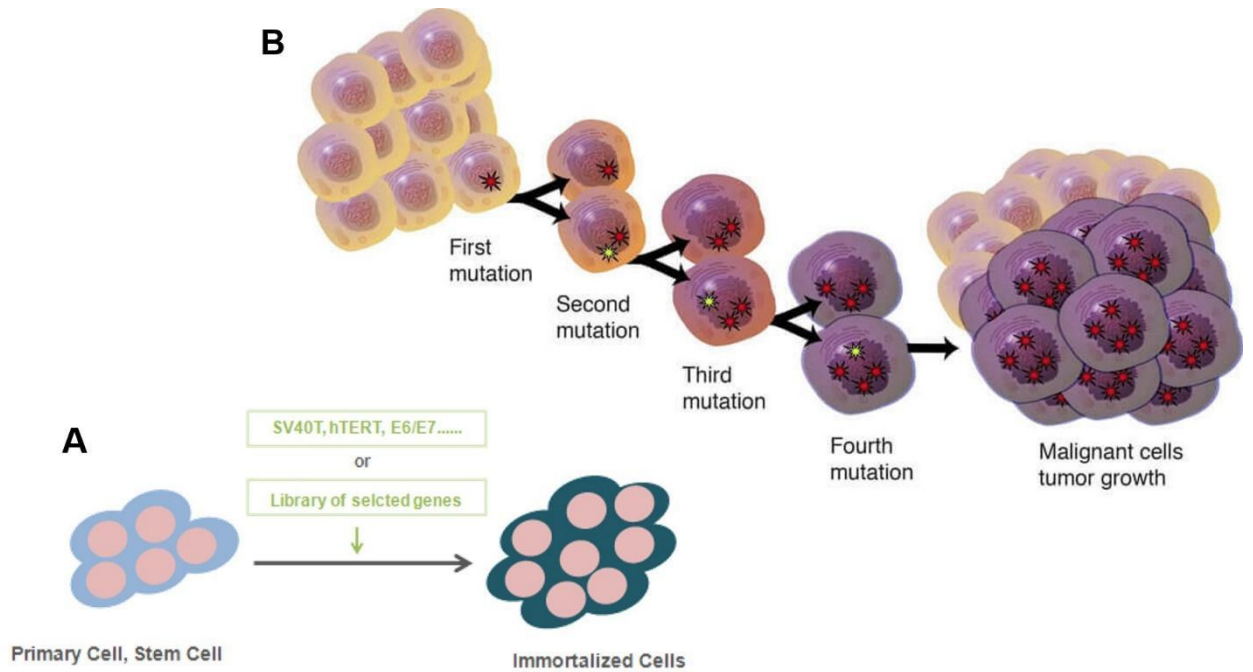


Figure 1. Cell immortalization; A) introducing genes, which can activate an oncogene or suppress the activity of tumor suppressor gene and promote cell division; B) by spontaneous mutation in the process of dividing and multiplying (adopted from creative bioarray; <https://www.creative-bioarray.com/>).

Working on cell lines provides a cost-effective approach, offering the advantage of studying cells in a controlled environment. However, cell lines may not fully represent the biology of the original tissue, and they may be subject to genetic drift or other changes over time⁷. Accumulation of genetic changes may cause genetic and phenotypic heterogeneity in different cells of the same cell line as every cell division can introduce a number of different mutations. It can eventually lead to differences in characteristics such as growth rate, response to stimuli and gene expression profiles. In addition, certain immortalized cell lines may have specific requirements for culture conditions, such as specific media formulations, growth factors, or supplements. Some cell lines may exhibit a slow growth rate, be more sensitive to changes in culture conditions, or require specialized techniques for successful propagation⁸.

Cancer cell lines serve as invaluable tools with diverse applications in drug discovery. These cell lines, which are derived from various cancer types, provide researchers with a simplified yet representative model of the complex cellular and molecular characteristics of cancers. Their ability to be cultured and manipulated in the laboratory setting allows for efficient screening of compounds, assessment of drug efficacy, and investigation into the underlying mechanisms of cancer development and progression⁹. Drug repositioning, also

known as drug repurposing, constitutes a strategic approach to drug discovery wherein existing drugs that were originally developed for one medical condition are identified and repurposed for the treatment of another condition. This approach capitalizes on the known pharmacological profiles of these drugs, potentially accelerating the drug development process and minimizing risks associated with safety and efficacy. Drug repositioning is particularly relevant to cancer research due to the intricate nature of cancer biology and the urgent need for effective therapies. By using cancer cell lines, researchers can screen a variety of existing drugs against these cell lines to identify compounds with unexpected anti-cancer properties ⁹.

In the context of drug repositioning, one effective method to assess the impact of drugs on cell lines involves the sequencing of treated and untreated cell lines to investigate their response. Among the cost-efficient techniques available, microarray stands out as a reasonable option. Microarrays are versatile tools that have revolutionized various fields of molecular biology and genomics, enabling high-throughput analysis of a multitude of biological molecules. Within the realm of microarray technology, there are several distinct types including Chromatin immunoprecipitation microarrays (ChIP-chip) ¹⁰, Genomic DNA Microarrays ¹¹, gene Expression Microarrays ¹² and protein Microarrays ¹³ highlighting their unique applications and functionalities.

Chromatin Immunoprecipitation Microarrays (ChIP -chip) ¹⁰ merge chromatin immunoprecipitation (ChIP) with microarray technology to map protein-DNA interactions across the genome. By selectively isolating DNA fragments associated with specific proteins and analyzing them on a microarray chip, researchers gain insights into gene regulation and epigenetic modifications. Genomic DNA Microarrays ¹¹ are designed to detect genetic variations within an organism's DNA, including copy number variations (CNVs) and single nucleotide polymorphisms (SNPs). They are invaluable for genotyping studies, helping researchers explore genetic distinctions among individuals, populations, or species, and unraveling the genetic underpinnings of various biological phenomena. Gene Expression Microarrays ¹² enable the simultaneous measurement of thousands of genes within a biological sample, shedding light on gene activity under specific conditions. They are instrumental in transcriptomics research, facilitating the comprehension of mRNA expression patterns in response to environmental cues, treatments, or developmental stages. Their applications span diverse domains, including identifying disease biomarkers, probing gene regulation networks, and characterizing intervention or environmental impacts. Protein Microarrays ¹³ provide a platform for the systematic examination of protein interactions, binding partners, and

expression levels. They enable the profiling of specific proteins or antibodies in biological samples, aiding in the elucidation of protein-protein interactions, antigen-antibody relationships, and protein expression patterns. These microarrays find applications in proteomics research, biomarker discovery, and drug development, offering insights into the complexities of protein biology. These diverse microarray methods are invaluable in advancing our understanding of biological processes and are indispensable tools for researchers across various scientific disciplines.

Gene expression microarray is ideal for investigating the effect of drugs in the context of drug repositioning. The analysis begins with the extraction of RNA from both treated and untreated cell lines. This RNA is then converted into labeled complementary DNA (cDNA) using reverse transcription. The cDNA is then applied to the microarray chip, where it hybridizes with the immobilized probes. After washing to remove any non-specific binding, the chip is scanned to quantify the intensity of fluorescent signals, revealing the level of gene expression (Figure 2). The obtained data is then subjected to rigorous bioinformatics analysis to identify differentially expressed genes, which may indicate the impact of the drugs on the cellular pathways. This approach allows researchers to profile how specific drugs influence the cellular response at the molecular level ¹⁴.

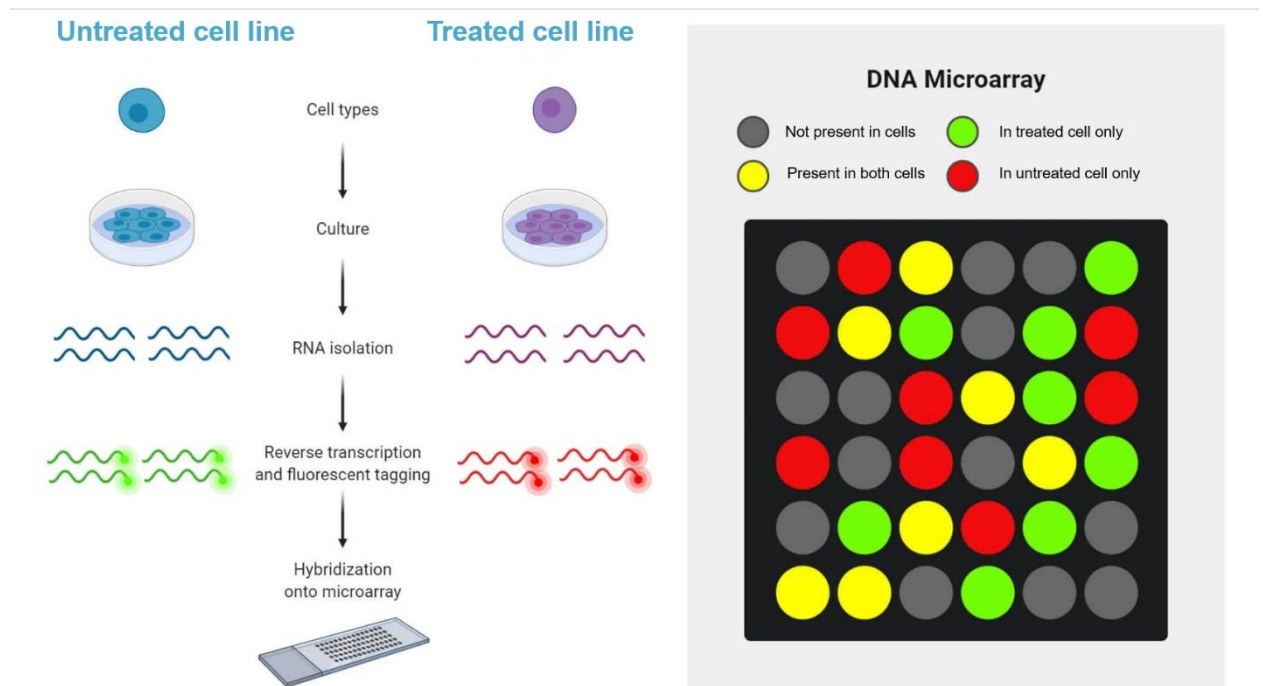


Figure 2. A general scheme of the methodology of gene expression microarray method in the concept of drug repositioning. RNA is extracted from treated and untreated cell lines and converted into fluorescent labeled cDNA. The cDNA is applied to a microarray chip, where it hybridizes with

immobilized probes. The chip is scanned to quantify fluorescent signals, revealing differences in gene expression levels between treated and untreated cells in downstream bioinformatics analysis.

Primary cell sorting

Primary cell sorting involves physically separating cells based on specific characteristics, such as size, shape, surface markers (usually surface proteins), combined with fluorescent labeling. This method is highly accurate and can isolate cells with high purity ¹⁵. However, primary cell sorting is also labor-intensive and can be relatively expensive, especially if specialized equipment is required ¹⁶. Some protein markers lack specificity, and employing them for cell sorting leads to contamination by inclusion of undesired cell types. In addition, some cells do not have unique surface protein markers ¹⁷. Besides, factors like cell stress, physical manipulation, or exposure to certain reagents during the sorting process can induce alterations in gene expression levels ¹⁸.

There are many methods of cell sorting, each with its own principles and applications. One of the most popular is Fluorescence-Activated Cell Sorting (FACS) employs fluorescent labels to bind to cell surface markers, enabling sorting based on distinct fluorescence patterns (Figure 3). FACS is a pivotal technique in modern cell biology and immunology that enables the isolation and purification of specific cell populations based on their distinct fluorescent properties. FACS combines flow cytometry and cell sorting technologies to analyze and sort individual cells from heterogeneous populations with remarkable precision. The process involves labeling target cells with fluorescent markers that bind to specific molecules, such as antibodies binding to cell surface proteins or dyes targeting intracellular components. As cells flow through a narrow stream in a flow cytometer, they pass through a laser beam that excites the fluorescent molecules, causing them to emit light at characteristic wavelengths. The emitted light is then detected and measured, providing information about the cells' fluorescence intensity and distribution. FACS takes this a step further by allowing the real-time sorting of cells based on their fluorescence properties. Electrically charged plates situated near the stream of cells generate an electric field that can deflect individual cells into separate collection tubes, effectively isolating the desired subpopulations for further analysis or experimentation. This technique has numerous applications in immunology, cancer research, stem cell studies, and various other fields, where the isolation of specific cell types is critical for advancing our understanding of cellular processes and disease mechanisms.

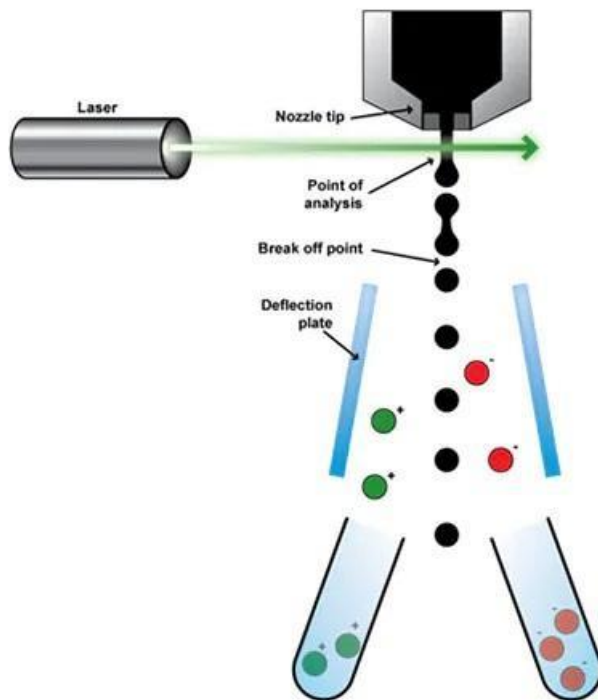


Figure 3. The principle of Fluorescence-Activated Cell Sorting. Each cell is channeled into a distinct droplet upon exiting the nozzle, and this droplet is electrically charged according to the cell's fluorescence. Employing deflection plates, the machine guides cells towards designated collection tubes. For instance, cells stained with Fluorescein isothiocyanate (FITC) and held within droplets receive a positive charge, prompting their attraction to the left and collection in corresponding tubes. Conversely, cells stained with PE and enclosed in droplets acquire a negative charge, propelling them towards the right for collection. Following sorting, the isolated cell populations are subjected to further analysis for verification, after which the sorted cells can be cultivated for subsequent experiments.

Figure adopted from <https://www.abcam.com>

Some other examples of primary cell sorting are: Magnetic-Activated Cell Sorting (MACS) uses magnetic beads coated with antibodies to separate cells with specific markers. Microfluidic Cell Sorting uses intricate channel networks to manipulate cells by physical properties like size. Dielectrophoresis (DEP) Sorting uses electric fields to sort cells based on their electrical characteristics. Pneumatic Cell Sorting relies on pressure differentials in microfluidic channels to sort cells gently. Finally, Density Gradient Centrifugation separates cells by their buoyant densities using centrifugal forces. These methods enable precise isolation of specific cell types, or sub-cellular structures for diverse applications¹⁹.

Single cell RNA sequencing

Single-cell RNA sequencing (scRNA-seq) is a cutting-edge molecular biology technique that enables the comprehensive analysis of gene expression profiles at the individual cell level within complex biological systems. Unlike traditional bulk RNA sequencing that averages gene expression across populations of cells, scRNA-seq dissects heterogeneity by isolating and sequencing the RNA content of individual cells. One of the widely used scRNA-seq platforms is 10xGenomics Chromium. This technology involves the isolation of single cells, the conversion of their RNA into complementary DNA (cDNA), and subsequent amplification and sequencing of this cDNA. By identifying the unique gene expression patterns of thousands of individual cells, scRNA-seq provides insights into cellular diversity, cell states, developmental trajectories, and disease mechanisms. The resulting data generates high-dimensional datasets that demand sophisticated computational analyses to uncover meaningful biological information, advancing our understanding of cellular biology with unprecedented granularity²⁰.

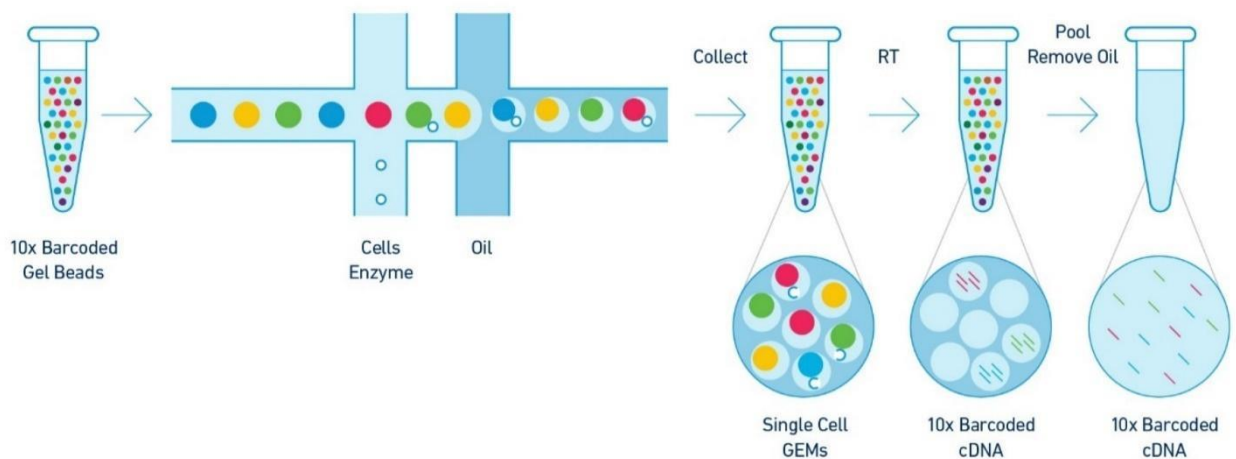


Figure 4. Schematic view of 10xGenomics single cell RNA sequencing Chromium

workflow. Individual cells (or nuclei) are combined with reagents and a solitary Gel Bead that contains uniquely barcoded oligonucleotides. These components are confined within nanoliter-sized Gel Bead in Emulsion (GEM) droplets using the GemCode™ Technology. Within each GEM, cellular lysis occurs, and barcoded reverse transcription is conducted on polyadenylated mRNA from each individual cell. This process takes place simultaneously within multiple GEMs. Following these steps, high-quality next-generation sequencing libraries are generated collectively in a single reaction and are compatible with Illumina sequencers. Image provided by 10x Genomics.

Based on the length and location of the start site of gene capturing in sequencing of cDNA, scRNA-seq can be distinguished to full-length, 3' End and 5' End. Full-Length scRNA-

seq captures entire RNA molecules. This method's ability to sequence full transcripts enables the identification of various gene isoforms, alternative splicing and intricate expression regulation. These advantages make this approach valuable for studying cellular heterogeneity and gene regulation complexity. However, it is technically demanding and resource-intensive due to the requirement for full-length cDNA synthesis and specialized sequencing equipment. On the other hand, 3' End scRNA-seq focuses on the 3' end of transcripts, allowing higher throughput and reduced complexity. While it provides less information about isoforms and splicing, it is cost-effective for profiling gene expression across many cells. Similarly, 5' End scRNA-seq targets the 5' end to uncover transcription start sites, promoter usage, and regulatory events. This method is valuable for understanding transcriptional initiation and regulatory elements while remaining cost-effective and suitable for large-scale studies²¹.

Different methods for single cell RNA profiling are available. Smart-seq methods, exemplified by Smart-seq2, involve full-length sequencing of individual cells, providing comprehensive insights into gene expression, including isoforms. This approach is ideal for full-length gene capturing. Drop-seq, on the other hand, employs microdroplets to encapsulate single cells with barcoded primers, capturing the 3' end of transcripts and enabling high-throughput gene expression profiling. Similarly, the 10x Genomics Chromium System utilizes gel bead-based partitioning within droplets to encapsulate cells and barcoded beads, predominantly capturing the 3' end of transcripts and facilitating high-throughput sequencing for gene expression analysis²⁰. Single cell profiling is also available through single-nuclei RNA sequencing (snRNA-seq) which use only nuclei instead of whole cells. Using snRNA-seq has advantages over scRNA-seq. This is because the process of tissue cryopreservation ruptures the cell membranes; however, nuclear membranes remain intact during the freeze–thaw cycle. Furthermore, it has been shown that the RNA-seq of single nuclei is highly representative of transcriptional profiles from the entire cells. However, since it focuses on nuclei, information about cytoplasmic RNA and cell morphology is lost²².

scRNA-seq/snRNA-seq is relatively expensive, and data analysis is complex and time-consuming²³. For instance, preprocessing the data involves quality control, filtering, and normalization. To handle the high-dimensional nature of the datasets, dimensionality reduction techniques become essential. Identifying cell-types within the data can be intricate due to rare cell-types and subtle differences. Finally, interpretation of the results necessitates expert knowledge to relate gene expression patterns to known functions or biological processes. Addressing these challenges involves employing advanced computational tools, statistical

methods, and domain expertise to gain deeper insights into cellular heterogeneity and spatial organization ²⁴.

Spatial transcriptomics

Spatial transcriptomics (ST) is an advanced molecular technique that provides a spatially contextualized understanding of gene expression within complex tissues, facilitating the mapping of molecular activity in its native spatial arrangement. While scRNA-seq focuses on analyzing gene expression profiles at the single-cell level, irrespective of spatial location, ST preserves the spatial context by capturing RNA molecules directly from tissue sections, utilizing spatially barcoded arrays. This enables the identification of gene expression profiles within a small number of cells while retaining their positional information. Using the advantages of scRNA-seq (higher resolution in distinguishing cells) and ST (preserving spatial information) one can combine these two methods to track the spatial location of single cells more accurately.

The 10x Genomics Visium is one of the widely used methods for ST begins with receiving the tissue, which can be either fresh frozen or embedded in optimal cutting temperature (OCT) blocks. Tissue sections are then cut and mounted onto specially designed Visium arrays. Reverse transcription reagents are added to the tissue sections, capturing polyadenylated mRNA. During this step, spatial barcodes are incorporated into the cDNA, preserving the spatial information of each transcript. The tissue sections are then dissociated from the arrays, and the cDNA is pooled and amplified. The resulting cDNA libraries are sequenced using next-generation sequencing technology. The sequencing data is processed using computational tools provided by 10x Genomics, aligning the reads to a reference genome while retaining the spatial information from the barcodes (Figure 5). This yields a spatially resolved gene expression profile, allowing researchers to map gene expression patterns within the tissue and gain insights into its molecular and cellular composition.

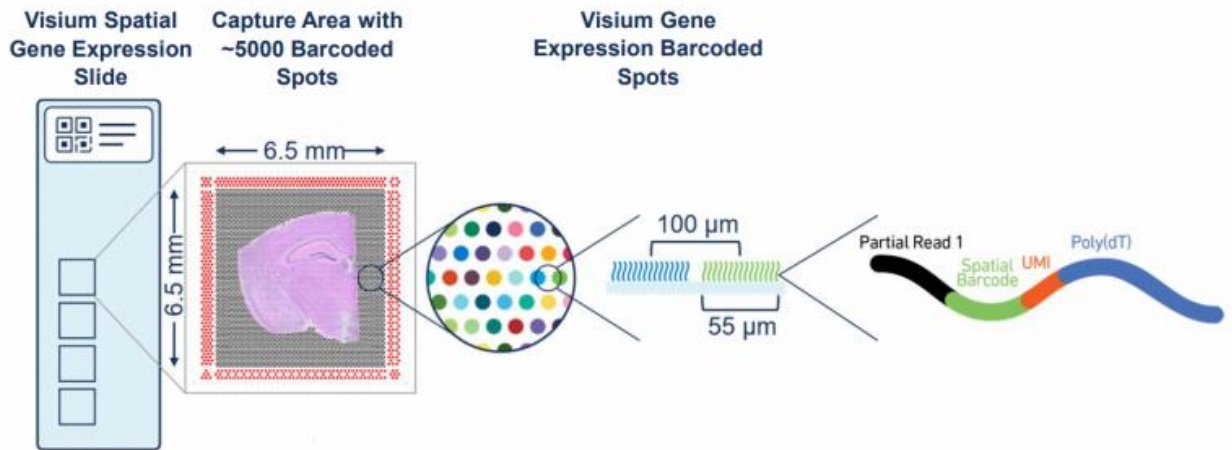


Figure 5. A schematic view of the 10x Genomics Visium ST protocol. It requires either fresh frozen tissue or OCT embedded tissue with intact morphology and high RNA quality, with a maximum tissue block size of 6.5x6.5 mm (the fresh frozen protocol is illustrated here). Each block contains around 5000 barcoded spots with 55 μ m diameter and 100 μ m distance from center to center between neighboring spots.

In 10x Genomics Visium Spatial Transcriptomics, each probe of barcoded spots (Figure 5) serves a specific purpose. Partial read 1 captures a portion of the gene or transcript sequence within the spot, revealing the genes present. The spatial barcode uniquely identifies the spot's location, allowing for spatial mapping of gene expression. Unique molecular identifiers (UMIs) distinguish between multiple copies of the same mRNA molecule, ensuring accurate quantification. Finally, the Poly(dT) sequence captures the polyadenylated tails of mRNA molecules, facilitating the isolation and sequencing of mRNA, all of which collectively enable the acquisition of spatially resolved gene expression data within a tissue or sample ²⁵.

ST encounters significant challenges including the difficulty of accurately detecting low-expression transcripts due to limited sensitivity, as well as the complex task of precisely identifying distinct cell types within tissues, especially closely related subtypes because of low spatial resolution. The intricate nature of ST datasets, characterized by high-dimensional information, demands advanced computational expertise for meaningful analysis and interpretation. Biases arising from sample preparation procedures, such as tissue handling and staining variability, can impact the quality and reproducibility of results. Moreover, ST experiments can be resource-intensive, involving substantial costs and longer timeframes, while offering lower throughput than traditional bulk RNA sequencing. The inherent heterogeneity of tissues poses a challenge in capturing the full cellular diversity comprehensively. Validating ST findings and ensuring reproducibility can be hindered by the absence of well-established

reference datasets for spatial transcriptomics, making it essential to work towards refining the technique to overcome these challenges and unlock its full potential. The challenges and solutions for different ST platforms will be discussed in more details in the following sections.

Challenges and solutions

PAPER I

Despite the insights provided by scRNA-seq and ST methods into cell-type heterogeneity, spatial distribution of cells and their gene expression, there are several challenges associated with data generation and analysis. Although ST is a powerful new technique for capturing patterns of spatial distribution of gene expression, it also has a drawback of its design. A 10x Genomics Visium Gene Expression slide used for ST experiments consists of two or four tissue-capture areas (6.5 mm x 6.5 mm), divided into 4992 spots, each 55 μm in diameter. Every spot contains oligonucleotide probes with unique sequence barcodes that encode spatial information of gene expression data. Due to their size, spots may encompass the expression profiles of several cells. Consequently, this diminishes the accuracy of distinguishing neighboring cell-types. While the Visium platform is one of the widely used approaches in spatial transcriptomics, there are other techniques and platforms available, such as Slide-seq²⁶, MERFISH²⁷, STARmap²⁸ and Xenium²⁹. Slide-seq also face similar limitations as Visium in capturing single-cell data. For both Visium and Slide-seq, this can be addressed by several methods including integration with other scRNA-seq datasets^{22,30}. MERFISH, STARmap and Xenium provide higher spatial resolution than Visium and Slide-seq. However, these three methods require the design and synthesis of specific oligonucleotide probes for targeted RNA detection, which can be challenging. Despite the complexity, MERFISH, STARmap and Xenium enable the identification and localization of individual cells and their gene expression profiles at subcellular levels. On the other hand, Visium and Slide-seq generally offer higher throughput compared to the MERFISH, STARmap and Xenium.

For different tissue types, 10x Genomics recommend specific tissue thicknesses to be studied using Visium platform. Choosing the thickness beyond this recommendation for different purposes such as easier sectioning, may have drawbacks. For instance, using hematoxylin and eosin (H&E) staining for tissue sections out of the recommended thickness range may lead to reduced quality and accuracy of the staining. This is particularly evident when dealing with thicker tissues, as the H&E stain struggles to evenly penetrate and distribute throughout the tissue. Consequently, uneven staining and poor visualization of cellular structures can occur, adversely affecting the accuracy and reliability of downstream analyses³¹.

Besides, thicker tissues are more challenging to permeabilize uniformly. This difficulty arises because mRNA extraction requires the use of chemicals to break down cell membranes and release the mRNA molecules, which can be less effective in penetrating thicker tissues. As a result, mRNA extraction efficiency may decrease with increasing tissue thickness, leading to lower quality data and reduced sensitivity in downstream analyses ³².

While selecting the recommended tissue section by 10x Genomics for certain tissue types has its benefits, there are also potential drawbacks. In addition to the previously recognized two-dimensional challenges associated with the larger size of spots compared to cells in ST, we have identified three-dimensional issues related to the volume of specific cell-types. For instance, when working on the tissue obtained from human brain, the recommended tissue thickness by 10xVisium is 10 μm while the size of neuronal nuclei in the human brain is around 20 μm . Consequently, the obtained ST data lack the full profile of neurons due to the incomplete incision of neuronal cells during the cryosectioning process of 10xVisium protocol. Increasing the tissue thickness to capture the full profile of neurons presents challenges related to staining and permeabilization. Therefore, it is essential to explore alternative approaches to address this issue effectively.

In spatial transcriptomics, consecutive tissue sections refer to the sequential slicing of a tissue sample into thin sections to study the spatial distribution of gene expression patterns. By examining gene expression profiles across multiple sections, researchers can gain insights into the distribution and arrangement of different cell-types within the tissue. The goal is to identify and characterize similar patterns of cell-types across the sections, which provides information about tissue architecture, cell-cell interactions, and the spatial context of gene expression patterns. Furthermore, there is another crucial shared characteristic among consecutive tissue sections that can help address the challenge of incomplete profiling of neurons in the human brain. Consecutive slices have complementary transcriptomics information that can be used to rectify neuronal profiles. To do so, the current analytical pipelines for ST data analysis needs to be improved to correct incomplete gene expression of cells before downstream analysis.

PAPER II

Chronic and acute myeloid leukemia (CML/AML) evade immune responses and induce immunosuppression. Patients with CML and AML exhibit dysfunctional immune cells (CD8+ T cells, NK cells) alongside suppressive myeloid cells and immunosuppressive regulatory T cells (Tregs). High Treg levels predict poorer treatment outcomes and shorter survival. While targeting Tregs directly for elimination seems beneficial, it's challenging and may cause

autoimmune adverse events. Identifying factors that drive Treg expansion in leukemias offers a potential alternative target. Extracellular vesicles (EVs), essential for intercellular communication, play a crucial role in immune modulation. In solid tumors, EVs inhibit T cell activity and promote Treg expansion. In myeloid neoplasms, including leukemias, EVs from leukemic cells enhance leukemic growth, drug resistance, and modify the bone marrow niche. AML-derived EVs have shown to inhibit CD8⁺ T cell function, yet the role of EVs in promoting Treg expansion has been less explored¹⁵.

To study the effect of EVs on Tregs, selecting the best approach is crucial. scRNA-seq is a powerful technique and provides high-resolution insights into cellular heterogeneity and can reveal previously unseen subpopulations within a cell type. However, one of the limitations of scRNA-seq is the relatively lower number of genes that can be captured per cell compared to total captured genes in bulk transcriptomics. This limitation arises from the technical challenges of amplifying and sequencing RNA from single cells, which can lead to incomplete coverage of the transcriptome³³. On the other hand, bulk transcriptomics has a high coverage for capturing gene profiles from the tissue but lacks the single cell resolution. To address this challenge, primary cell sorting can be combined with bulk transcriptomics to offer a comprehensive overview of gene expression in a specific cell type. This approach is valuable when investigating the global changes in gene expression within a cell population under a specific condition. Hence, the combination of primary cell sorting and bulk transcriptomics offers a more comprehensive view of the effect of EVs on Tregs compared to scRNA-seq.

PAPER III

The sequencing of the human genome has revolutionized our understanding of genetics and disease, allowing us to identify genetic risk factors for various conditions. However, having a list of genes associated with diseases is not enough to fully comprehend their roles and mechanisms. To gain deeper insights, researchers need to understand how genes function within cells. This requires perturbing genes and observing the resulting effects.

Connectivity Map (CMap) is a resource of the profiles of 3 cancer cell lines treated with 164 drugs using Affymetrix microarray. Before CMap, comprehensive resources for studying the effects of genetic and chemical perturbations on cells have been lacking. CMap improved our insights of how genes and chemicals influence cellular functions. It leverages a compendium of gene expression profiles, which are like snapshots of the activity levels of genes in different conditions. By comparing these profiles, researchers can uncover connections between genetic and chemical perturbations that might otherwise go unnoticed. Examples of

CMap use include the anthelmintic drug parabendazole as an inducer of osteoclast differentiation³⁴, celastrol as a leptin sensitizer³⁵, compounds targeting COX2 and ADRA2A as potential diabetes treatments³⁶, small molecules that mitigate skeletal muscular atrophy³⁷ and spinal muscular atrophy³⁸, and new therapeutic hypotheses for the treatment of inflammatory bowel disease³⁹ and cancer⁴⁰. However, the challenge lay in generating a comprehensive dataset due to the high cost of traditional gene expression techniques. To address this, LINCS L1000 was introduced, a high-throughput and cost-effective approach to gene expression profiling.

Library of integrated network-based cellular signatures (LINCS) currently comprises of over two million gene expression profiles of chemically perturbed human cell lines at a variety of time points and doses⁴¹. These data were produced using the LINCS L1000 method, which is an array-based transcriptomic profiling and measures a reduced representation of the transcriptome (~1000 genes called landmarks) and the rest of the transcriptome will be imputed. It has been shown that L1000 is highly reproducible, comparable to RNA sequencing, and suitable for computational inference of the expression levels of 81% of non-measured transcripts⁴². An instance of possible challenges for using the LINCS L1000 profiles is that providing such a huge amount of data requires computationally intensive approaches for data mining. Therefore, developing a systematic method to accurately and efficiently extract the relevant information of effect of drug treatments is crucial⁹.

Aims

The purpose of the presented study was:

- To assess the usefulness and benefits of different wet lab techniques in obtaining distinct cell types. This exploration encompassed the examination of various methods, such as vectorizing tissue sections and dissociating tissues as two separate methods (PAPER I), performing primary cell sorting (PAPER II), and using immortalized cells (PAPER III).
- To demonstrate the feasibility of utilizing transcriptomics methods to integrate and analyze data obtained from different wet lab techniques. For instance, aligning the information for each cell-type to the associated coordinates in vectorized tissue sections (ST), employing gene barcode labeling for cell profiling and categorization (scRNA-seq) (PAPER I), gene profiling in bulk derived from separated cell-types using primary cell sorting (PAPER II) and finally, using probe hybridization to capture a subset of genes in immortalized cell lines (PAPER III).
- To showcase the effectiveness of these transcriptomic approaches in enhancing spatial analysis (PAPER I), unraveling disease mechanisms (PAPER II), and identifying potential therapeutic strategies (PAPER III).

Material and Methods

In our three studies ^{9, 15, 22}, we addressed three different approaches to obtain cell-type resolution of transcriptomics profiles. These methods include scRNA-seq combined with ST to study human brain (PAPER I) ²², primary cell sorting and bulk transcriptomics downstream analysis on CML Tregs (PAPER II) ¹⁵, and array-based transcriptomics profiles generated by LINCS L1000 method (PAPER III) ⁴¹ from various immortalized cancer cell lines ⁹.

PAPER I: scRNA-seq and ST of postmortem brain samples

In terms of the scRNA-seq and ST methods, we utilized the modified 10x Genomics Visium Spatial Gene Expression method to analyze the profiles of consecutive sections from fresh-frozen brain tissues. Accordingly, we used the orbitofrontal neocortex (ON) and temporal neocortex (TN) samples from two subjects. Both subjects were considered healthy controls, as the aim was to investigate the reputability of our findings in this research in different subjects. Tissue specimens were provided by Harvard University and Massachusetts Alzheimer's Disease Research Center and all experimental procedures were conducted in accordance with Independent Bioethics Committee for Scientific Research at Medical University of Gdansk (consent No. NKBBN/564-108/2022). The brain-tissue slices were placed onto a Visium Gene Expression slide (10x Genomics) and fixed according to the 10x Genomics protocol (doc. CG000239 Rev. C). Next, the slides were divided into two via a piece of silicone gasket. Subsequently, we stained the tissue by two methods; hematoxylin and eosin, as well as combination hematoxylin and Congo red; the latter designed to detect possible amyloid deposits. We imaged the slides at 20x magnification using brightfield settings (Olympus cellSens Dimension software). Afterwards, the tissue was permeabilized, using conditions described in manufacturer protocol. The mRNA was released and bound to spatially barcoded capture probes on the slide. Next, cDNA was synthesized from captured mRNA, and sequencing libraries were prepared. Samples were loaded and pooled according to the protocol (doc. CG000239 Rev C) and sequenced in the standard Illumina pair-end constructs, using Illumina's NextSeq 550 System. Data pre-processing was done using SpaceRanger to first, convert (from BCL to fastq format using *spaceranger mkfastqc* function) and second, align the obtained profiles to reference genome (GRCH38) and calculate the number of gene counts (*spaceranger count* function). SpaceRanger is a computational tool designed to preprocess and analyze spatial transcriptomics data generated by the 10x Genomics Visium platform (Figure 5). It performs data organization and quality control to transform complex, raw data into a structured format, allowing researchers to examine gene expression patterns across different

regions of a biological sample. Essentially, it's a crucial tool for unraveling how genes are active in specific locations within tissues or samples. The raw gene counts were processed using Seurat (version 4.0.3). Seurat is an R package used in genomics to analyze individual cells. It helps researchers understand how genes behave in different cells, identify cell types, and gain insights into various biological processes, making it a valuable tool for studying diseases and development. The spatial transcriptomics data used in this study are available at the GEO data repository under the GSE184510 accession number and are accessible upon request.

We hypothesized that using single tissue sections of human brain in ST analysis, there is a source of batch effect which originates from incomplete neuronal gene profiles. This is because of the larger size of neuronal nuclei in comparison with the thickness of tissue slices and can be corrected using consecutive slices data integration (CSDI). To test this hypothesis, we used Seurat R package to perform dimensionality reduction (*RunPCA* and *RunUMAP* Seurat functions with default parameters), clustering (*FindClusters* Seurat function with default parameters) and label transferring from snRNA-seq datasets to ST before and after CSDI (to see the parameters and utilized functions see ²²). Dimensionality reduction in Seurat is a computational technique that simplifies intricate single-cell genomics data by compressing it into a lower-dimensional format while preserving essential information. It enables researchers to create visualizations and plots that reveal hidden patterns, identify distinct cell types, and highlight critical genes influencing cellular behaviors. This streamlined representation facilitates the understanding of complex biological processes, such as disease mechanisms or tissue development, by providing a more interpretable view of gene expression within individual cells, ultimately enhancing the insights gained from single-cell genomics datasets. For label transferring and deconvolution of ST spots, we used publicly available snRNA-seq dataset (deposited in GEO with GSE129308 accession number) as RNA-seq of single nuclei is highly representative of transcriptional profiles from the entire cells ⁴³. We introduced CSDI as a complementary method for data correction with the potential of removing unknown batch effects.

PAPER II: primary cell sorting of CML Tregs

Plasma (source of primary EVs) was obtained from whole blood of 10 leukemic (7 CML and 3 AML; before starting treatment and at diagnosis) patients. EVs from plasma were first isolated by size exclusion chromatography. Then EVs were further isolated using differential ultracentrifugation protocol, cell culture conditioned medium was first centrifuged at 160xg (5 minutes), 320xg (5minutes) and 1300xg (20 minutes), to deplete cells and cellular debris.

Further, high-speed ultracentrifugation steps were performed: 10.000xg for 40 minutes to deplete medium/large EVs, 100.000xg for 90 minutes to pellet small EVs, after which EVs were resuspended in PBS and washed by another ultracentrifugation step, for 90 minutes at 100.000xg. Ultracentrifugation was performed using 45Ti fixed-angle rotor and Optima XPN-100 ultracentrifuge (Beckman Coulter).

Human lymphocytes were obtained from buffy coats of healthy donors (different donor each experiment) from the Regional Center for Blood Donation and Blood Care in Warsaw, Poland (in accordance with the Declaration of Helsinki and Polish regulations). Peripheral blood mononuclear cells (PBMCs) were isolated by density gradient centrifugation (Lymphoprep, STEMCELL). T cells, including Tregs, were sorted using BD FACS Aria II.

Regulatory T cells (Tregs) from *ex vivo* cultures, treated with CML EVs (from three patients) and nontreated (from three patients), were sorted to obtain pure population of viable, CD4⁺CD25^{hi}CD127^{lo} Tregs. Viability of sorted cells was verified at 97%. Sorted cells were washed and frozen in TRI Reagent (Sigma-Aldrich). RNA was isolated using Total RNA Mini column purification kit (A&A Biotechnology). Sequencing libraries were prepared using NEB Next Ultra II Directional RNA library prep kit for Illumina. Samples were sequenced using Illumina NextSeq 500, 75-bp single-end reads (Genomics Core Facility at EMBL, Heidelberg). The sequenced reads were aligned to hg38 genome using Hisat2 (version 2.1.0) with default settings. The numbers of reads aligned to each gene and the differential expression were computed with python HTSeq script (version 0.11.2) and DESeq2 R package (version 1.28.1), respectively. Genes that had significant (adj. p-value < 0.05) changes in their expression levels (log-fold-change > 1) were called differentially expressed. The Gene Ontology analysis was performed with Bioconductor package ClusterProfiler (version 3.16.1). Analysis of transcription factor binding motifs (TFBM) was performed using PSCAN software, by referencing -950 to +50 bp regions of DEGs to JASPAR 2018_NR database. JASPAR is a well-known open-access database of transcription factor binding profiles. It provides information about the DNA binding preferences of transcription factors, which are proteins that regulate gene expression. Researchers use JASPAR to study how transcription factors interact with specific DNA sequences, which is crucial for understanding gene regulation. The gene expression data are available under GEO number GSE180883.

PAPER III: developing pipeline for data extraction from LINCS L1000 database

In the third project, we benefited from the data generated by LINCS L1000 platform. Subramanian et al. 2017⁴¹ as part of the NIH LINCS Consortium, developed a new, low-cost, high throughput reduced representation expression profiling method that is termed L1000. With the LINCS L1000 platform, now more than 2 million profiles from hundreds of cells and thousands of drug perturbagens are publicly available⁴⁴. In this method, a reduced representation of the transcriptome (~1000 genes called landmarks) are measured and the rest of the transcriptomics are imputed. Apart from its high reproducibility and comparability to RNA sequencing, Subramanian et al. 2017 showed the method's potential in discovering the mechanism of action of small molecules, functionally annotating genetic variants of disease genes, and providing valuable insight for clinical trials⁴¹.

From the LINCS L1000 database, we aimed to retrieve normalized gene expression profiles of landmark genes and imputed transcripts for four different cancer cell lines including A549, HEPG2, MCF7, and HT-29, treated with drug in comparison to control (DMSO). To achieve this objective, we utilized the Slinky R package (version 1.8.0) to parse normalized gene expression profiles of landmark genes and imputed transcripts for a total of 12,328 genes and over 900 drugs in each of the four cancer cell lines.⁴⁵ In our work, we extracted the data for both control (treated with DMSO) and experimental conditions (treated with various drugs) with the highest standard dose (10 μ m) and longest time points (24 h). Log fold changes (LFC) were computed through the NumPy library (version 1.19.1) in Python 3.7.6. To avoid undefined LFC values due to division by zero or log₂ transformation of non-positive numbers, one was added to gene expression values for both treatment and control before transformation. As a result, we produced LFCs for four different cancer cell lines across about 12,000 genes and 900 drugs. Using the LFC matrixes as an input, we developed a method to extract the drug information which can deregulate particular genes of interest (Figure 1 in PAPER III). To test our pipeline, we investigated the drugs that can decrease the level of Heparan Sulfate (HS) and Chondroitin Sulfate (CS) at the surface of all four cancer cell lines to improve the performance of anticancer peptides.

To find a reference statistical method to compute gene-gene correlations, three methods (Spearman (SP), Pearson (PE), and Kendall tau (KE)) were evaluated using the A549 LFC matrix. Top 100 and 500 co-expressed gene pairs were subjected to functional analysis. The method which produces the most enriched terms in both Gene Ontology (GO) and KEGG pathway analysis was chosen as the reference statistical method. The reference statistical

method was applied to all four cancer cell lines. The correlations with experimentally validated HS and CS genes were extracted. As the first filtration step, only common co-expressed genes with experimentally supported HS and CS genes in all four gene-gene correlation data frames (corresponding to four cancer cells) were considered. The expression profiles of these genes were extracted from LFC matrixes of all four cancer cell lines. Drugs that resulted in significant upregulation or downregulation of the majority of the selected genes were categorized into two separate data frames: one for downregulated genes and another for upregulated genes. As the second filtration step, drugs that jointly led to up or down-regulation of most selected genes in all four cancer cells were extracted. These drugs were proposed for down/upregulation of HS and CS for wet-lab validations.

Results

PAPER I

By employing CSDI (Consecutive Slices Data Integration) in spatial transcriptomics (ST) profiles of successive tissue sections in the human brain, the spot clustering and label transferring from the single-nucleus RNA sequencing (snRNA-seq) dataset can be enhanced. This improvement proposed by us allows more precise and biologically significant outcomes.

Stuart et al. 2019⁴⁶ developed the CSDI to correct the transcriptomic profiles of consecutive slices using anchors representing spots with similar gene expression profile from two consecutive slices. This is used to pair spots from the two slices. At the same time, the transcriptomic differences between pairs of spots in anchors are used to correct datasets from both consecutive sections. We conducted spatial gene expression analysis in human postmortem, fresh frozen tissue sections. Two anatomical regions, the Orbitofrontal Neocortex (ON) and the Temporal Neocortex (TN) from two adult male donors were investigated. From each region of both subjects, one pair of consecutive slices (eight slices in total) were prepared (Figure 1 in PAPER I). We performed our data analysis in two parallel approaches. First, by considering each tissue section of consecutive slices as an independent object (Figure 2 in PAPER I) and second, by taking into account that consecutive slices may have complementary information and needs to be integrated before downstream analysis (Figure 4 in PAPER I). We performed the integration using CSDI method. Hence, we could compare the results obtained from these two approaches with each other and with histological and morphological information of human cerebral cortex as reference.

We performed spot clustering to classify the spots with similar expression profiles, and distinguish distinct cellular layers. In both above-mentioned approaches (first: before and second: after CSDI), we could illustrate the grey matter (GM) and white matter (WM) which was consistent with histological images, however the pattern of cluster between consecutive slices was inconsistent in the first method. Given the expected architectural similarity between two successive slices of the cerebral cortex, the question was raised that where the inconsistency between the pattern of clusters originates from.

To compare these two approaches further, and to better understand the identified brain layers, we integrated the measured expression profiles of ST from both approaches (before and after CSDI) with a previously described snRNA-seq dataset⁴⁷. As single nucleus profiles

contain greater number of genes than in our ST profiles, the integration of these two datasets (ST and snRNA-seq) allowed us to perform the spot annotation more precisely. Using predefined cell-type annotations in snRNA-seq—including oligodendrocytes, astrocytes, and neurons—the ST spots were labeled. Before CSDI, we could not confidently annotate neurons in GM, which is incompatible with histological image (Figure 2 in PAPER I), while after CSDI, neurons were located in the GM in all eight tissue sections (Figure 4 in PAPER I). Besides, more neuronal layers in GM could be unveiled using the second approach which is consistent with morphology of human cerebral cortex.

Cell bodies of neurons are mainly found in the GM. However, in the first approach, during our label transferring, the spots marked as neurons received weak probability values in the GM. Besides, the pattern of spot clusters in this region (GM) were mainly inconsistent between consecutive slices. This is an important concern, which led us to hypothesize that using information from a single section of tissue may lead to inaccurate interpretation of clusters and cell-types. The differences between annotations obtained for the spots before and after CSDI can be attributed to the fact that the size of neuronal nuclei is larger than the thicknesses of the tissue sections used in the ST protocol. Accordingly, a single slice will capture incomplete transcriptomic neuronal context. CSDI provides a robust means of rectification of this misinterpretation. Hence, the corrected signals of all types of nuclei can be obtained. Consequently, the label transferring from snRNA-seq to ST is made consistent with the histological findings only after CSDI. Ultimately, one can study the spatial distribution of different cell-types more precisely.

PAPER II

Using bulk transcriptomics data from primary sorted cells in myeloid leukemia, we measured the average gene expression across regulatory T cells (Tregs). We observed that leukemic extracellular vesicles (EVs) expand pro-leukemic FOXP3+ Tregs which result in evading immune system surveillance, inducing immunosuppression, inferior response to chemotherapy, leukemia relapse and shorter survival.

In tumors, immunosuppressive milieu can induce expression of Treg-specific transcription factor Foxp3 in non-regulatory, CD4⁺CD25⁻ conventional T cells and turn them into CD25^{hi}Foxp3⁺ induced regulatory T cells⁴⁸. Therefore, we studied the impact of leukemic

EVs on Foxp3 induction. We performed ex vivo cultures of purified (sorted) human CD4⁺CD25^{hi}CD127^{lo} Tregs together with EVs released by CML-K562 cells (CML EVs).

Analysis of Tregs by RNA sequencing revealed significant remodeling of the transcriptome and elevated expression of 356 genes due to treatment with CML EVs, as well as influence on biological processes, such as RNA metabolism. We analyzed genes described as characteristic for Tregs in cancer⁴⁹⁻⁵⁰ and observed a visible trend of upregulated expression for *CCR4*, *TFRC*, *TNFRSF1B* (encoding *TNFR2*), *ENTPD1* (*CD39*), *TNFRSF8* (*CD30*), *IL1R1*, *HAVCR2* (*TIM-3*), and *TGFBI* (Figure 4 and supplementary Figure 11 in PAPER II). However, in most cases, the difference was not statistically significant, therefore we additionally verified these observations on protein level. Analysis of transcription factor-binding motifs (TFBMs) of differentially expressed genes identified several transcription factors potentially engaged in modulation of Tregs by leukemic EVs, such as *EGR1*, *EGR3*, *ZBTB7A* (*LRF*), *E2F4*, or *TFDP1*. Overall, RNA sequencing further signified that leukemic EVs affect Treg, by global remodeling of gene expression, including upregulation of genes responsible for immunosuppressive function. Analysis of transcription factor-binding motifs pinpointed a set of transcription factors that modulate these changes in Tregs and maybe relevant for immunosuppression in myeloid leukemias.

PAPER III

We developed a drug repositioning pipeline to analyze array-based transcriptomics data generated for hundreds of cancer/normal cell lines treated with thousands of drugs in the LINCS L1000 project. We used this method to propose the drugs which can promote the broad utilization of anticancer peptides (ACPs).

The LINCS L1000 project as a new gene expression profiling method has provided an excellent opportunity to study the mechanism of action of small molecules, functionally annotate genetic variants of disease genes, and inform clinical trials by collecting gene expression profiles for thousands of drugs at a variety of time points, doses, and cell lines⁴². Taking the massive amount of data produced by LINCS L1000 into account, parsing the data would be computationally intensive. To address this issue, various methods have been developed^{45,51} In this study, we developed a new method to extract the desired information of the effect of drug treatments in gene level (Figure 1 in PAPER III). In order to validate our method, we tackled an existing issue in the field of ACPs. The efficacy of the positively-charged

ACPs, as an alternative/complementary strategy to conventional chemotherapy, is inhibited by elevated levels of negatively-charged cell-surface components, such as negatively-charged HS and CS, which trap the peptides and prevent their contact with the cell membrane and consequent pore formation and cell lysis⁹. Using our method, we proposed the FDA approved drugs which can promote the broad utilization of anticancer peptides by decreasing the level of HS and CS.

To do so, we determined LFC values for drug compared to control for each gene in each cancer cell line. LFC describes how much expression values change between these two conditions. To find the best statistical method which discovers the most meaningful co-expression correlations in our datasets, the top 100 and 500 pairs of co-expressed genes from A549 LFC matrix were determined by SP, PE, and KE methods and were subjected to GO and KEGG pathway analysis (Table 2 in PAPER III). PE outperformed SP and KE based on the number of significantly enriched terms (adjusted p-value < 0.05). Concerning the top 100 co-expressed gene pairs, both KEGG and GO pathway analysis depict better KE performance compared with PE and SP. However, looking at the top 500 co-expressed gene pairs, PE depicts considerably more enriched terms and generally more involved genes in enriched terms in both KEGG and GO pathway analysis. Hence, we chose PE as the reference statistical method to compute gene-gene correlations for the remaining cancer cell lines (i.e., HEPG2, HT29, and MCF7).

Considering those gene-gene correlations that appeared in all four cell lines and involving genes already known in the literature to be associated with HS and CS⁵², top 10 correlations with each HS and CS lab-validated genes were chosen. Assuming that the in silico-driven gene-gene associations are common between all four cancer cell lines, biological correlations between these genes could be expected.

To investigate the pathways related to selected genes, we conducted KEGG and GO analysis (Figure 3 in PAPER III). GO analysis identified Golgi lumen, which is significant as this is where EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyses the synthesis of HS⁵³. In addition to this, collagen-containing extracellular matrix, where HS and CS are available⁵⁴ was the highest enriched term. On the other hand, KEGG pathway analysis revealed glycosaminoglycan biosynthesis of HS and CS as the top two enriched terms.

The expression profile of selected HS and CS co-expressed genes were extracted from LFCs of all four cancer cells. The heatmap was used to identify the drugs that cause down-regulation of these genes in all four cancer cell lines. Thus, a list of potential drugs which can decrease the level of HS and CS at the surface of cancer cell lines was proposed for further wet-lab validations (Table 4 in PAPER III).

Conclusion

Cell-type resolution transcriptomic methods, such as single-cell RNA sequencing, spatial transcriptomics, bulk transcriptomics from primary sorted cells, and microarray profiles from immortalized cell lines, are powerful tools for studying gene expression patterns of cell-types and their functional implications. These techniques have provided valuable insights into cellular heterogeneity, spatial organization, and the molecular mechanisms underlying various biological processes and diseases. It is important to acknowledge that each method presents its own set of advantages and challenges that need to be considered in their application.

In the analysis of human brain tissue, we identified a source of batch effect which was incomplete ST profiles of cells, i.e., neurons, larger than thickness of tissue sections. We suggested a method, named CSDI, to rectify the incomplete profiles and consequently improve spot clustering and label transferring. In the bulk transcriptomics study of primary sorted myeloid leukemia cells, we explored the impact of leukemic extracellular vesicles on regulatory T cells and gained insights into immunosuppression mechanisms. Furthermore, we developed a drug repositioning pipeline using microarray-based LINCS L1000 data to identify potential drugs that could enhance the utilization of anticancer peptides. These findings highlight the importance of employing advanced computational tools, statistical methods, and domain expertise to fully leverage the capabilities of transcriptomics in unraveling the complexity of gene expression and its implications in biological systems.

REFERENCES

1. Cooper, G.; Adams, K., *The cell: a molecular approach*. Oxford University Press: 2022.
2. Kulkarni, A.; Anderson, A. G.; Merullo, D. P.; Konopka, G., Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Current opinion in biotechnology* **2019**, *58*, 129-136.
3. Victorelli, S.; Passos, J. F., Telomeres and Cell Senescence - Size Matters Not. *EBioMedicine* **2017**, *21*, 14-20.
4. Wang, Y.; Chen, S.; Yan, Z.; Pei, M., A prospect of cell immortalization combined with matrix microenvironmental optimization strategy for tissue engineering and regeneration. *Cell & Bioscience* **2019**, *9* (1), 7.
5. Jiang, M.; Min, Y.; Debusk, L.; Fernandez, S.; Strand, D. W.; Hayward, S. W.; Lin, P. C., Spontaneous immortalization of human dermal microvascular endothelial cells. *World journal of stem cells* **2010**, *2* (5), 114-20.
6. Goodspeed, A.; Heiser, L. M.; Gray, J. W.; Costello, J. C., Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. *Molecular cancer research : MCR* **2016**, *14* (1), 3-13.
7. Kaur, G.; Dufour, J. M., Cell lines: Valuable tools or useless artifacts. *Spermatogenesis* **2012**, *2* (1), 1-5.
8. Shen, Y.; Schmidt, B. U. S.; Kubitschke, H.; Morawetz, E. W.; Wolf, B.; Käs, J. A.; Losert, W., Detecting heterogeneity in and between breast cancer cell lines. *Cancer convergence* **2020**, *4* (1), 1.
9. Mohammadi, E.; Tahmoorespur, M.; Benfeitas, R.; Altay, O.; Javadmanesh, A.; Lam, S.; Mardinoglu, A.; Sekhavati, M. H., Improvement of the performance of anticancer peptides using a drug repositioning pipeline. *Biotechnology journal* **2022**, *17* (1), 2100417.
10. Lee, T. I.; Johnstone, S. E.; Young, R. A., Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nature protocols* **2006**, *1* (2), 729-48.
11. Trevino, V.; Falciani, F.; Barrera-Saldaña, H. A., DNA Microarrays: a Powerful Genomic Tool for Biomedical and Clinical Research. *Molecular Medicine* **2007**, *13* (9), 527-541.
12. Slonim, D. K.; Yanai, I., Getting started in gene expression microarray analysis. *PLoS computational biology* **2009**, *5* (10), e1000543.
13. Sutandy, F. X.; Qian, J.; Chen, C. S.; Zhu, H., Overview of protein microarrays. *Current protocols in protein science* **2013**, *Chapter 27* (1), Unit 27.1.
14. Xiang, C. C.; Chen, Y., cDNA microarray technology and its applications. *Biotechnology Advances* **2000**, *18* (1), 35-46.
15. Swatler, J.; Turos-Korgul, L.; Brewinska-Olchowik, M.; De Biasi, S.; Dudka, W.; Le, B. V.; Kominek, A.; Cyranowski, S.; Pilanc, P.; Mohammadi, E.; Cysewski, D.; Kozłowska, E.; Grabowska-Pyrzewicz, W.; Wojda, U.; Basak, G.; Mieczkowski, J.; Skorski, T.; Cossarizza, A.; Piwocka, K., 4-1BBL-containing leukemic extracellular vesicles promote immunosuppressive effector regulatory T cells. *Blood Advances* **2022**, *6* (6), 1879-1894.

16. Rahmanian, N.; Bozorgmehr, M.; Torabi, M.; Akbari, A.; Zarnani, A.-H., Cell separation: Potentials and pitfalls. *Preparative Biochemistry & Biotechnology* **2017**, *47* (1), 38-51.
17. Gundry, R. L.; Boheler, K. R.; Van Eyk, J. E.; Wollscheid, B., A novel role for proteomics in the discovery of cell-surface markers on stem cells: Scratching the surface. *Proteomics. Clinical applications* **2008**, *2* (6), 892-903.
18. Box, A.; DeLay, M.; Tighe, S.; Chittur, S. V.; Bergeron, A.; Cochran, M.; Lopez, P.; Meyer, E. M.; Saluk, A.; Thornton, S.; Brundage, K., Evaluating the Effects of Cell Sorting on Gene Expression. *Journal of biomolecular techniques : JBT* **2020**, *31* (3), 100-111.
19. Witek, M. A.; Freed, I. M.; Soper, S. A., Cell Separations and Sorting. *Analytical chemistry* **2020**, *92* (1), 105-131.
20. Jovic, D.; Liang, X.; Zeng, H.; Lin, L.; Xu, F.; Luo, Y., Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical and Translational Medicine* **2022**, *12* (3), e694.
21. Hahaut, V.; Pavlinic, D.; Carbone, W.; Schuierer, S.; Balmer, P.; Quinodoz, M.; Renner, M.; Roma, G.; Cowan, C. S.; Picelli, S., Fast and highly sensitive full-length single-cell RNA sequencing using FLASH-seq. *Nature Biotechnology* **2022**, *40* (10), 1447-1451.
22. Mohammadi, E.; Chojnowska, K.; Bieńkowski, M.; Kostecka, A.; Koczkowska, M.; Żmijewski, M. A.; Jąkowski, M.; Ingelsson, M.; Filipowicz, N.; Olszewski, P.; Davies, H.; Wierzbicka, J. M.; Hyman, B. T.; Dumanski, J. P.; Piotrowski, A.; Mieczkowski, J., Size matters: the impact of nucleus size on results from spatial transcriptomics. *Journal of Translational Medicine* **2023**, *21* (1), 270.
23. Hwang, B.; Lee, J. H.; Bang, D., Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine* **2018**, *50* (8), 1-14.
24. Du, J.; Yang, Y.-C.; An, Z.-J.; Zhang, M.-H.; Fu, X.-H.; Huang, Z.-F.; Yuan, Y.; Hou, J., Advances in spatial transcriptomics and related data analysis strategies. *Journal of Translational Medicine* **2023**, *21* (1), 330.
25. Li, X.; Wang, C.-Y., From bulk, single-cell to spatial RNA sequencing. *International Journal of Oral Science* **2021**, *13* (1), 36.
26. Rodriques, S. G.; Stickels, R. R.; Goeva, A.; Martin, C. A.; Murray, E.; Vanderburg, C. R.; Welch, J.; Chen, L. M.; Chen, F.; Macosko, E. Z., Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **2019**, *363* (6434), 1463-1467.
27. Chen, K. H.; Boettiger, A. N.; Moffitt, J. R.; Wang, S.; Zhuang, X., Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **2015**, *348* (6233), aaa6090.
28. Wang, X.; Allen, W. E.; Wright, M. A.; Sylwestrak, E. L.; Samusik, N.; Vesuna, S.; Evans, K.; Liu, C.; Ramakrishnan, C.; Liu, J.; Nolan, G. P.; Bava, F. A.; Deisseroth, K., Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **2018**, *361* (6400).
29. Henley, R.; Rapticavoli, N.; Janesick, A.; Shelansky, R.; Kim, A.; Hensel, J.; Meschi, F.; Farahani, N.; Kumar, V.; Qian, X., 95 Characterization of human breast cancer tissue with the Xenium In Situ platform reveals a novel marker for invasiveness. *BMJ Specialist Journals*: 2022.

30. Dong, R.; Yuan, G.-C., SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biology* **2021**, *22* (1), 145.
31. Li, Y.; Li, N.; Yu, X.; Huang, K.; Zheng, T.; Cheng, X.; Zeng, S.; Liu, X., Hematoxylin and eosin staining of intact tissues via delipidation and ultrasound. *Sci Rep* **2018**, *8* (1), 12259.
32. Williams, C. G.; Lee, H. J.; Asatsuma, T.; Vento-Tormo, R.; Haque, A., An introduction to spatial transcriptomics for biomedical research. *Genome Medicine* **2022**, *14* (1), 68.
33. Hegenbarth, J.-C.; Lezsoche, G.; De Windt, L. J.; Stoll, M., Perspectives on Bulk-Tissue RNA Sequencing and Single-Cell RNA Sequencing for Cardiac Transcriptomics. *Frontiers in Molecular Medicine* **2022**, *2*.
34. Brum, A. M.; van de Peppel, J.; Van Der Leije, C. S.; Schreuders-Koedam, M.; Eijken, M.; van der Eerden, B. C.; van Leeuwen, J. P., Connectivity Map-based discovery of parabendazole reveals targetable human osteogenic pathway. *Proceedings of the National Academy of Sciences* **2015**, *112* (41), 12711-12716.
35. Liu, J.; Lee, J.; Hernandez, M. A. S.; Mazitschek, R.; Ozcan, U., Treatment of obesity with celastrol. *Cell* **2015**, *161* (5), 999-1011.
36. Zhang, M.; Luo, H.; Xi, Z.; Rogaeva, E., Drug repositioning for diabetes based on 'omics' data mining. *PloS one* **2015**, *10* (5), e0126082.
37. Dyle, M. C.; Ebert, S. M.; Cook, D. P.; Kunkel, S. D.; Fox, D. K.; Bongers, K. S.; Bullard, S. A.; Dierdorff, J. M.; Adams, C. M., Systems-based discovery of tomatidine as a natural small molecule inhibitor of skeletal muscle atrophy. *Journal of Biological Chemistry* **2014**, *289* (21), 14913-14924.
38. Farooq, F.; Balabanian, S.; Liu, X.; Holcik, M.; MacKenzie, A., p38 Mitogen-activated protein kinase stabilizes SMN mRNA through RNA binding protein HuR. *Human molecular genetics* **2009**, *18* (21), 4035-4045.
39. Dudley, J. T.; Sirota, M.; Shenoy, M.; Pai, R. K.; Roedder, S.; Chiang, A. P.; Morgan, A. A.; Sarwal, M. M.; Pasricha, P. J.; Butte, A. J., Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Science translational medicine* **2011**, *3* (96), 96ra76-96ra76.
40. Singh, A. R.; Joshi, S.; Zulcic, M.; Alcaraz, M.; Garlich, J. R.; Morales, G. A.; Cho, Y. J.; Bao, L.; Levy, M. L.; Newbury, R., PI-3K inhibitors preferentially target CD15+ cancer stem cell population in SHH driven medulloblastoma. *PloS one* **2016**, *11* (3), e0150836.
41. Subramanian, A.; Narayan, R.; Corsello, S. M.; Peck, D. D.; Natoli, T. E.; Lu, X.; Gould, J.; Davis, J. F.; Tubelli, A. A.; Asiedu, J. K.; Lahr, D. L.; Hirschman, J. E.; Liu, Z.; Donahue, M.; Julian, B.; Khan, M.; Wadden, D.; Smith, I. C.; Lam, D.; Liberzon, A.; Toder, C.; Bagul, M.; Orzechowski, M.; Enache, O. M.; Piccioni, F.; Johnson, S. A.; Lyons, N. J.; Berger, A. H.; Shamji, A. F.; Brooks, A. N.; Vrcic, A.; Flynn, C.; Rosains, J.; Takeda, D. Y.; Hu, R.; Davison, D.; Lamb, J.; Ardlie, K.; Hogstrom, L.; Greenside, P.; Gray, N. S.; Clemons, P. A.; Silver, S.; Wu, X.; Zhao, W. N.; Read-Button, W.; Wu, X.; Haggarty, S. J.; Ronco, L. V.; Boehm, J. S.; Schreiber, S. L.; Doench, J. G.; Bittker, J. A.; Root, D. E.; Wong, B.; Golub, T. R., A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **2017**, *171* (6), 1437-1452.e17.
42. Subramanian, A.; Narayan, R.; Corsello, S. M.; Peck, D. D.; Natoli, T. E.; Lu, X.; Gould, J.; Davis, J. F.; Tubelli, A. A.; Asiedu, J. K., A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **2017**, *171* (6), 1437-1452. e17.

43. Grindberg, R. V.; Yee-Greenbaum, J. L.; McConnell, M. J.; Novotny, M.; O'Shaughnessy, A. L.; Lambert, G. M.; Araúzo-Bravo, M. J.; Lee, J.; Fishman, M.; Robbins, G. E., RNA-sequencing from single nuclei. *Proceedings of the National Academy of Sciences* **2013**, *110* (49), 19802-19807.
44. Kwee, I.; Martinelli, A.; Khayal, L. A.; Akhmedov, M., metaLINCS: an R package for meta-level analysis of LINCS L1000 drug signatures using stratified connectivity mapping. *Bioinformatics Advances* **2022**, *2* (1).
45. Kort, E. J.; Jovinge, S., Streamlined analysis of LINCS L1000 data with the slinky package for R. *Bioinformatics* **2019**, *35* (17), 3176-3177.
46. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W. M., 3rd; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R., Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177* (7), 1888-1902.e21.
47. Zhang, Y.; Fung, I. T. H.; Sankar, P.; Chen, X.; Robison, L. S.; Ye, L.; D'Souza, S. S.; Salinero, A. E.; Kuentzel, M. L.; Chittur, S. V., Depletion of NK cells improves cognitive function in the Alzheimer disease mouse model. *The Journal of Immunology* **2020**, *205* (2), 502-510.
48. Le Dieu, R.; Taussig, D. C.; Ramsay, A. G.; Mitter, R.; Miraki-Moud, F.; Fatah, R.; Lee, A. M.; Lister, T. A.; Gribben, J. G., Peripheral blood T cells in acute myeloid leukemia (AML) patients at diagnosis have abnormal phenotype and genotype and form defective immune synapses with AML blasts. *Blood, The Journal of the American Society of Hematology* **2009**, *114* (18), 3909-3916.
49. Alvisi, G.; Brummelman, J.; Puccio, S.; Mazza, E. M.; Tomada, E. P.; Losurdo, A.; Zanon, V.; Peano, C.; Colombo, F. S.; Scarpa, A., IRF4 instructs effector Treg differentiation and immune suppression in human cancer. *The Journal of clinical investigation* **2020**, *130* (6), 3137-3150.
50. Plitas, G.; Konopacki, C.; Wu, K.; Bos, P. D.; Morrow, M.; Putintseva, E.; Chudakov, D.; Rudensky, A., Regulatory T cells exhibit distinct features in human breast cancer. *Immunity* **2016**, *45* (5), 1122-1134.
51. Duan, Q.; Flynn, C.; Niepel, M.; Hafner, M.; Muhlich, J. L.; Fernandez, N. F.; Rouillard, A. D.; Tan, C. M.; Chen, E. Y.; Golub, T. R., LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. *Nucleic acids research* **2014**, *42* (W1), W449-W460.
52. Fadnes, B.; Rekdal, Ø.; Uhlin-Hansen, L., The anticancer activity of lytic peptides is inhibited by heparan sulfate on the surface of the tumor cells. *BMC cancer* **2009**, *9*, 1-13.
53. McCormick, C.; Duncan, G.; Goutsos, K. T.; Tufaro, F., The putative tumor suppressors EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyzes the synthesis of heparan sulfate. *Proceedings of the National Academy of Sciences* **2000**, *97* (2), 668-673.
54. Sarrazin, S.; Lamanna, W. C.; Esko, J. D., Heparan sulfate proteoglycans. *Cold Spring Harbor perspectives in biology* **2011**, *3* (7), a004952.

RESEARCH

Open Access



Size matters: the impact of nucleus size on results from spatial transcriptomics

Elyas Mohammadi^{1†}, Katarzyna Chojnowska^{1†}, Michał Bienkowski², Anna Kostecka¹, Magdalena Koczkowska¹, Michał A. Zmijewski³, Marcin Jąkałski¹, Martin Ingelsson^{4,5,6}, Natalia Filipowicz¹, Paweł Olszewski¹, Hanna Davies⁷, Justyna M. Wierzbicka³, Bradley T. Hyman^{8,9,10}, Jan P. Dumanski^{1,7}, Arkadiusz Piotrowski^{1†} and Jakub Mieczkowski^{1†*}

Abstract

Background Visium Spatial Gene Expression (ST) is a method combining histological spatial information with transcriptomics profiles directly from tissue sections. The use of spatial information has made it possible to discover new modes of gene expression regulations. However, in the ST experiment, the nucleus size of cells may exceed the thickness of a tissue slice. This may, in turn, negatively affect comprehensive capturing the transcriptomics profile in a single slice, especially for tissues having large differences in the size of nuclei.

Methods Here, we defined the effect of Consecutive Slices Data Integration (CSDI) on unveiling accurate spot clustering and deconvolution of spatial transcriptomic spots in human postmortem brains. By considering the histological information as reference, we assessed the improvement of unsupervised clustering and single nuclei RNA-seq and ST data integration before and after CSDI.

Results Apart from the escalated number of defined clusters representing neuronal layers, the pattern of clusters in consecutive sections was concordant only after CSDI. Besides, the assigned cell labels to spots matches the histological pattern of tissue sections after CSDI.

Conclusion CSDI can be applied to investigate consecutive sections studied with ST in the human cerebral cortex, avoiding misinterpretation of spot clustering and annotation, increasing accuracy of cell recognition as well as improvement in uncovering the layers of grey matter in the human brain.

Keywords Spatial transcriptomics, Cerebral cortex, Neuronal nuclei, Consecutive tissue sections, Data integration

[†]Elyas Mohammadi and Katarzyna Chojnowska contributed equally to this study

[†]Arkadiusz Piotrowski and Jakub Mieczkowski jointly supervised this study

*Correspondence:

Jakub.Mieczkowski

jakubm@gumed.edu.pl

¹ 3P-Medicine Laboratory, Medical University of Gdańsk, 80 210 Gdańsk, Poland

² Department of Pathomorphology, Medical University of Gdańsk, 80 210 Gdańsk, Poland

³ Department of Histology, Medical University of Gdańsk, 80 210 Gdańsk, Poland

⁴ Department of Public Health and Caring Sciences, Geriatrics, Rudbeck Laboratory, Uppsala University, 751 85 Uppsala, Sweden

⁵ Krembil Brain Institute, University Health Network, Toronto, ON M5G 2C4, Canada

⁶ Department of Medicine and Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, ON M5S 1A8, Canada

⁷ Department of Immunology, Genetics and Pathology and Science for Life Laboratory, Uppsala University, 751 85 Uppsala, Sweden

⁸ Department of Neurology, Massachusetts General Hospital, Boston, MA 02114, USA

⁹ Massachusetts Alzheimer's Disease Research Center, Charlestown, MA 02129, USA

¹⁰ Harvard Medical School, Boston, MA 02115, USA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

The spatial transcriptomics concept has been introduced as a combination of massively parallel sequencing and microscopic imaging [1]. This method is an attractive approach in studies of normal development and in clinical translational research. Visium Spatial Gene Expression (ST) is one of the technologies developed around this concept. ST is a next-generation molecular profiling method dedicated to unraveling the transcriptomic architecture of the tissue. The application of ST for mapping the transcriptome with the morphological context has been proven successful in many fields [2].

Although ST is a powerful new technique for capturing patterns of spatial distribution of gene expression, it also has a drawback of its design. A Visium Gene Expression slide consists of two or four tissue-capture areas (6.5 mm × 6.5 mm), divided into 4992 spots, each 55 µm in diameter. Every spot contains oligonucleotide probes with unique sequence barcodes that encode spatial information in gene expression data (Asp et al., 2020). Due to their size, spots may encompass the expression profiles of several cells. Consequently, this diminishes the accuracy of distinguishing neighboring cell types. This can be addressed by several methods [3, 4], including integration with other single-cell analyses [5]. The most popular methods for the integration rely on so called anchors, which represent similar gene expression patterns.

The importance of the anchor-based data integration in distinct single-cell modalities (i.e., spatial transcriptomics and single nucleus RNA sequencing data [snRNA-seq]) has been investigated previously [5]. However, the application of Consecutive Slices Data Integration (CSDI) in ST analysis using the anchor-based approach remains unexplored. We investigated the effects of CSDI on spot clustering and cell-type annotation using both snRNA-seq and ST technologies in human cerebral cortex samples. By applying the CSDI to ST, we aimed to evaluate whether a single slice of tissue would be sufficient for ST analysis or whether consecutive slices would be required. We found that without CSDI, the pattern of obtained spot clusters between consecutive slices is inconsistent, and the cell-type annotation does not match the microscopic characterisation of the slice. These issues were resolved by employing CSDI, and layer-structure of grey matter of the human brain was unveiled.

Methods

Data acquisition

We utilized the modified 10 × Genomics Visium Spatial Gene Expression method to analyze the profiles of consecutive sections from fresh-frozen brain tissues. Accordingly, we used the orbitofrontal (ON) and

temporal neocortex (TN) samples from two subjects. Tissue specimens were provided by Harvard University and Massachusetts Alzheimer's Disease Research Center and all experimental procedures were conducted in accordance with Independent Bioethics Committee for Scientific Research at Medical University of Gdansk (consent No. NKBBN/564-108/2022). The brain-tissue slices were placed onto a Visium Gene Expression slide (10x Genomics) and fixed according to the 10x Genomics protocol (doc. CG000239 Rev. C). Next, the slides were divided into two via a piece of silicone gasket. Subsequently, we stained the tissue by two methods—hematoxylin and eosin, and hematoxylin and Congo red—to detect eventual amyloid deposits. We imaged the slides at 20× magnification using brightfield settings (Olympus cellSens Dimension software). Afterward, the tissue was permeabilized, using conditions according to manufacturer protocol. The mRNA was released and bound to spatially barcoded capture probes on the slide. Next, cDNA was synthesized from captured mRNA, and sequencing libraries were prepared. Samples were loaded and pooled according to the protocol (doc. CG000239 Rev. C) and sequenced in the standard Illumina pair-end constructs, using Illumina's NextSeq 550 System.

Visium data processing

Four pairs (consecutive slices) of ST raw data (BCL files) from two postmortem brain samples were converted to fastq files using 10xGenomics software Space Ranger version 1.2.1 and its *spaceranger mkfastq* function. Subsequently, reads were aligned to the human genome-reference sequence (GRCH38) using the STAR method, and spatial feature counts were generated using the *spaceranger count* function. Because an inverted microscope was used for imaging, all images were flipped horizontally prior to being applied to the Space Ranger.

Data preprocessing and normalization

All outputs from *spaceranger count* were read as Seurat objects using *LoadIOX_Spatial* function of Seurat version 4.0.3. Prior to data normalization, the percentages of mitochondrial genes were calculated by the *PercentageFeatureSet* function. Then, the spots with a number of spatial features of more than 7000 and less than 200 were removed; spots that encompassed more than 15% of mitochondrial genes also were omitted from the downstream analyses. Standard normalization was performed using the *NormalizeData* function and the *LogNormalize* method using default parameters. Variable features for each object were determined using the *FindVariableFeatures* function and *VST* method. Next, the data were scaled and regressed out for the percentage of mitochondrial genes using the *ScaleData* function.

Dimensionality reduction and clustering

Dimensionality reduction was completed using the *RunPCA* function. Prior to clustering, nearest neighbors were determined by the *FindNeighbors* function with default parameters. After this, the *FindClusters* function determined the clusters by a shared nearest-neighbor (SNN) modularity optimization-based clustering algorithm (The resolution was arbitrarily set to 0.3).

Consecutive slices data integration

Dimensionality reduction for consecutive slices was completed jointly through diagonalized canonical correlation analysis (CCA). Using the *FindIntegrationAnchors* function, mutual nearest neighbors (MNNs) were found in this shared low-dimensional space and were termed anchors (for more details, see Stuart et al. [5]). The *IntegrateData* function was considered for CSDI using precomputed anchor sets. The integrated consecutive slices were saved as transcriptomics-corrected objects. The same workflow for dimensionality reduction and clustering was applied to integrated objects. Finally, the eight Seurat objects (four pairs of consecutive slices) before and after CSDI (16 in total) were saved as RDS files to be compared from different perspectives.

Label transferring from snRNA-seq to ST

The previously annotated snRNA-seq dataset was obtained from scREAD, a publicly available snRNA-seq database [6] (<https://bmbbls.bmi.osumc.edu/scread/>). The causes of death for the two donors were Alzheimer's disease (Subject A) and stroke (Subject B). Hence, snRNA-seq profiles with AD01104 and AD01102 scREAD data IDs for Alzheimer's and non-Alzheimer's disease were retrieved [7]. The raw-sequencing data and the digital-expression matrices obtained using the 10xGenomics software Cell Ranger are available in the NCBI's Gene Expression Omnibus (GSE129308) and are accessible through the GEO Series accession number GSM3704357-GSM3704375 (Otero-Garcia et al., 2020).

Data normalization and dimensionality reduction with the same parameters as the ST data were conducted for the two snRNA-seq datasets. By considering snRNA-seq as our reference and ST data as query datasets, anchors were found, and precomputed cell labels were transferred using the *FindTransferAnchors* (identifying shared cell/spot states present across different datasets) and *TransferData* functions, respectively. Label transferring and clustering were completed twice for each of the ST objects—once before and once after CSDI—to investigate the effect of CSDI on label transferring and clustering.

Results

We conducted spatial gene expression analysis in human postmortem, fresh frozen tissue sections. Two anatomical regions, the Orbitofrontal Neocortex (ON) and the Temporal Neocortex (TN) from two adult male donors were investigated (Fig. 1). From each region of both

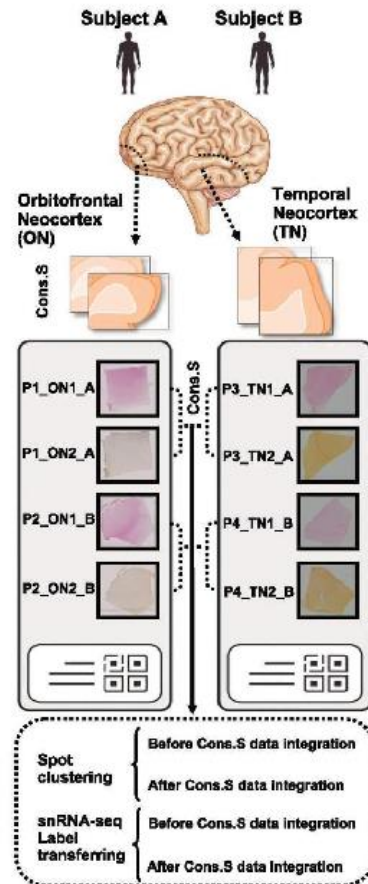


Fig. 1 Schematic view of experimental workflow and in silico analytical pipeline. Four pairs (P1–P4) of consecutive slices of human postmortem brains were obtained from two distinct anatomical locations (ON and TN). Two types of computational analyses were performed (spot clustering and snRNA-seq label transferring) and further divided into additional subtypes (before and after CSDI for each category). The names of slices are as follows: P, pairs of consecutive-sections; ON (1–2), Orbitofrontal Neocortex (section number in pair of consecutive slices); TN (1–2); Temporal Neocortex (section number in pair of consecutive slices); A and B: two studied subjects; Cons. S: consecutive sections

subjects, one pair of consecutive slices (eight slices in total) were prepared. We cut the ON and TN tissue into 10–12 μm thick sections. Each sample was sequenced to a median depth of 187 million reads, corresponding to a mean of 3300 unique molecular identifiers (UMIs) and a mean of 2058 genes per spot.

Identifying distinct cell types and their annotation using single tissue section

Figure 2A shows the distinction between the grey matter (GM) and the white matter (WM). The border between GM and WM was established histologically based on cellular composition and arrangement (Fig. 2A, Z1, Z2,

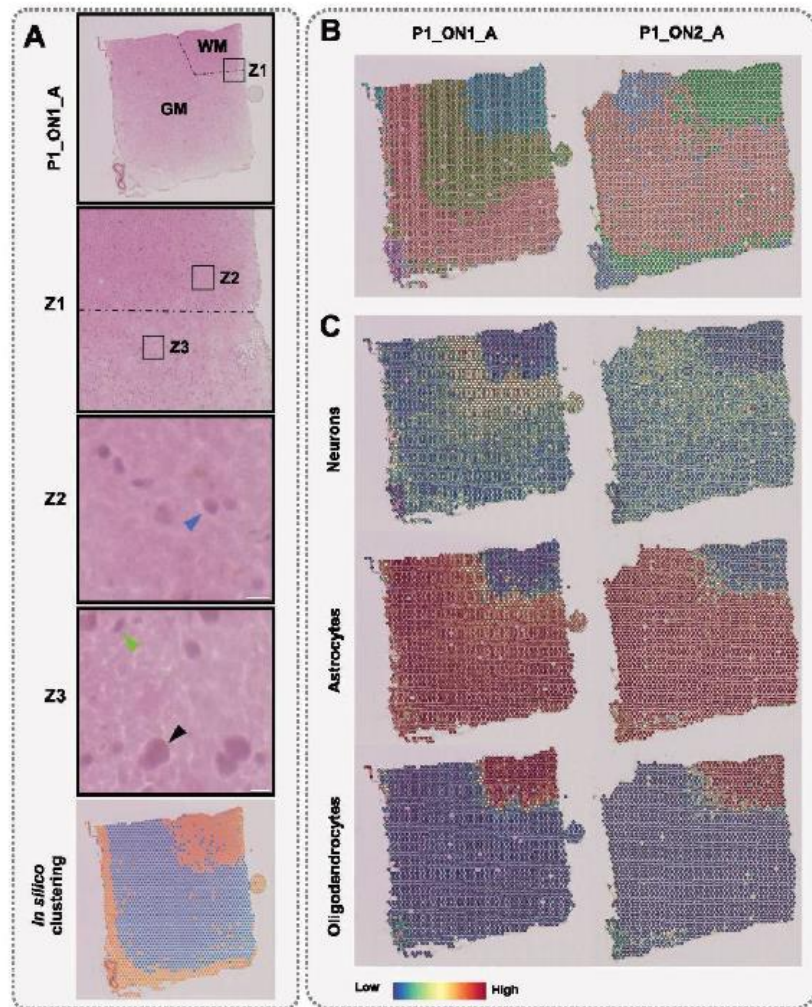


Fig. 2 Results from spatial transcriptomics analysis using single tissue sections. **A** Top, histological image of orbitofrontal neocortex (ON) with marked white matter (WM) and grey matter (GM); Z1, zoomed-in image of the border between WM and GM; Z2, Blue arrow points to an oligodendrocyte nucleus; Z3, Black and Green arrows represent nuclei of neurons and astrocytes, respectively. In Z2 and Z3, white scale bars represent 10 μm. Bottom, unsupervised classification of spots. **B** The ST spots clustering before CSDI. **C** Label transferring before CSDI. The name of the sample encodes number of the section (P1/P2), number of slice (ON1/ON2), and patient id (A/B)

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

and Z3). We used an unsupervised method to investigate whether categorizing the ST spots based on their transcriptomics profile could represent structural layers of the brain. Subsequently, we compared the obtained groups with histological images of tissue slices to assess the obtained clusterization and classification (Fig. 2A). Thus, we confirmed the general consistency of GM and WM patterns revealed by histologic and transcriptomic methods.

We performed spot clustering using the steps recommended by Satija et al. [7] in order to cluster the spots with similar expression profiles, and distinguish distinct cellular layers. The resulting clusters revealed the separation of subcortical WM and cortical GM. More detailed morphological layers of the brain were also unveiled through the more detailed clustering (Fig. 2B). Considering the expected similarity of architecture between two consecutive slices of the cerebral cortex, we should observe the very similar pattern of clusters. However, this consistency was vague, and the layered structure of GM in P1_ON2_A could not be observed (Fig. 2B). We observed that although, the use of a single section of brain tissue with the ST method can be informative, it may also have critical limitations in spot clustering. To overcome this, we decided to use external gene expression data set and anchor-based integration method.

To better understand the identified brain layers, we integrated the measured expression profiles with a previously described snRNA-seq dataset [8]. As single nucleus profiles contain greater number of genes than in our ST profiles, the integration of these two data sets allowed us to perform the spot annotation more precisely. Using predefined cell-type annotations in snRNA-seq—including oligodendrocytes, astrocytes, and neurons—the ST spots were labeled (see “Methods” for details). The pattern of transferred labels is shown in P1_ON1_A and P1_ON2_A as an example (Fig. 2C). The locations of oligodendrocytes and astrocytes were primarily identified in WM and GM, respectively, in line with brain structure (Fig. 2A). However, we could not confidently annotate neurons in GM, which is incompatible with histology (Fig. 2A). In summary, a single slice of brain tissue using the ST method is informative but has limitations in distinguishing cell types using label transferring as well as in spot clustering.

The effects of CSDI on identifying distinct cell types and their annotations

Stuart et al. [5] developed the CSDI to correct the transcriptomic profiles of consecutive slices using anchors representing spots with similar gene expression profile from two consecutive slices. This is used to pair spots from the two slices. At the same time, the transcriptomic

differences between pairs of spots in anchors are used to correct datasets from both consecutive sections.

We decided to apply the CSDI method due to the heterogeneity of the brain in terms of size of nuclei among different cell types (Fig. 3). On average, the size of a nucleus from a neuron in GM (about 20 μm) is much larger than the thickness of tissue section (10–12 μm). Consequently, a single tissue section will encompass only part of nuclei for essentially all neurons present in the studied sample. This restriction will also apply to other smaller nuclei, although to a lesser extent. Thus, for all cells present in a studied brain tissue, it will cause partial loss of transcriptomic signals. Taking “P1_ON1_A” and “P1_ON2_A” as consecutive slices of ON as an example, we could identify all the morphological layers of the brain [9] only after CSDI (Fig. 4A and Additional file 1: Fig. S1). In conclusion, CSDI can resolve the issue of inconsistency of the pattern of clustering between consecutive slices. Moreover, by applying the same parameters (see “Methods” for details), we can identify more neuronal layers in GM [10] (Fig. 4A and Additional file 1: Fig. S1).

Cell bodies of neurons are mainly found in the GM (Fig. 2A). However, during our label transferring, the spots marked as neurons received weak probability values in the GM (Fig. 2C). This is an important concern, which led us to hypothesize that using information from a single section of tissue may lead to inaccurate interpretation of clusters and cell types. Our approach to transferring cell labels from snRNA-seq to ST before and after CSDI revealed different results, which provides support for the above hypothesis. These differences are much more pronounced in GM, where the spots recognized as neurons, or astrocytes are the dominating cell types (Fig. 4B). Accordingly, we compared the annotations with consideration for the size of nuclei and the structural

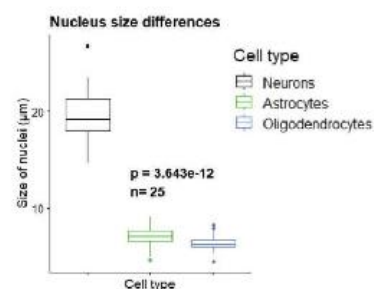


Fig. 3 The nucleus size heterogeneity among investigated cell types. The boxplot shows significant differences in size of nuclei in the human cerebral cortex (Kruskal–Wallis test). This is compatible with histological differences of nuclei size in Fig. 2A, Z2 and Z3 which the scale bars denote 10 μm

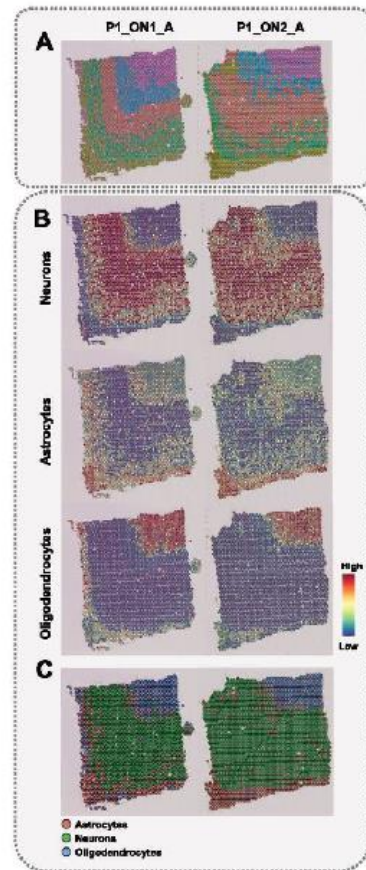


Fig. 4 Improvement of the spot clustering and annotation using CSDI method. **A** Clustering after CSDI unveiled the GM layers and resolved the inconsistency between the pattern of clusters in consecutive slices. **B** snRNA-seq Label transferring after CSDI. The annotation probability is shown as a scheme for three different cell types. **C** The classification of spots through label transferring is shown by considering the highest probability of annotation for each spot after CSDI

layers of the brain (WM and GM). Prior to CSDI, the annotation of neurons and astrocytes received low and high probabilities, respectively, in the area of the GM, while the nuclei of oligodendrocytes were mostly visible in the WM (Fig. 2C and Additional file 2: Fig. S2). Interestingly, after CSDI, the likelihood of the annotation of neurons was increased due to the gain of neuronal transcriptomic profiles (Fig. 4B), which is consistent with the histological imaging (Fig. 2A). It is noteworthy that

we did not observe any changes in the probability of annotation for oligodendrocytes in the WM before and after CSDI, which is also in agreement with histological structure of the cerebral cortex. Among the transferred labels from snRNA-seq to ST, neurons and astrocytes are mainly available in GM. Consequently, no signal fluctuation will occur before and after CSDI in WM. In both situations, the WM would preferentially be annotated with oligodendrocytes (Figs. 2C, 4B, and Additional file 2: Fig. S2).

The differences between annotations obtained for the spots before and after CSDI can be attributed to the fact that the size of neuronal nuclei is much bigger than astrocytic nuclei [11] (Fig. 3). Their size is actually larger than the thicknesses of the tissue sections used in the ST protocol. Accordingly, a single slice will capture incomplete transcriptomic neuronal context. CSDI provides a robust means of rectification of this misinterpretation. Hence, the corrected signals of all types of nuclei can be obtained. Consequently, the label transferring from snRNA-seq to ST is made consistent with the histological findings (Fig. 2A). Ultimately, one can study the spatial distribution of different cell types more precisely.

We evaluated the results from two independent spot-categorization methods used in this study: label transferring and spot clustering. Hence, we classified the spots using the transferred labels (Fig. 4C and Additional file 3: Fig. S3) and compared them with the spot clustering results represented in Fig. 4A. We observed that the green cluster in Fig. 4C represents the three distinguished neuronal layers in Fig. 4A (blue, red, and green clusters). We were not capable of labeling the neuronal layers in Fig. 4C as the utilized reference snRNAseq dataset did not distinguish neuronal subtypes. Similarly, the blue cluster in Fig. 4C represents oligodendrocytes in Fig. 4A (purple cluster). Through determining the locations of neurons and oligodendrocytes in both methods, we demonstrated that the results from both spot-categorization methods are consistent with the histological images (Fig. 2A).

We assessed the improvement of annotation for neurons before and after CSDI. A considerable rise in the number of defined neurons was observed after data integration (Fig. 5A and Additional file 4: Fig. S4). To investigate the accuracy of changes in spot labeling, we computed the differentially expressed genes (DEGs) in neurons versus oligodendrocytes and astrocytes before (if available) and after CSDI. In some of the tissue sections, no neurons could be identified before CSDI (Fig. 5 and Additional file 4: Fig. S4). Next, we applied the DEGs to Gene Ontology (GO) analysis (Fig. 5B and Additional file 4: Fig. S4). According to the enriched terms in GO analysis, neurons were identified accurately after data

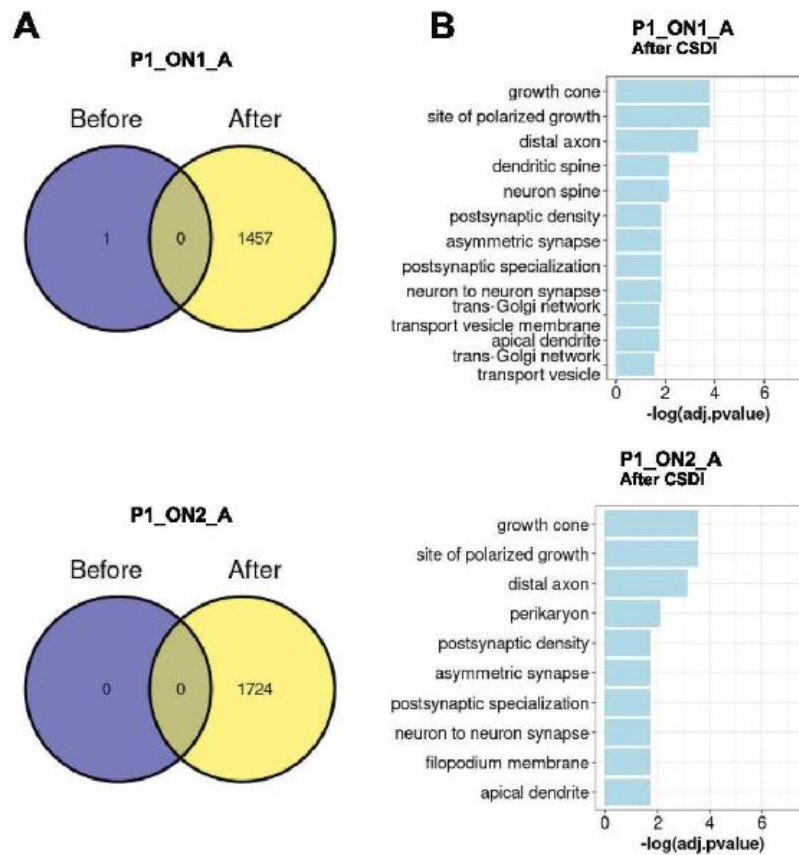


Fig. 5 Improvement of the results from DEGs in spots assigned to neurons, before and after CSDI. **A** Intersection of the spots assigned to neurons before (purple circle) and after (yellow circle) CSDI. **B** Gene ontology analysis is applied to DEGs in neurons versus oligodendrocytes and astrocytes

integration. We showed that by applying CSDI, mislabeled neuronal spots will be rectified. Thus, more accurate biologically meaningful results can be achieved.

Discussion

We studied the impact of the CSDI method on spatial gene expression analysis and evaluated the effect of CSDI on the improvement of clustering and label transferring [5]. The application of CSDI was motivated by the two issues we observed in the results of the basic spatial transcriptomic analysis. Firstly, in the GM, we observed inconsistencies between the patterns of clusters in consecutive slices (Fig. 4A). Secondly, we failed to recreate the expected layered structure of GM (Fig. 4A).

According to the study conducted by Maynard et al. [10], data correction in consecutive slices was performed using the data-refinement step of Space Ranger. Hence, the spatial topography of gene expression in the human dorsolateral prefrontal cortex was defined. The pattern of determining clusters was consistent in all pairs of consecutive slices, a phenomenon we observed only after applying the CSDI. Moreover, using CSDI, we distinguished cortical and subcortical WM layers. Thus, we showed that the expected consistency of the pattern of clustering between consecutive slices can be achieved with CSDI similar to Space Ranger. We investigated the results of the clustering and label transferring, with and without CSDI utilization. Simultaneously, we compared the consistency

of the results obtained with the topographic organization of the cerebral cortex. We observed the improvement of clustering and the label transferring after applying CSDI. The superior performance of using CSDI is likely related to the size of nuclei in different cell types as the determining parameter. The sizes of the nuclei of certain neurons are much larger than the nuclei of astrocytes and oligodendrocytes (e.g., neurons from the pyramidal layer (Figs. 2A and 3) of the cerebral cortex) [11]. In the human brain, the size of neuronal nuclei may exceed the tissue thickness recommended in the cryosectioning step of the ST protocol (10 μm) [12]. This may jeopardize capturing the whole transcriptomics profile using a single slice.

High resolution spatial methods [13] and/or experiments involving tissue sections or entire organ cross-sections from small animals are virtually free from the risk of losing the transcriptomics content of cells. Akeret et al. [14], studied single tissue sections of mice brains using 10 \times Visium spatial transcriptomics without any problem in spot labeling. The reason could be due to the fact that in mice, the average diameter of neuronal soma derived from the cortical pyramidal layer does not exceed 10 μm [15]. Hence, our approach is specifically applicable to tissues composed of cells with nuclei sizes exceeding the minimum thickness of the section required for the 10 \times Genomics Visium spatial transcriptomics experiment.

We used a combination of ST and snRNA-seq technologies to unveil the cerebral-cortex structure and related cell types. The ST preserves the spatial location of gene expression. However, its resolution at the level of the spot, as well as in terms of the number of captured genes, is nominally lower than the single-nuclei/single-cell transcriptomics [16]. The lower resolution results from the size of spots in ST expression slides (55 μm in diameter). Accordingly, each spot may encompass the transcriptomic profiles of multiple cells. The ST data integration with snRNA-seq/scRNA-seq is considered a deconvolution method to unravel the underlying cell types in each ST spot. In this context, using snRNA-seq has advantages over scRNA-seq. This is because the process of tissue cryopreservation ruptures the cell membranes; however, nuclear membranes remain intact during the freeze-thaw cycle [17]. Furthermore, it has been shown that the RNA-seq of single nuclei is highly representative of transcriptional profiles from the entire cells. This fact is specifically relevant to postmortem brain samples after long-term storage at -80°C [17]. Hence, we utilized the prelabeled snRNA-seq to deconvolute the ST spots.

To confirm the deconvolution of ST spots and defined cell types, we compared our annotation with neuropathological findings. Astrocytes play a vital role in delivering energy to neurons via the astrocyte-neuron lactate shuttle [18]. Hence, astrocytic nuclei are spatially

located beside perikarya (Fig. 2A), mostly placed in the GM [19]. In Fig. 4B, the GM is annotated for both neurons and astrocytes, corresponding to the previous findings [18]. According to the shape of oligodendrocytic nuclei—which are round with visible halos [20]—the annotation of WM for oligodendrocytes corresponds with the expected normal morphology of the brain cross-section [21] (Fig. 2A). These concepts are consistent with our histological (WM and GM) (Fig. 2A) and cell-type (astrocytes, neurons, and oligodendrocytes) (Fig. 4C) classifications.

An alternative solution to resolve the low resolution of the ST method is to decrease the size of barcoded spots in gene expression slide glasses. However, as we addressed in our study, ST results would be affected by the size of neuronal nuclei because the origin of the problem is not the sizes of spots but the thickness of the tissue slices. Accordingly, by decreasing the sizes of capture spots, deconvolution methods may no longer be required anymore; however, the need for CSDI remains.

In summary, the transcriptomic profiles of ST consecutive slices may need to be corrected prior to further analysis. Correcting the datasets simply for the depth of sequencing using normalization methods (e.g., log normalization) cannot remove all the unknown batch effects of consecutive slices. Data correction can be performed during the data-processing step by Space Ranger using the *spaceranger aggr* function or during the analysis steps using CSDI. In Space Ranger, the transcriptomic profile of consecutive slices will be aggregated, normalized to the same sequencing depth. Then, the feature-barcode matrices and the analysis of the combined data can be recomputed. In CSDI, the spots with similar transcriptomics profiles in two datasets will be anchored. Using the anchors, the transcriptomics profile of consecutive slices will be corrected, and one can proceed with the downstream analysis. Consequently, the results of clustering and annotation will be improved after data correction. Therefore, more trustable biological findings can be achieved. A general comparison between CSDI and *Space Ranger aggr* is shown in Additional file 4: Fig. S4. The pattern of clusters after applying CSDI is more consistent in consecutive slices than Space Ranger (Additional file 5: Fig. S5, A), while in label transferring both methods perform equally (Additional file 5: Fig. S5, B). However, more in depth analysis is required to show the outperformance of one over the other.

The study has potential limitations. First, while spatial transcriptomic technology allowed us to define the spatial location of cell types in human brain tissues, the resolution was limited to 1–10 cells per spot [22]. This means that spatial transcriptomic analysis should be taken with caution and possibly benefit from the computational

approaches like CSDI. In this study, to annotate the ST spots more precisely, we integrated snRNA-seq and ST profiles. Second, we could only examine the role of neuronal nucleus size from human brain on results from ST. Although, we acknowledged that this is not a concern in mouse brains [14, 15], we do not have enough relevant information about other species. In addition, we exclusively used the 10 μm thickness for the tissue sections of the human brain, as recommended by manufacturer. However, a wider range of tissue thicknesses could provide a more comprehensive understanding of how neuronal nucleus size may affect the ST analysis.

Conclusion

We observed that the thickness of tissue sections may be an important factor of the spatial transcriptomic analysis. In particular, the recommended in the cryosectioning step of the ST protocol (10 μm) [12] may not be sufficient to capture the entire transcriptomics profile of human brain tissue due to the large size of neuronal nuclei (about 20 μm). Importantly, this limitation can be overcome by using CSDI, which adjust the transcriptomic profiles prior to further analysis. The amendment leads to improved annotation results and more reliable biological findings.

Limitations of the study

Despite that ST is a powerful technology, the current study has the following limitations: (1) we couldn't attribute the gene changes to a particular cell type among multiple cell types captured in spots, therefore we need to use deconvolution methods to decipher the spots content; (2) the size of neuronal nuclei exceed the recommended thickness of tissue slices, thus the transcriptomics profile of spots encompassing neurons might be misinterpreted (3) the size of slide's capture area did not allow us to study any anatomic region completely.

Abbreviations

ST	Spatial gene expression
CSDI	Consecutive slices data integration
snRNA-seq	Single nucleus RNA sequencing
ON	Orbitofrontal neocortex
TN	Temporal neocortex
SNN	Shared nearest-neighbor
CCA	Canonical correlation analysis
MNNs	Mutual nearest neighbors
DEGs	Differentially expressed genes
GO	Gene Ontology

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-023-04129-z>.

Additional file 1: Figure S1. Improvement of the spot clustering through CSDI. The colors represent the results of unsupervised clustering performed either before (A) or after (B) application of CSDI. (A) The expected consistency of the pattern of spot clusters between consecutive slices is missed. (B) The consistency of clustering after consecutive slices data integration is improved. The patients IDs are marked above the pictures. The designations "before" and "after" refer to pictures made either before or after CSDI respectively.

Additional file 2: Figure S2. Improvement of the label transferring by CSDI. The annotation probability is shown as a scheme for three different cell types. Label transferring from snRNA-seq to ST consecutive slices is shown before and after data integration. Before integration, the probability of spot annotations for neurons is not compatible with tissue histology. After integration, the probability of the presence of neurons increased in the GM of the cerebral cortex.

Additional file 3: Figure S3. A comparison between the pattern of spot classification after CSDI through spot clustering and label transferring methods. The consistency between the results from these two methods, which is compatible with the histological information of tissue slices, supports the accuracy of the spot categorization.

Additional file 4: Figure S4. Improvement of the results from DEGs in spots assigned to neurons, before and after CSDI. Venn diagrams represent the intersection of spots assigned to neurons before (purple circle) and after (yellow circle) CSDI. In both situations, DEGs in neurons versus oligodendrocytes and astrocytes (if available) were applied to GO analysis and the results are shown in barplots.

Additional file 5: Figure S5. A comparison between CSDI and Space Ranger aggr in improvement of spot clustering and label transferring. Using CSDI, A) the pattern of clusters are more consistent versus Space Ranger aggr while in B) label transferring both methods perform equally.

Author contributions

Conceptualization, AP, JPD, JM, KC, EM, MB and PO; Methodology, KC, EM and MB; Software, EM, JM and MZ; Validation, KC, EM, MB, AK and MK; Formal analysis, EM, KC, AP and JM; Investigation, KC, EM, JM, AK and MK; Resources, JPD, AP, BTH, MI, MZ, and JW; Data Curation, EM; Writing—original draft, EM, KC, JM and AP; Writing—review and editing, EM, KC, JM, AP, JPD, MB, AK, BTH, MI, MK, MJ, NF, MZ, JW, PO and HD; Visualization, KC and EM; Supervision, JM, AP, JDM, MB and MJ; Project administration, JPD, AP, NF and HD; Funding acquisition, AP, JPD and BTH. All authors read and approved the final manuscript.

Funding

This work was supported by the Foundation for Polish Science under the International Research Agendas Program financed from the Smart Growth Operational Program 2014–2020 (Grant Agreement No. MAB/2018/6). This study was also partly supported by grants from the Swedish Cancer Society, the Swedish Research Council, Hjärtfonden, and Alzheimerfonden, to JPD, and P30AG062421/AG/NIA NIH HHS/United States, to B.T.H.

Availability of data and materials

The snRNA-seq data used in this study are publicly available at the scREAD database (<https://bmbis.bmi.osu.edu/scread/>) under AD01104 and AD01102 scREAD data IDs. Additionally, the snRNA-seq raw data and the digital expression matrices obtained using the 10x Genomics software Cell Ranger are available in the NCBI's Gene Expression Omnibus (GSE129308) and are accessible through the GEO series accession numbers GSM3704357–GSM3704375. The spatial transcriptomics data used in this study are privately available at the GEO data repository under the GSE184510 accession number. These data can be available from the authors upon a reasonable request.

Code availability

The scripts used in this study are developed by the R programming language version 4.1.0 and have been deposited in a public GitHub repository (https://github.com/jakubmie/ST_snRNA-seq).

Declarations**Ethics approval and consent to participate**

Tissue specimens were provided by Harvard University and Massachusetts Alzheimer's Disease Research Center and all experimental procedures were conducted in accordance with independent Bioethics Committee for Scientific Research at Medical University of Gdansk (consent No. NK88N/564-108/2022).

Consent for publication

Not applicable.

Competing interests

All authors declare no competing interests.

Received: 4 January 2023 Accepted: 12 April 2023

Published online: 21 April 2023

References

- Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. *Nature*. 2021;596:211–20.
- Asp M, Bergenstråhle J, Lundeberg J. Spatially Resolved transcriptomes—next generation tools for tissue exploration. *BioEssays*. 2020;42:1900221.
- Dong R, Yuan G-C. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biol*. 2021;22:145.
- Elosua-Bayes M, Nieto P, Mereu E, Gut I, Heyn H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res*. 2021;49:e50–e50.
- Stuart T, Butler A, Hoffman R, Hafemeister C, Papalexi F, Mauck WM III, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive integration of single-cell data. *Cell*. 2019;177:1888–1902. e1821.
- Jiang J, Wang C, Qi R, Fu H, Ma Q. scREAD: a single-cell RNA-Seq database for Alzheimer's disease. *iScience*. 2020;23: 101769.
- Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol*. 2015;33:495–502.
- Zhang Y, Fung ITH, Sankar P, Chen X, Robison LS, Ye L, D'Souza SS, Salinero AT, Kuentzel ML, Chittur SV. Depletion of NK cells improves cognitive function in the Alzheimer disease mouse model. *J Immunol*. 2020;205:502–10.
- Zhao F, Stone MR, Ren X, Guenther J, Smythe KS, Pulliam T, Williams SR, Uyingco CR, Taylor SEB, Nghiem P et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol*. 2021;39:1375.
- Maynard RR, Collado-Torres L, Weber LM, Uyingco C, Barry BK, Williams SR, Cattalini JL, Tran MN, Besich Z, Tippiani M. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci*. 2021;24:425–36.
- von Economo C, Tripathi LC. Cellular structure of the human cerebral cortex. Basel: Karger; 2009.
- 10xGenomics: Visium Spatial Gene Expression Optimized Tissues. <https://www.support.10xgenomics.com/spatial-gene-expression/tissue-optimization/doc/specifications-visium-spatial-gene-expression-optimized-tissues>. 2020.
- Rodrigues SG, Sickels RR, Goeva A, Martin CA, Murray F, Vanderburg CR, Welch J, Chen LM, Chen F, Macosko EZ. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science*. 2019;363:1463–7.
- Akeret K, Hugelschöfer M, Schaefer DJ, Buzzi RM. Spatial transcriptome data from coronal mouse brain sections after striatal injection of heme and heme-hemopexin. *Data Brief*. 2022;41: 107866.
- Kim EJ, Juavinett AL, Kyubwa EM, Jacobs MW, Callaway EM. Three types of cortical layer 5 neurons that differ in brain-wide connectivity and function. *Neuron*. 2015;88:1253–67.
- Navarro JF, Crozeau DL, Jurek A, Andrusivova Z, Yang B, Wang Y, Ogedegbe B, Riaz T, Szein M, Desler C. Spatial transcriptomics reveals genes associated with dysregulated mitochondrial functions and stress signaling in Alzheimer disease. *iScience*. 2020;23: 101556.
- Grindberg RV, Yee-Greenbaum JL, McConnell M, Nowotny M, O'Shaughnessy AL, Lambert GM, Araúzo-Brevo MJ, Lee J, Fishman M, Robbins GE. RNA-sequencing from single nuclei. *Proc Natl Acad Sci*. 2013;110:19802–7.
- Siracusa R, Fusco R, Cuzzocrea S. Astrocytes: role and functions in brain pathologies. *Front Pharmacol*. 2019;10:1114.
- Duchatel RJ, Shannon Wolkert C, Tooney PA. White matter neuron biology and neuropathology in schizophrenia. *Npj Schizophrenia*. 2019;5:10.
- El Sharouny S, Shaaban M, Elsayed R, Tahaf A, Abd ElWahed M. N-acetylcysteine protects against cuprizone-induced demyelination: histological and immunohistochemical study. *Folia Morphologica*. 2021;81:280–93.
- Hofmann K, Rodriguez-Rodriguez R, Gaebler A, Casals N, Scheller A, Kuerschner I. Astrocytes and oligodendrocytes in grey and white matter regions of the brain metabolize fatty acids. *Sci Rep*. 2017;7:10779.
- Ji AL, Rubin AJ, Tirone K, Jiang S, Reynolds DL, Meyers RM, Guo MG, George BM, Molitorik A, Bergenstråhle J. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell*. 2020;182(497–514): e422.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com

Gdańsk, on 28th of August 2023

Prof. dr hab. Arkadiusz Piotrowski
(professional title, name and surname)

STATEMENT

As the co-author of the work entitled **Size matters: the impact of nucleus size on results from spatial transcriptomics** I declare hereby that my own substantive contribution to the preparation, implementation and analysis of this study and presentation of experimental work in the form of above mentioned publication is **10% and comprise supervision and conceptualization of the project, formal analysis, providing resources, manuscript writing and editing, project administration and acquisition of the funding.**

At the same time, I consent to the submission of the above-mentioned work by **Elyas Mohhamadi** as part of a doctoral dissertation in the form of a thematically coherent collection of articles published in scientific journals.

I hereby declare that an independent and separable part of the above work shows the individual contribution of **Elyas Mohhamadi** when developing the concept, performing the experimental part, developing and interpretation of the results of this work.


.....
(co-author signature)

Gdańsk, on 28th of August 2023

Prof. Jan Dumanski
(professional title, name and surname)

STATEMENT

As the co-author of the work entitled **Size matters: the impact of nucleus size on results from spatial transcriptomics** I declare hereby that my own substantive contribution to the preparation, implementation and analysis of this study and presentation of experimental work in the form of above mentioned publication is **8% and comprise conceptualization of the project, providing resources, manuscript writing and editing, project administration and funding acquisition.**

At the same time, I consent to the submission of the above-mentioned work by **Elyas Mohhamadi** as part of a doctoral dissertation in the form of a thematically coherent collection of articles published in scientific journals.

I hereby declare that an independent and separable part of the above work shows the individual contribution of **Elyas Mohhamadi** when developing the concept, performing the experimental part, developing and interpretation of the results of this work.


.....
(co-author signature)

Gdańsk, on 28th of August 2023

dr hab. Jakub Mieczkowski
(professional title, name and surname)

STATEMENT

As the co-author of the work entitled **Size matters: the impact of nucleus size on results from spatial transcriptomics** I declare hereby that my own substantive contribution to the preparation, implementation and analysis of this study and presentation of experimental work in the form of above mentioned publication is **10% and comprise supervision and conceptualization of the project, formal analysis, investigation, manuscript writing and editing.**

At the same time, I consent to the submission of the above-mentioned work by **Elyas Mohhamadi** as part of a doctoral dissertation in the form of a thematically coherent collection of articles published in scientific journals.

I hereby declare that an independent and separable part of the above work shows the individual contribution of **Elyas Mohhamadi** when developing the concept, performing the experimental part, developing and interpretation of the results of this work.



.....
(co-author signature)

Gdańsk, on 10th of August 2023

dr Katarzyna Chojnowska
(professional title, name and surname)

STATEMENT

As the co-author of the work entitled **Size matters: the impact of nucleus size on results from spatial transcriptomics** I declare hereby that my own substantive contribution to the preparation, implementation and analysis of this study and presentation of experimental work in the form of above mentioned publication is **22% and comprise conceptualization and methodology of the experiments, formal analysis, experimental part, validation, manuscript writing and visualization of the results.**

At the same time, I consent to the submission of the above-mentioned work by **Elyas Mohhamadi** as part of a doctoral dissertation in the form of a thematically coherent collection of articles published in scientific journals.

I hereby declare that an independent and separable part of the above work shows the individual contribution of **Elyas Mohhamadi** when developing the concept, performing the experimental part, developing and interpretation of the results of this work.


.....
(co-author signature)

4-1BBL-containing leukemic extracellular vesicles promote immunosuppressive effector regulatory T cells

Julian Swatler,¹ Laura Turos-Korgul,¹ Marta Brewinska-Olchowik,¹ Sara De Biasi,² Wioleta Dudka,^{1,3} Bac Viet Le,^{1,4} Agata Kominek,¹ Salvador Cyranowski,^{5,6} Paulina Pilanc,⁵ Elyas Mohammadi,⁷ Dominik Cysewski,⁸ Ewa Kozłowska,⁹ Wioleta Grabowska-Pyrzewicz,¹⁰ Urszula Wojda,¹⁰ Grzegorz Basak,¹¹ Jakub Mieczkowski,⁷ Tomasz Skorski,⁴ Andrea Cossarizza,^{2,12} and Katarzyna Piwocka¹

¹Laboratory of Cytometry, Nencki Institute of Experimental Biology, Warsaw, Poland; ²Department of Medical and Surgical Sciences for Children & Adults, University of Modena and Reggio Emilia, Modena, Italy; ³Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany; ⁴Fels Cancer Institute for Personalized Medicine, Lewis Katz School of Medicine at Temple University, Philadelphia, PA; ⁵Laboratory of Molecular Neurobiology, Nencki Institute of Experimental Biology, Warsaw, Poland; ⁶Postgraduate School of Molecular Medicine, Medical University of Warsaw, Warsaw, Poland; ⁷SP Medicine Laboratory, Medical University of Gdansk, Gdansk, Poland; ⁸Laboratory of Mass Spectrometry, Institute of Biochemistry and Biophysics, Warsaw, Poland; ⁹Department of Immunology, Faculty of Biology, University of Warsaw, Warsaw, Poland; ¹⁰Laboratory of Preclinical Testing of Higher Standard, Nencki Institute of Experimental Biology, Warsaw, Poland; ¹¹Department of Hematology, Transplantation and Internal Medicine, Medical University of Warsaw, Warsaw, Poland; and ¹²National Institute for Cardiovascular Research, Bologna, Italy

Key Points

- Rab27a-dependent secretion of leukemic EVs promotes leukemia engraftment and immunosuppressive potential of Treg cells in vivo.
- Leukemic EVs containing 4-1BBL protein promote eTregs displaying specific signature (CD39, CCR8, CD30, TNFR2, CCR4, TIGIT, and IL21R).

Chronic and acute myeloid leukemia evade immune system surveillance and induce immunosuppression by expanding proleukemic Foxp3⁺ regulatory T cells (Tregs). High levels of immunosuppressive Tregs predict inferior response to chemotherapy, leukemia relapse, and shorter survival. However, mechanisms that promote Tregs in myeloid leukemias remain largely unexplored. Here, we identify leukemic extracellular vesicles (EVs) as drivers of effector proleukemic Tregs. Using mouse model of leukemia-like disease, we found that Rab27a-dependent secretion of leukemic EVs promoted leukemia engraftment, which was associated with higher abundance of activated, immunosuppressive Tregs. Leukemic EVs attenuated mTOR-S6 and activated STAT5 signaling, as well as evoked significant transcriptomic changes in Tregs. We further identified specific effector signature of Tregs promoted by leukemic EVs. Leukemic EVs-driven Tregs were characterized by elevated expression of effector/tumor Treg markers CD39, CCR8, CD30, TNFR2, CCR4, TIGIT, and IL21R and included 2 distinct effector Treg (eTreg) subsets: CD39⁺CCR8^{hi}TNFR2^{hi} eTreg1 and CD39⁺TIGIT^{hi} eTreg2. Finally, we showed that costimulatory ligand 4-1BBL/CD137L, shuttled by leukemic EVs, promoted suppressive activity and effector phenotype of Tregs by regulating expression of receptors such as CD30 and TNFR2. Collectively, our work highlights the role of leukemic extracellular vesicles in stimulation of immunosuppressive Tregs and leukemia growth. We postulate that targeting of Rab27a-dependent secretion of leukemic EVs may be a viable therapeutic approach in myeloid neoplasms.

Introduction

Myeloid neoplasms, including chronic and acute myeloid leukemia (CML/AML), are characterized by evasion of antileukemic effector immune response and induction of immunosuppression. In CML and AML

Submitted 20 September 2021; accepted 15 January 2022; prepublished online on *Blood Advances* First Edition 7 February 2022; final version published online 17 March 2022. DOI 10.1182/bloodadvances.2021006195.

Transcriptomics data were deposited under Gene Expression Omnibus number GSE180883. Proteomics data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository, identifier PXD027240.

Requests for data sharing may be submitted to Katarzyna Piwocka (k.piwocka@nencki.edu.pl).

The full-text version of this article contains a data supplement.

© 2022 by The American Society of Hematology. Licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), permitting only noncommercial, nonderivative use with attribution. All other rights reserved.

patients at diagnosis, effector cells of the immune system (CD8⁺ T cells, NK cells) are exhausted and dysfunctional.¹⁻³ Simultaneously, suppressive immune subsets, including suppressive myeloid cells⁴ and Foxp3⁺ regulatory T cells (Tregs), dominate the microenvironment.⁴ Tregs are increased in peripheral blood^{5,6} and bone marrow (BM) of leukemic patients.^{7,8} Importantly, high levels of Tregs in blood and BM of AML patients predict inferior response to chemotherapy, disease relapse, and shorter survival.^{9,10} Depletion of Tregs in a mouse model of MLL-AF9 AML reduced leukemic burden and increased survival.¹¹ In CML, Tregs were shown to decrease in patients responding to tyrosine kinase inhibitors, and low Treg levels were associated with treatment-free remission.¹²

Although direct Treg targeting and subsequent elimination would seem like a viable therapeutic strategy, precise depletion of Tregs is difficult to achieve in patients, and it may lead to severe autoimmune adverse events.¹³ Downregulation of proleukemic Tregs and improved therapeutic outcome could be achieved by targeting factors that drive expansion and activity of Treg subsets in leukemias. Even though some candidates, including coinhibitory PD-1/PD-L1, Gal9/TIM-3 pathways, and IDO enzyme, were implicated in Tregs expansion,^{3,14-16} Treg-driving mechanism have been poorly investigated in myeloid neoplasms.

Extracellular vesicles (EVs) have recently emerged as important mediators of intercellular communication. EVs are particles released by all types of cells and present in body fluids. They are divided into small exosomes, medium microvesicles, and large apoptotic bodies.¹⁷ EVs have been largely implicated in both immune cell activation and immunosuppression.¹⁸ In solid tumors, EVs inhibit antitumor activity of T cells¹⁹ and are responsible for induction and expansion of Tregs.^{20,21} In leukemias, including myeloid neoplasms, leukemic EVs promote (in an autocrine manner) growth and drug resistance of leukemic cells,^{22,23} as well as modify stromal and vascular components of the BM niche.^{24,25} Until now, immunomodulatory properties toward T cells have only been described for AML-derived EVs, which inhibit effector function of CD8⁺ cytotoxic T lymphocytes.²⁷

Using mouse *ex vivo* models, we previously observed involvement of CML-derived leukemic EVs in regulation of suppressive activity of murine Tregs and Foxp3 expression.²⁸ Here, we use *ex vivo* models with human lymphocytes and EVs to show that leukemic EVs, containing 4-1BBL/CD137L/TNFSF9 protein, promote induction and immunosuppressive polarization of human Tregs by modulating mTOR and STAT5 signaling. We identify specific subsets of effector Tregs (eTreg), characterized by expression of novel tumor Treg markers, such as CD30, CCR8, TNFR2, CD39, and TIGIT. Finally, *in vivo* in a mouse model of CML-like disease, Rab27a-dependent leukemic EVs promote leukemic engraftment, which was associated with higher abundance of activated, immunosuppressive Tregs. Our results indicate that inhibition of leukemic EVs secretion, to attenuate Tregs, may contribute to improved therapeutic outcome in myeloid leukemias.

Methods

Cell lines

Human cell lines CML-K562 (ATCC, CCL-243) and AML-MOLM-14 (DSMZ, ACC777) were cultured in supplemented Iscove's modified Dulbecco's medium or RPMI 1640 media, respectively.

Murine 32D BCR-ABL1⁻GFP⁺ cells were obtained from 32D BCR-ABL1⁺ cells and cultured as previously described.²⁸ To obtain stable knock-out cell lines without expression of Rab27a or 4-1BBL, clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) technology was used, using "all-in-1" plasmids encoding guide RNA, Cas9, and red fluorescent protein (RFP) (Merck-Sigma-Aldrich).

Primary human samples from patients

Plasma (source of primary EVs) was obtained from whole blood of 10 leukemic (7 CML and 3 AML; before starting treatment and at diagnosis) patients and 10 healthy donors. Blood was processed as described in supplemental Methods. Material from patients was collected with their informed written consent, under the approval of the Bioethics Committee of the Medical University of Warsaw (KB/107/2018) and Ethics and Bioethics Committee of the Cardinal Stanislaw Wyszyński University in Warsaw (KEIB-19/2017), and in accordance with the Declaration of Helsinki and Polish regulations.

Extracellular vesicles isolation and characterization

EVs released by leukemic cells were isolated from cell culture medium, conditioned for 24 hours by either K562 (CML EVs) or MOLM-14 (AML EVs) cells (seeded in media with EVs-depleted fetal bovine serum). EVs were isolated by differential ultracentrifugation as previously described.¹⁷ Pelleted EVs were resuspended in either nonsupplemented AIM V medium (for functional assay with T cells) or phosphate-buffered saline (for EVs characterization). EVs characterization and uptake analysis (supplemental Figure 1A-F) was performed as previously described¹⁷ (described in detail in supplemental Methods). For all experiments (unless specified in dose-dependent studies), T cells were treated with either 3 × 10⁹ CML EVs or 5 × 10⁹ AML EVs, equivalent of EVs released by 1 × 10⁸ cells. EVs from plasma were isolated using size-exclusion chromatography (qEV original 35 nm columns, Izon) and fractions 7 to 9 (supplemental Figure 1G-H) were pooled and concentrated using Amicon Ultra-2 10kDa (Merck). Isolation and characterization of EVs was performed according to guidelines of the International Society for Extracellular Vesicles²⁹ and the EV-TRACK consortium³⁰ (EV-TRACK database entry EV210187).

Primary T cell (Treg) isolation and culture

Human lymphocytes were obtained from buffy coats of healthy donors (different donor each experiment) from the Regional Center for Blood Donation and Blood Care in Warsaw, Poland (in accordance with the Declaration of Helsinki and Polish regulations). Peripheral blood mononuclear cells (PBMCs) were isolated by density gradient centrifugation (Lymphoprep, STEMCELL). T cells, including Tregs, were sorted using BD FACS Aria II (gating strategy: supplemental Figure 2A). Sorted T cells (CD4⁺CD25^{hi}CD127^{lo} Treg, CD4⁺CD25⁻ conventional T cells/Tconv) were cultured in AIM V medium (Gibco), stimulated with antibodies anti-CD3 (coated wells, 5 μg/mL; Biolegend) and anti-CD28 (soluble, 1 μg/mL; Biolegend), and supplemented with IL-2 (100 IU/mL for Tconv/Foxp3 induction, 50 IU/mL for Tregs; Peprotech). Cultures were maintained for 6 days to analyze Foxp3 induction in Tconv, 5 days to analyze Tregs, and 18 hours to analyze phosphorylation of signaling molecules in T cells.

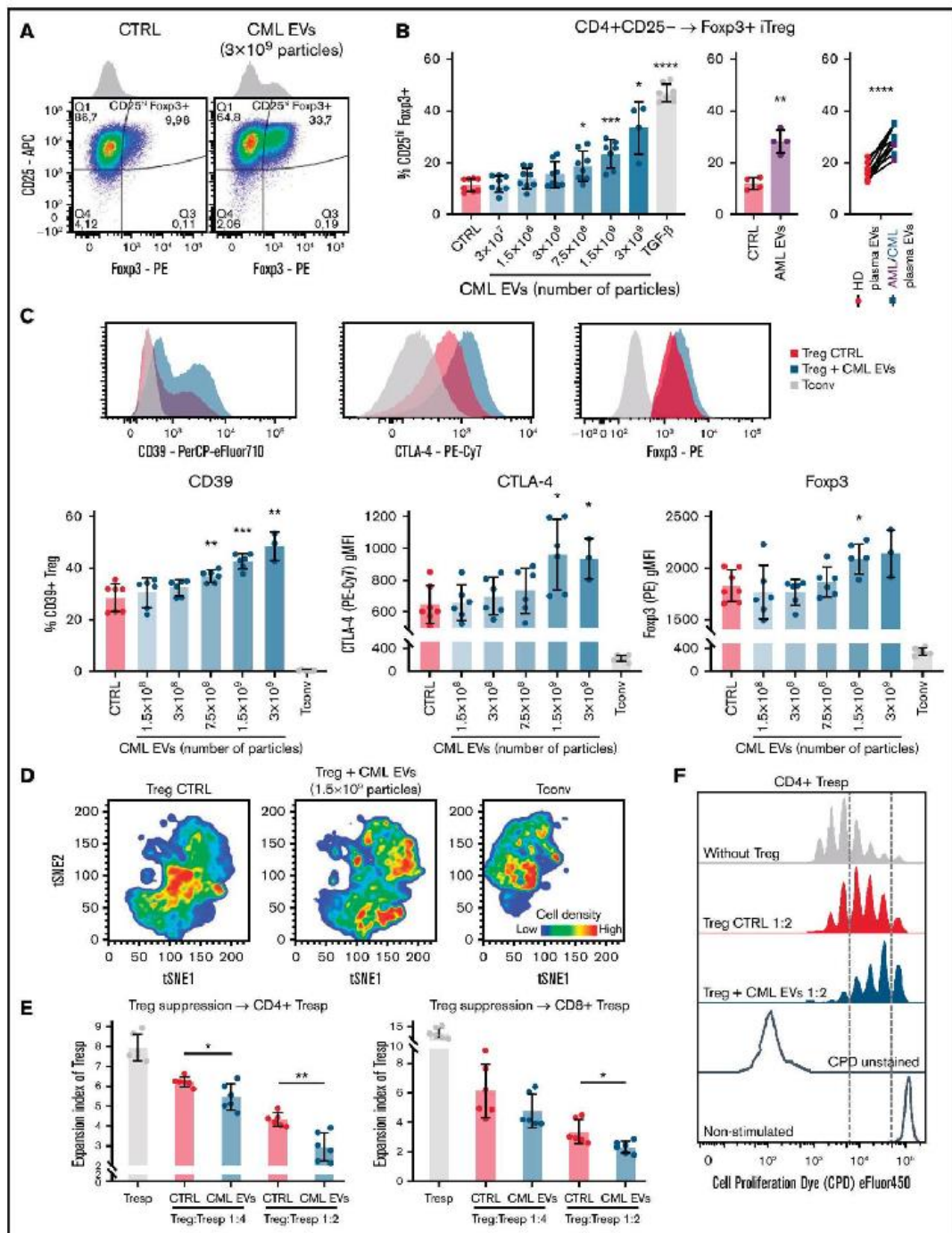


Figure 1. Leukemic EVs promote induction, suppressive phenotype, and activity of human Tregs. (A) Representative dot plots and histograms of iTregs induced from CD4⁺CD25⁻ Tconv in control conditions and after treatment with CML EVs (K562-derived). (B) Induction of Fcγ3⁺ cells (iTregs) from CD4⁺CD25⁻ Tconv cells by

In vivo mouse model of CML

To study CML in vivo in immunocompetent animals, murine 32D BCR-ABL1⁺GFP⁺ cells, either wild-type (wt) or Rab27a^{-/-}, were injected into genetically-matched male C3H mice, 8 to 10 weeks old (Figure 2A). All experimental procedures were performed according to the guidelines of Poland's National Ethics Committee for Animal Experimentation and approved by the First Local Ethics Committee for Animal Experimentation in Warsaw (635/2019, 1059/2020). In brief, 1×10^6 32D BCR-ABL1⁺GFP⁺ cells were injected intraperitoneally into nonirradiated animals. Control, nonleukemic mice were mock injected with NaCl. Following 2 months, blood, BM, and spleens were isolated and used for subsequent analyses. Development of leukemia-like disease was assessed by analysis of engraftment of GFP⁺ cells by flow cytometry. Animals that had GFP⁺ cells engrafted into BM, blood, and spleen (supplemental Figure 5A) were further analyzed.

High resolution spectral flow cytometry and data analysis

For 23-color phenotyping of Tregs treated with leukemic EVs, cells from ex vivo cultures were stained with surface antibody cocktail (including viability dye, supplemental Table 1) in Brilliant Stain Buffer Plus (BD) for 30 minutes. Cells were fixed and permeabilized for intracellular staining by eBioscience Foxp3/Transcription Factor Staining Buffer Set (Invitrogen). Intracellular proteins were stained for 30 minutes. Samples were acquired using CYTEK Aurora spectral flow cytometer and analyzed in FlowJo (BD). For manual analysis, cells were gated as shown in supplemental Figure 12A. For computational analyses, each sample was downsampled to obtain 7500 Treg. For tSNE and FlowSOM,³¹ all samples were concatenated and processed using specific plugins in FlowJo v10. For FlowSOM, cells were clustered into 6 populations (49 nodes), based on 15 parameters (Figure 5B). Generated clustering strategies were then used to cluster and quantify detected populations in individual samples and experimental conditions. Remaining flow cytometry assays and antibody specifics are described in supplemental Methods and Tables.

Statistical analysis

Data were plotted and statistics were performed using GraphPad Prism v9. Statistical tests used are indicated in figure legends. Unless indicated otherwise, statistics were performed in comparison with control (CTRL). Significant differences ($P < .05$) are marked

on graphs with asterisks (* $P < .05$, ** $P < .01$, *** $P < .001$, **** $P < .0001$).

Results

EVs released by myeloid leukemia cells induce Foxp3 and upregulate suppressive phenotype and activity of Tregs

In tumors, immunosuppressive milieu can either expand and drive effector polarization of already differentiated Treg or induce expression of Treg-specific transcription factor Foxp3 in nonregulatory, CD4⁺CD25⁻ conventional T cells and turn them into CD25^{hi}-Foxp3⁺ induced regulatory T cells (iTregs) (CD4⁺CD25⁻ → Foxp3⁺ iTreg).³² Therefore, we studied the impact of leukemic EVs on Foxp3 induction, phenotypic changes, and suppressive activity of human Tregs. We performed ex vivo cultures of purified (sorted) human CD4⁺CD25^{hi}CD127^{lo} Tregs or CD4⁺CD25⁻ Tconv, together with EVs released by CML-K562 cells (CML EVs) or AML-MOLM-14 cells (AML EVs) (characterized in supplemental Figure 1A-C). Interaction of EVs with human T cells/Tregs was confirmed by tracking fluorescent signal of carboxyfluorescein succinimidyl ester (CFSE)-labeled EVs in a culture with lymphocytes (supplemental Figure 1D-F).

CML EVs induced expression of Foxp3 in CD4⁺CD25⁻ cells, similarly to TGF- β , a widely recognized Foxp3 inducer (Figure 1A-B). Foxp3 induction was also observed after treatment with AML EVs and primary EVs from plasma of leukemic patients (as compared with healthy donor EVs) (Figure 1B). Induced expression of Foxp3 distinguished a separate population of iTreg (Figure 1A).

Treatment of sorted CD4⁺CD25^{hi}CD127^{lo} Tregs with CML EVs led to elevated expression of molecules responsible for suppressive activity: CD39, CTLA-4, and Foxp3 (Figure 1C). Unsupervised analyses of flow cytometric data based on these markers and activation molecule CD25 already indicated EVs-mediated polarization of Tregs into heterogeneous cell states (Figure 1D; supplemental Figure 3B) and expansion of highly suppressive subsets, as revealed by FlowSOM clustering (supplemental Figure 3C-D). Using "nonconditioned medium" control we confirmed that effects observed in our experiments were specific to CML EVs (supplemental Figure 4A-C). Consequently, a functional in vitro suppression assay confirmed that Tregs treated with CML EVs more potently

Figure 1 (continued) leukemic EVs, either released by CML-K562 cells (left panel), AML-MOLM-14 cells (middle panel, 5×10^9 particles), or primary patients' plasma EVs, compared with healthy donors' (HD) plasma EVs (right panel). For CML-K562, data are from 4 experiments, $n = 8$ (except $n = 4$ for 3×10^9 CML EVs); for AML-MOLM-14 from 4 experiments, $n = 4$. Mean \pm SD is presented, unpaired t test with Welch's correction, compared with CTRL. For plasma EVs, $n = 10$ CML/AML patients (8 AML, 7 CML) and 10 healthy donors. Pairing was done for samples that were used to treat the same batch of primary CD4⁺CD25⁻ Tconv. Two-tailed paired t test. For panels A and B, cells were gated as in supplemental Figure 2B. (C) Influence of leukemic EVs (K562-derived) on key proteins that drive suppressive activity of Tregs: CD39, CTLA-4, and transcription factor Foxp3. Non-Treg CD4⁺Tconv (CD25⁻) cells were used as negative controls. Representative histograms are shown (for CD39, 3×10^9 CML EVs; for CTLA-4 and Foxp3, 1.5×10^9 CML EVs). Data are from 3 experiments, $n = 6$ ($n = 3$ for 3×10^9 particles). (D) Unsupervised tSNE clustering of Tregs, Treg + CML EVs (K562-derived, 1.5×10^9 particles), and Tconv, based on CD39, CTLA-4, CD25, and Foxp3. In each group, 80 000 cells were clustered, 10 000 from each replicate (obtained by downsampling). For panels C and D, cells were gated as in supplemental Figure 3A. (E) In vitro suppressive activity of control and leukemic EV (K562-derived)-treated Tregs, pronounced as expansion index (EI) of responder cells (Tresp). Lower expansion index corresponds to higher suppressive activity. Data are from 3 experiments, $n = 6$. (F) Representative proliferation profiles of CD4⁺ responder cells in an in vitro suppression assay with control and CML EV (K562-derived)-treated Tregs. For panels C and E, mean \pm SD is presented, unpaired t test with Welch's correction, compared with CTRL. * $P < .05$, ** $P < .01$, *** $P < .001$, **** $P < .0001$. SD, standard deviation.

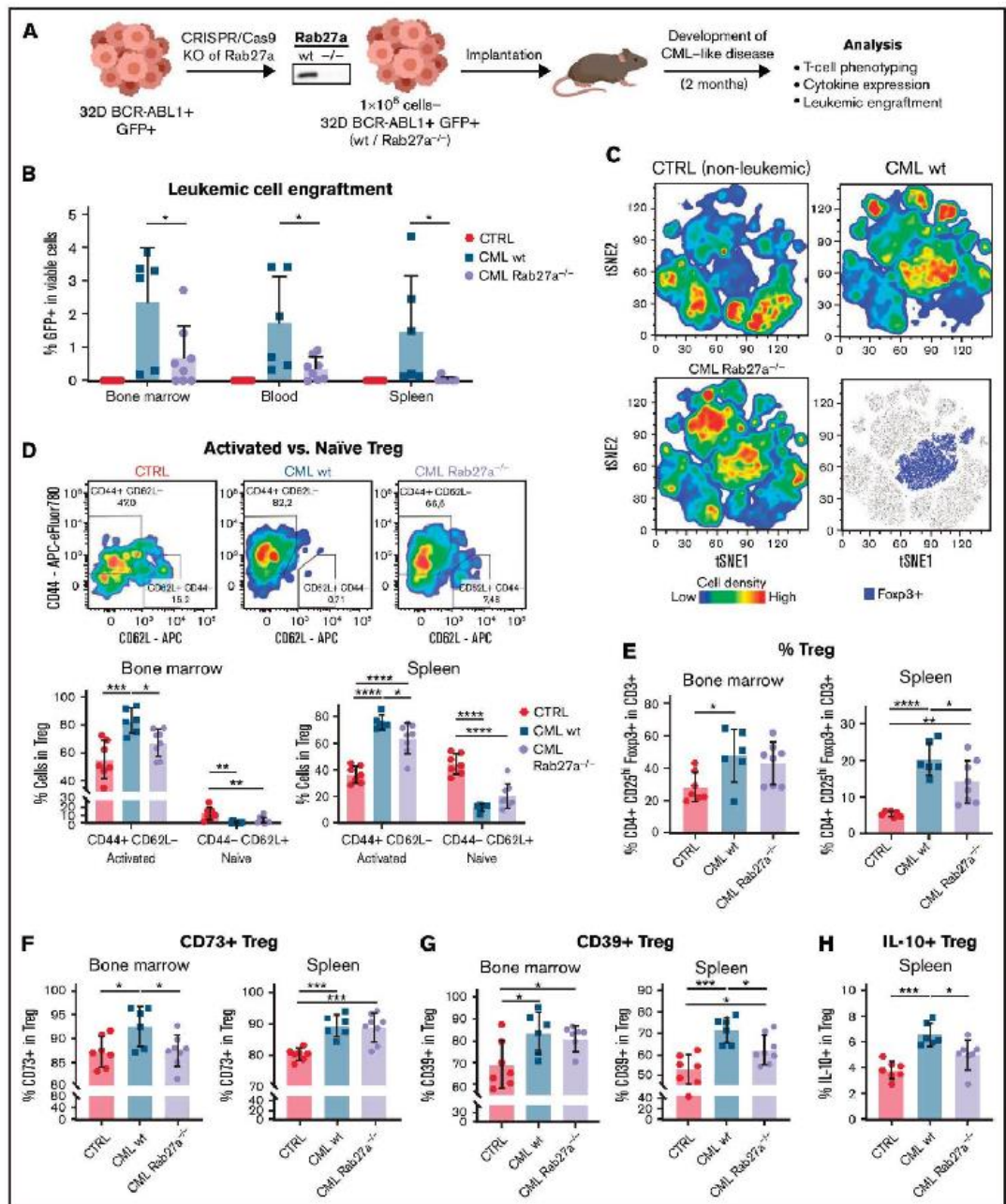


Figure 2. Rab27a-dependent leukemic EVs promote engraftment of CML cells and activation, amount, and suppressive phenotype of Tregs in vivo in mice. (A) Experimental pipeline of experiments using in vivo model of CML/leukemia-like disease. (B) Leukemic spreading and engraftment of leukemic (GFP⁺) cells in BM, blood, and spleen of mice with leukemia-like disease. (C) Unsupervised tSNE clustering of CD3⁺ T cells from spleens of control, healthy mice (CTRL), and animals with

inhibited proliferation of effector responder CD4⁺ and CD8⁺ T cells (Tresp), confirming superior suppressive activity (Figure 1E-F).

Overall, we demonstrate that leukemic EVs directly induce differentiation and promote suppressive phenotype and activity of human Tregs.

Rab27a-dependent secretion of leukemic EVs promotes Treg activity and leukemic engraftment in vivo

Rab27a is a significant regulator of EVs biogenesis, and Rab27a deficiency downregulates secretion of EVs, though it is not the only protein engaged in EVs secretion.³³ Due to its major function, Rab27a deficiency is sufficient to downregulate EVs secretion and has been used to study effect of continuous EVs secretion in in vivo models of cancer.¹⁹ To evaluate physiological relevance of leukemic EVs, stable Rab27a-deficient (Rab27a^{-/-}) 32D BCR-ABL1⁺GFP⁺ cells, which secrete 30% to 40% less EVs (supplemental Figure 5C-D), were used in an in vivo model of CML-like disease (Figure 2A).

Diminished release of EVs by Rab27a^{-/-} BCR-ABL1-expressing cells led to reduced engraftment of leukemic cells into blood, spleen, and BM (Figure 2B), whereas Rab27a deficiency did not affect their clonogenic potential, proliferation, and cell cycle (supplemental Figure 5E). These results suggest that leukemic EVs modulated disease development by affecting other cells in the microenvironment, including Tregs.

We analyzed Tregs in mice, focusing on spleen and BM: tissues where CML develops and which encompass full spectrum of cellular Treg interactions. Representative tSNE visualization of multiparameter flow cytometry data has already confirmed that in Rab27a^{-/-} leukemia-like disease, Foxp3⁺ Tregs cluster differently (predominantly as cells expressing lower CD25, CD39, CD44), as compared with wt leukemia and healthy animals (Figure 2C; supplemental Figure 6A). Detailed analysis of CD25^{hi}Foxp3⁺ Tregs has revealed that Rab27a-dependent secretion of EVs significantly promotes expansion of CD44⁺CD62L⁻ activated Tregs in both BM and spleen, contrary to naive CD44⁻CD62L⁺ Tregs (Figure 2D). Moreover, Rab27a deficiency resulted in lower Treg numbers in spleen, as well as diminished expression on Tregs of suppressive CD39, IL-10 in spleen and CD73 in BM (Figure 2E-H). Differences between spleen and BM in some phenotypic features of Treg in Rab27a^{-/-} leukemia may be an outcome of different tissue context, as BM consists primarily of myeloid cells, whereas T and B cells are most abundant in the spleen. This may also be relevant in terms of CD39/CD73 modulation, as these ectoenzymes may be expressed by other cells in BM/spleen niche.³⁴ Overall, significant changes in the amount of activated CD44⁺CD62L⁻ Tregs in both tissues show that leukemic EVs expand Tregs with immunosuppressive

potential in vivo, even though leukemic EVs did not entirely affect Treg phenotype or Treg numbers in the BM.

In our model, Rab27a^{-/-} CML did not significantly influence other immunosuppressive cells, such as Bregs or myeloid suppressive cells (supplemental Figure 8A-D), which additionally exhibited low abundance, compared with mouse models of solid tumors.^{35,36} This implicated direct modulation of Tregs by Rab27a-dependent leukemic EVs rather than indirectly via B cell-T-cell, or macrophage-T cell interactions.

Altogether, in vivo data strongly support our hypothesis that secretion of leukemic EVs and EVs-mediated upregulation of activated, immunosuppressive Tregs facilitate development of myeloid leukemias.

Furthermore, pharmacological targeting of EVs secretion by inhibition of Rab27a or nSMase2 has attenuated expansion of Tregs in cultures of human PBMCs with myeloid leukemia cell lines K562 and MOLM-14 (supplemental Figure 9). This suggests that effects similar to genetic targeting of Rab27a (described above) may also be achieved by pharmacological inhibition of EVs secretion.

Leukemic EVs drive Tregs by modulating mTOR and STAT5 signaling and remodeling of the transcriptome

Foxp3 and Tregs undergo complex molecular regulation.^{37,38} To identify molecular drivers involved in modulation of Tregs by leukemic EVs, we analyzed activation of signaling pathways crucial for regulation of Foxp3 and Treg biology (by phospho-specific flow cytometry)³⁸ in both human Tregs and CD4⁺CD25⁻ Tconv differentiating into iTregs, as well as performed transcriptomic analysis of human Tregs.

Leukemic (both CML and AML) EVs downregulated phosphorylation of mTOR and its downstream effector protein S6, parallel to upregulation of phosphorylated STAT5, in both Tregs and CD4⁺CD25⁻ Tconv upon Foxp3 induction by leukemic EVs (Figure 3A-D). Such changes have been described as favorable for Treg differentiation, stability, and suppressive function, as activated STAT5 binds the Foxp3 promoter to drive its transcription, whereas mTOR pathway inhibits Foxp3 expression.^{38,39} On the other hand, phosphorylation of p38, p65/RelA, and SMAD2/3 were not changed (supplemental Figure 10A-B). To validate whether signaling changes induced by leukemic EVs promote genetic and functional stability of Tregs, we analyzed demethylation of the Treg-specific demethylated region (TSDR) in the Foxp3 gene and secretion of both non-Treg-specific and immunoregulatory cytokines.⁴⁰ TSDR was demethylated in both CTRL and leukemic EVs-treated Tregs (Figure 3E). However, CML EVs blocked secretion of IL-6, IL-17A, and IFN- γ , preventing polarization into unstable, Th1/Th17-like subsets (Figure 3F), even though secretion of immunoregulatory IL-10 and TGF- β was unaffected

Figure 2 (continued) leukemia-like disease, induced by wt or Rab27a^{-/-} CML cells. The bottom right graph shows localization of Foxp3⁺ cells on the tSNE map. Data from 3 to 4 mice (per group) from single experiment were used as representative groups. In each group, 30 000 viable CD3⁺ T cells were clustered, 7500 to 10 000 from each animal (obtained by downsampling in FlowJo). (D) Distribution of activated (CD44⁺CD62L⁻) and naïve (CD44⁻CD62L⁺) Treg subsets in BM and spleen of mice bearing leukemia-like disease. Representative density plots showing expression of CD44 and CD62L by Tregs in the BM are shown. Treg amount (E) and expression of CD73 (F), CD39 (G), and IL-10 (H) on Tregs in BM and spleen of mice bearing leukemia-like disease. In each graph, data are presented as mean \pm SD. 1-way ANOVA with Tukey's posttest, **P* < .05, ***P* < .01, ****P* < .001, *****P* < .0001. *N* = 6 to 8 animals per group from 3 different experiments (different litters/groups of animals and leukemic cells' injections). Gating strategy for Treg phenotyping is shown in supplemental Figure 7A-B. SD, standard deviation.

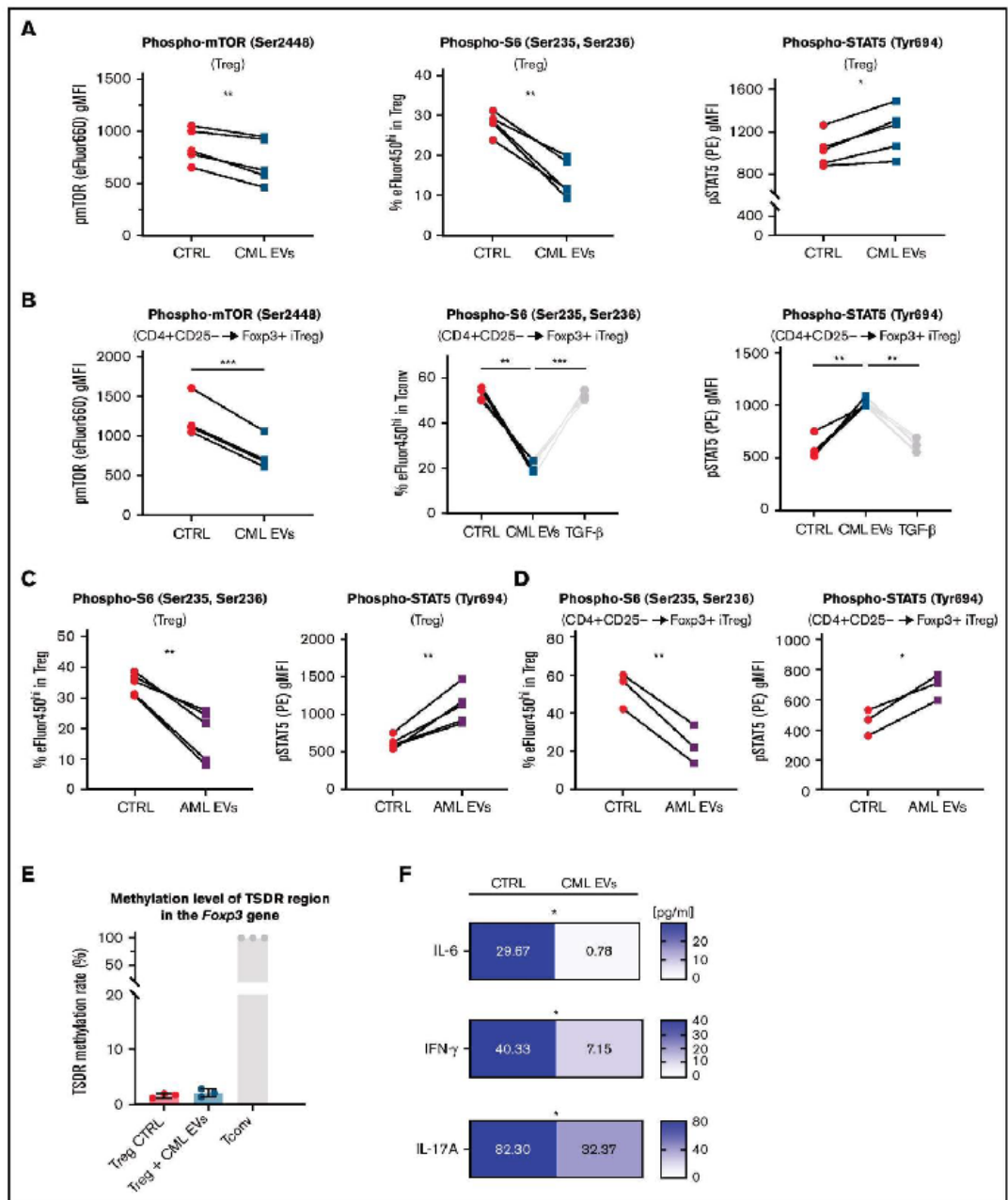


Figure 3. Leukemic EVs inhibit mTOR-S6 and upregulate STAT5 signaling in both human Tregs and CD4⁺CD25⁻ Tconv (differentiating into iTregs) and maintain stability of human Tregs. (A) Analysis of phosphorylation of mTOR, S6, and STAT5 in Tregs after treatment with CML EVs (K562-derived). Data are from 5 experiments, n = 5. Single data points, connected for each experiment, are presented. (B) Analysis of phosphorylation of mTOR, S6, and STAT5 in Tconv differentiating into iTregs. (C) Analysis of phosphorylation of mTOR, S6, and STAT5 in Treg after treatment with AML EVs. (D) Analysis of phosphorylation of mTOR, S6, and STAT5 in CD4⁺CD25⁻ Tconv differentiating into iTregs. (E) Analysis of TSDR methylation level in the Foxp3 gene. (F) Analysis of cytokine levels in CTRL and CML EVs.

(supplemental Figure 10C). These results clearly show that leukemic EVs not only upregulate suppressive features of Tregs but also maintain their stability. Both processes are likely modulated by downregulated mTOR-S6 and upregulated STAT5 signaling.

Analysis of Tregs by RNA sequencing revealed significant remodeling of the transcriptome and elevated expression of 356 genes due to treatment with CML EVs (Figure 4A-B), as well as influence on biological processes, such as RNA metabolism (supplemental Figure 11B-C). We analyzed genes described as characteristic for Tregs in cancer⁴¹⁻⁴⁵ and observed a visible trend of upregulated expression for *CCR4*, *TFR3*, *TNFRSF1B* (encoding TNFR2), *ENTPD1* (CD39), *TNFRSF8* (CD30), *IL1R1*, *HAVCR2* (TIM-3), and *TGFB1* (supplemental Figure 11D). However, in most cases, the difference was not statistically significant, therefore we additionally verified these observations on protein level (Figure 5D,F). Analysis of transcription factor-binding motifs (TFBMs) of differentially expressed genes identified several transcription factors potentially engaged in modulation of Tregs by leukemic EVs, such as EGR1, EGR3, ZBTB7A(LRF), E2F4, or TFDP1 (Figure 4C; supplemental Figure 11E). Overall, RNA sequencing further signified that leukemic EVs affect Treg, by global remodeling of gene expression, including upregulation of genes responsible for immunosuppressive function. Analysis of transcription factor-binding motifs pinpointed a set of transcription factors that modulate these changes in Tregs and may be relevant for immunosuppression in myeloid leukemias.

Tregs driven by leukemic EVs are constituted by heterogeneous effector subsets and characterized by upregulated CD39, CCR8, CD30, TNFR2, CCR4, TIGIT, and IL21R

To evaluate changes in human Tregs on single cell protein level, we developed a 23-color panel for spectral flow cytometry to analyze markers of effector/tumor Treg.⁴¹⁻⁴⁹ tSNE clustering revealed that leukemic EVs promoted expansion of heterogeneous Treg subsets (Figure 5A). Using FlowSOM we clustered Tregs into 6 populations (Figure 5B). As 2 of them (Pop0 and Pop3) exhibited very similar expression pattern of analyzed markers and established 1 joint cluster on tSNE, we decided to merge them and analyze together as 1 population (Pop0+Pop3) (Figure 5B, marked in pink). Two identified populations were upregulated (Pop2, Pop4) and 1 was downregulated (Pop0+Pop3), by CML and AML EVs (Figure 5C,E). Both upregulated populations exhibited high expression of functional effector markers (Figure 5B; supplemental Figure 12B), thus we named them effector Treg1 (eTreg1/Pop2) and effector Treg2 (eTreg2/Pop4). eTreg1 (Pop2) could be distinguished by high expression of CD30, CCR8, TNFR2, whereas eTreg2 (Pop4) by

high expression of CD39 and TIGIT (Figure 5B; supplemental Figure 12B). As newly described effector subsets exhibited a distinct phenotype, it suggests specialized functions of eTreg1 and eTreg2 in the context of leukemic microenvironment. The EVs-downregulated population (Pop0+Pop3) had lower expression of markers such as CCR4, CD25, Foxp3, ICOS (supplemental Figure 12B), thus it constitutes a less-effector and probably less suppressive population.

Classical, manual gating analysis of the entire Treg population revealed that CML and AML EVs upregulated expression of several effector tumor Treg markers, identifying a specific leukemic EVs-driven Treg signature that includes elevated expression of CD39, CCR8, CD30, Foxp3, TNFR2, CCR4, TIGIT, and IL21R markers (Figure 5D,F, blue frame). Although LAG-3 and CD73 were highly upregulated, their expression was limited to <1% of Tregs. Altogether, these data demonstrate that leukemic EVs mediate development of specific subsets of eTreg cells and promote EVs-dependent signature of Tregs (elevated expression of CD39, CCR8, CD30, Foxp3, TNFR2, CCR4, TIGIT, and IL21R).

Finally, we verified whether primary EVs isolated from plasma of CML/AML patients promote similar effector polarization of human Tregs and the same specific marker signature of Tregs. The treatment of Tregs with primary EVs from the plasma of leukemic patients (compared with healthy donor EVs) led to elevated expression of the signature molecules CD39, CCR8, TNFR2, CCR4, TIGIT, IL21R, and CD30 (Figure 5G). This provides evidence supporting the hypothesis that leukemic EVs, present in the circulation (plasma), can influence immune cells in distant tissues to facilitate immunosuppression outside the BM.

Collectively, leukemic EVs, both released in vitro by CML/AML cells and of primary origin (plasma EVs), upregulate specific signature of Tregs (characterized by high expression of CD39, CCR8, TNFR2, CCR4, TIGIT, IL21R, and CD30) that includes 2 effector immunosuppressive Treg subsets.

Leukemic EVs contain 4-1BBL protein, which contributes to Treg activity and effector phenotype

Finally, we aimed to identify specific protein content of EVs that influences human Tregs. Mass spectrometry, followed by functional annotation of detected proteins, identified groups of proteins connected to immune response, such as TNF signaling, including 4-1BBL/TNFSF9/CD137L (Figure 6A-B). Presence of 4-1BBL in CML and AML EVs was confirmed by western blotting (Figure 6C). TNF receptor superfamily was recently implicated in Treg function, though by far mainly in the small intestine, colon, and

Figure 3 (continued) (A) iTregs (CD4⁺CD25⁻ → Foxp3⁺ iTreg as in Figure 1A-B) after treatment with CML EVs (K562-derived). Data are from 4 experiments, n = 4. Single data points, connected for each experiment, are presented. (B) Analysis of phosphorylation of S6 and STAT5 in Tregs after treatment with AML EVs (MOLM-14-derived). Data are from 5 experiments, n = 5. Single data points, connected for each experiment, are presented. For panels A and B, cells were gated as in supplemental Figure 3A, using antibodies conjugated with fluorochromes as in supplemental Table 1. (C) Analysis of phosphorylation of S6 and STAT5 in Tconv differentiating into iTregs (CD4⁺CD25⁻ → Foxp3⁺ iTreg as in Figure 1A-B) after treatment with AML EVs (MOLM-14-derived). Data are from 3 experiments, n = 3. Single data points, connected for each experiment, are presented. For panels C and D, cells were gated as in supplemental Figure 2B, using antibodies conjugated with fluorochromes as in supplemental Table 1. (E) Level of methylation of TSDR region in the *Foxp3* gene in Tregs treated with leukemic EVs (K562-derived; Tconv served as positive control). Data are from 3 experiments, n = 3. Mean ± SD is presented. (F) Secretion of effector/proinflammatory cytokines (pg/ml, calculated per 1 × 10⁵ cells; final concentration shown on graphs) detected in cultured medium of control Tregs or Tregs cultured with leukemic EVs (K562-derived). Data are from 6 experiments, n = 6. For all experiments in the figure, statistics was 2-tailed paired *t* test. **P* < .05, ***P* < .01, ****P* < .001. SD, standard deviation.

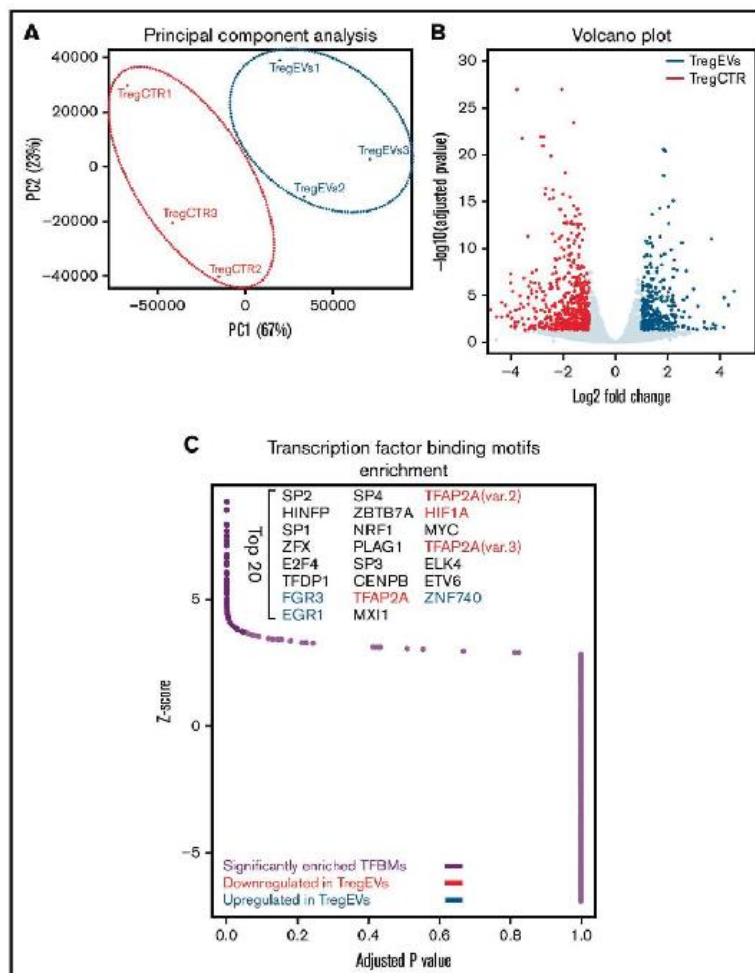


Figure 4. Leukemic extracellular vesicles (K562-derived) remodel human Treg transcriptome. (A) Principal component analysis (PCA) of transcriptomes of control (TregCTR) and CML EVs-treated Tregs (TregEVs). (B) Volcano plot comparing differences in gene expression between CTRL and CML EVs-treated Tregs (TregEVs). Treatment with CML EVs led to 356 upregulated (marked in blue) and 588 downregulated (marked in red) differentially expressed genes (DEGs). Genes were considered as differentially expressed when $-\log_{10}(\text{adjusted } P \text{ value}) < .05$ and $\log_2 \text{ fold change (LFC)} > 1$. (C) Analysis of transcription factor binding motifs (TFBMs) enriched in DEGs ($\text{LFC} > 1$), performed using PSCAN software. Significantly enriched TFBMs are indicated by dark purple dots, and top 20 hits are named. Transcription factors (TFs) that are differentially expressed at mRNA level are additionally indicated in either blue (upregulated by CML EVs) or red (downregulated by CML EVs). RNA sequencing was performed for $n = 3$ per group. For each replicate, Tregs obtained from a different donor were used. mRNA, messenger RNA.

during colitis.⁴⁶⁻⁴⁸ We hypothesized that it may also contribute to EVs-mediated eTreg polarization in myeloid leukemias. Using CRISPR/Cas9 mutagenesis, we generated 4-1BBL-deficient K562 CML cells, leading to the absence of 4-1BBL in CML cells and EVs (Figure 6D; supplemental Figure 13A). Tregs treated with 4-1BBL-deficient CML EVs no longer upregulated CD30, TNFR2, and LAG-3 (Figure 6E, but not the remaining receptors upregulated by leukemic EVs, such as CD39 or Foxp3; supplemental

Figure 13B-C), as well as exhibited weaker suppressive activity in a functional in vitro suppression assay (Figure 6F; supplemental Figure 13D). These data demonstrate the regulatory role of 4-1BBL in promotion of Treg activity and effector phenotype.

Overall, we postulate that TNF superfamily protein 4-1BBL contributes to effector immunosuppressive polarization of Tregs promoted by leukemic EVs.

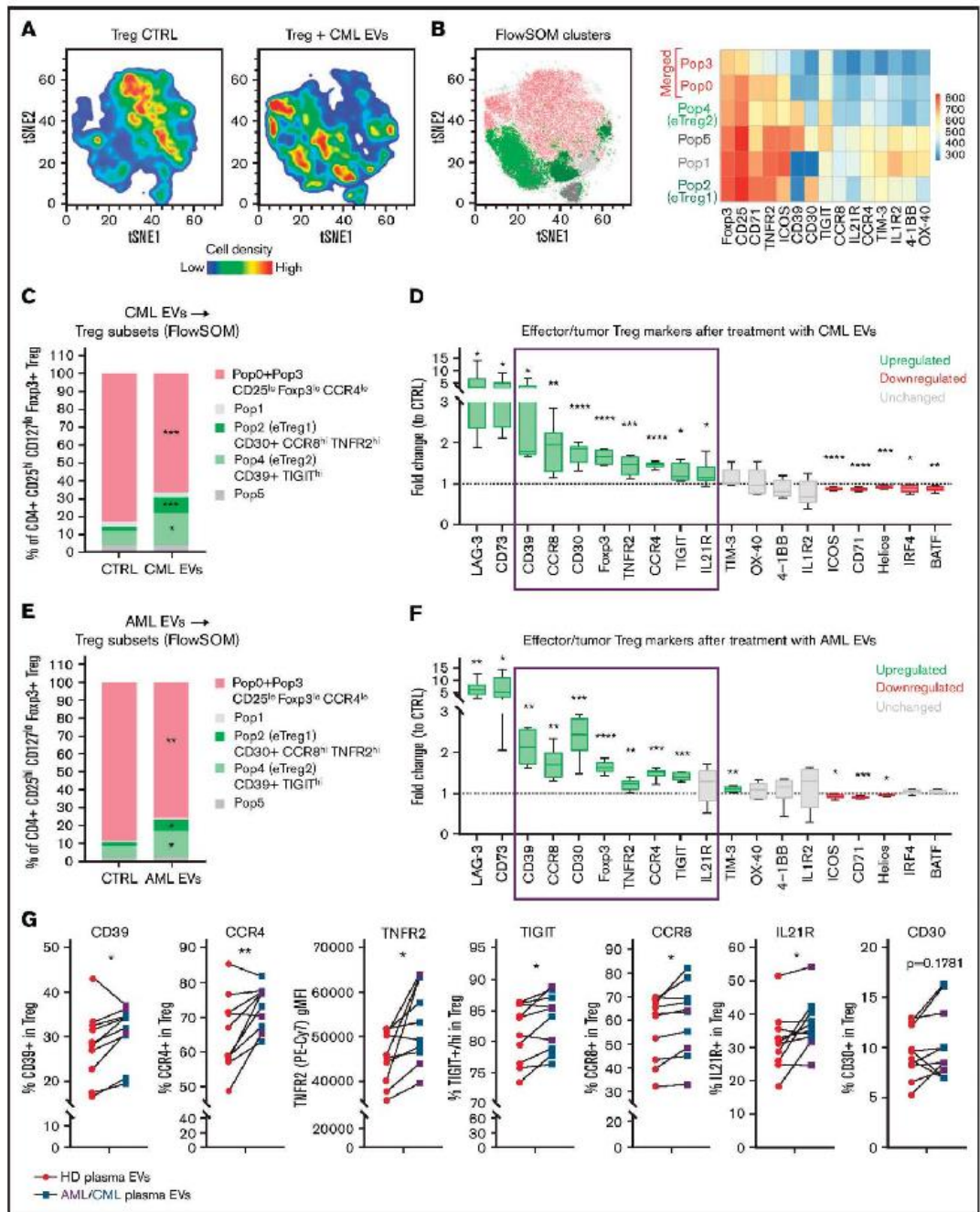


Figure 5. Human eTreg subsets and effector/tumor Treg signature are driven by leukemic EVs (A) Unsupervised tSNE clustering of CTRL and CML EV (K562-derived)-treated Tregs. (B) Unsupervised FlowSOM clustering of CTRL and CML EVs(K562-derived)-treated Tregs. Identified 5 to 6 FlowSOM clusters/populations were overlaid onto the tSNE map. Heatmap shows relative expression of selected markers by 6 FlowSOM populations. For panels A and B, data from 16 samples (8 per group).

Discussion

In the presented study, we report leukemic (CML and AML) extracellular vesicles as novel, significant drivers of human immunosuppressive Foxp3⁺ Tregs with an effector phenotype, including 2 distinct effector subsets (Figure 7). We show this by complementary approaches, using pure leukemic EVs released by human cell lines and primary EVs isolated from the plasma of leukemic patients, compared with EVs of healthy donors. Using the mouse model of CML-like disease, we showed that leukemic EVs-Tregs interaction facilitates the development of leukemia-like disease, which implicates importance for human leukemia.

EVs enable intercellular communication between distant tissues and cells. EVs may thus promote the growth of leukemic cells outside the BM by creating leukemia-permissive microenvironment in faraway tissues. It has already been established that Tregs and immunosuppression are critical elements that facilitate such leukemia-supporting conditions.^{3,9,11} We found that circulating, primary EVs from plasma of CML and AML patients induced Foxp3 and promoted effector signature of Tregs. This supports our hypothesis that leukemic EVs in circulation may drive proleukemic Tregs outside the BM and established a previously undescribed mechanism facilitating spreading of leukemic blasts.

Indeed, one of our critical findings, significant for myeloid neoplasms, is that EVs secretion facilitates leukemia development *in vivo* in mice, demonstrated in a model of leukemia-like disease induced by Rab27a^{-/-} CML cells with downregulated EVs secretion. Rab27a deficiency and diminished release of EVs *in vivo* partially reversed expansion of proleukemic, activated Tregs in the BM and spleen. In parallel, significantly lower engraftment of leukemic cells was observed. Importantly, we excluded the role of autocrine influence of EVs on growth and clonogenicity of leukemic cells and excluded the involvement of B cells or immune myeloid cells in the identified effects. These observations strongly suggest direct modulation of Tregs by CML EVs *in vivo* and the relevance of EVs-Tregs interaction for leukemia progression. Even if leukemic EVs have been shown to modulate other components of the BM niche,^{24,25} and we cannot entirely conclude that the effect depends solely on Tregs and immunosuppression, our data indicate significant involvement of Tregs. In a translational context, our *in vivo* discoveries provide rationale for therapeutic targeting of EVs/Rab27a in leukemias. Similar conclusions have been drawn by Poggio et al,¹⁹ based on mouse models of Rab27a^{-/-} colorectal and prostate cancers, where targeting EVs also targeted immune checkpoint molecule PD-L1. EVs secretion may be clinically targeted by pharmacological inhibitors of Rab27a⁴⁹ or other EV-regulatory hubs (nSMase2).⁵⁰ In our experiments, such treatment indeed attenuates expansion of

Tregs in cultures of human PBMCs and leukemic cells. Overall, our discoveries identified leukemic, Rab27a-dependent EVs and EVs-Tregs interaction as potential, previously unrecognized, therapeutic targets in myeloid neoplasms.

Recent studies of Tregs in cancer have identified new molecules specific for tumor Tregs, such as CCR8,⁴² CD30, IL21R,⁴⁴ and others,⁵¹ as well as specific eTreg subsets.⁴¹ Remarkably, our findings provide evidence that leukemic EVs contribute to the expansion of highly suppressive effector subsets of Tregs and promote specific effector signatures of Tregs, which has not been previously dissected in the cancer field. The eTreg1 subset we described had high expression of CCR8, CD30, and TNFR2, as well as transcription factor IRF4 (supplemental Figure 12B). It thus resembles an effector, IRF4-driven Treg population described in lung cancer.⁴¹ On the other hand, tissue and cellular context for these molecules in myeloid neoplasms remain to be elucidated, such as expression of CCR8 chemokine ligands in the leukemic BM. Presence of TNFR2, which is usually expressed by T cell receptor-activated Tregs,⁴⁴ implies interaction with antigen-presenting cells, such as tolerogenic dendritic cells.⁵² The ubiquitous expression of CD39 and TIGIT on the eTreg2 subset we identified suggests different function of this population in the leukemic microenvironment. TIGIT acts as a coinhibitory receptor, capable of inhibiting Th1- and Th17-polarized CD4⁺ T cells,⁵³ whereas CD39 is an ectoenzyme, converting adenosine triphosphate to adenosine (jointly with CD73), also to inhibit effector T cells.⁵⁴ Moreover, CD39 on Tregs can support hematopoietic stem cells in the BM,⁵⁵ which implies possible interaction of eTreg2 with leukemic stem cells. Thus, even though identified populations of Tregs have well-documented immunosuppressive phenotype and function, it would be important to dissect the precise relevance of eTreg1 and eTreg2 in the leukemic microenvironment, both functionally and spatially. The significance of leukemic EVs-Tregs interaction is further highlighted by our discoveries of molecular regulators. We identified transcription factors that may be responsible for driving eTreg or markers of eTregs in myeloid neoplasms. Some of them, such as EGR3,⁵⁶ E2F4,⁴⁹ ZBTB7A/LRF,⁵⁷ or TFDP1,⁴¹ have already been implicated in regulation of Tregs, as well as in tumors. We also pinpoint that although leukemic EVs modulate mTOR and STAT5 pathways, they do not engage TGF-β/SMAD signaling, which classically induces *de novo* Foxp3 expression,⁵⁸ implicating a new modality of Treg induction and modulation.

Finally, we detected a functional role of TNF superfamily member 4-1BBL/TNFSF9/CD137L in leukemic EVs and propose that its presence contributes to the amplified immunosuppressive potential of Tregs. According to ExoCarta database, 4-1BBL protein was previously not detected in EVs⁵⁹ but was identified on hematopoietic and progenitor cells in the BM.⁶⁰ Moreover, 4-1BB

Figure 5 (continued) 7500 events/sample, were concatenated and used to create tSNE map and FlowSOM clustering scheme. (C) Abundance of Treg subsets (identified by FlowSOM as in panel B) in CTRL and CML EV (K562-derived)-treated Tregs. (D) Expression (fold change to CTRL samples) of tumor Treg markers after treatment of Tregs with CML EVs (K562-derived). For TNFR2, Foxp3, IRF4, and BATF, gMFI was analyzed; for other markers, percent of positive cells. For panels C and D, data are from 4 experiments, n = 8. (E) Abundance of Treg subsets (identified by FlowSOM as in panel B) in CTRL and AML EV (MOLM-14-derived)-treated Tregs. (F) Expression (fold change to CTRL samples) of tumor Treg markers after treatment of Tregs with AML EVs (MOLM-14-derived). For TNFR2, Foxp3, IRF4, and BATF, gMFI was analyzed; for other markers, percent of positive cells. For panels E and F, data are from 4 experiments, n = 6. For panels C through F, statistics were unpaired *t* tests with Welch's correction. **P* < .05, ***P* < .01, ****P* < .001, *****P* < .0001. (G) Expression of selected tumor Treg markers after treatment of Tregs with primary patients' plasma EVs, compared with healthy donors' (HD) plasma EVs. For leukemic group, plasma from 3 AML and 7 CML patients was used. For TNFR2, gMFI was analyzed; for other markers, percent of positive cells. N = 10 CML/AML patients and 10 healthy donors. Pairing was done for samples that were used to treat the same batch of (primary) Tregs. Two-tailed paired *t* test. **P* < .05, ***P* < .01. Gating strategy for Treg phenotyping is shown in supplemental Figure 12A. gMFI, geometric mean fluorescence intensity.

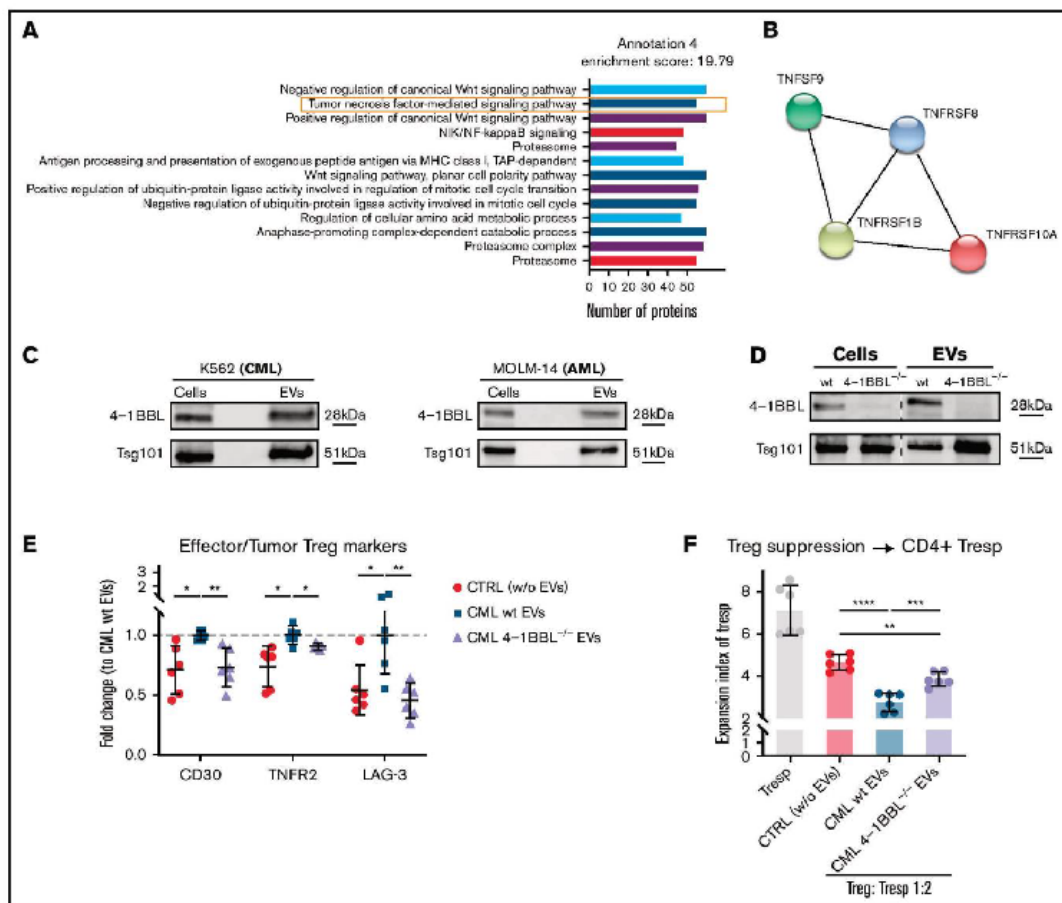


Figure 6. Leukemic EVs contain 4-1BBL/TNFSF9 protein, which contributes to effector phenotype and suppressive activity of human Tregs. (A) Protein groups identified (annotation 4 only, out of 18 statistically significant annotations; details in supplemental Table 4) in mass spectrometric profiling of CML EVs (K562-derived). Number of proteins in each group is indicated. Data were collected from 4 experiments ($n = 4$). (B) Proteins from TNF (TNFSF9/4-1BBL) and TNF receptor (TNFRSF1B/TNFR2, TNFRSF8/CD30, TNFRSF10A/DR4) superfamilies identified in proteomic analysis of CML EVs (K562-derived). (C) Western blot analysis of 4-1BBL/TNFSF9 protein in K562 (CML) and MOLM-14 (AML) cells and EVs. An equal amount of protein was loaded on gels. Data are representative for 4 experiments (CML EVs) and 2 experiments (AML EVs). (D) Western blot analysis of 4-1BBL/TNFSF9 protein in wt and 4-1BBL^{-/-} K562 (CML) cells and released EVs. An equal amount of protein was loaded on gels. Data from the same gel are presented (marked with dashed line). (E) Expression (fold change to CML wt EVs samples) of CD30, TNFR2, and LAG-3 after treatment of Tregs with CML EVs (K562-derived) from wt and 4-1BBL^{-/-} CML cells. For TNFR2, gMFI was analyzed; for CD30 and LAG-3, percent of positive cells. Data are from 3 experiments, $n = 6$. (F) In vitro suppressive activity of Tregs treated with either wt or 4-1BBL-deficient CML EVs (K562-derived), toward CD4⁺ responder T cells, pronounced as expansion index (EI) of responder cells (Tresp). Lower expansion index corresponds to higher suppressive activity. Data are from 3 experiments, $n = 6$. For panels E and F, mean \pm SD is presented, unpaired t -test with Welch's correction. * $P < .05$, ** $P < .01$, *** $P < .001$, **** $P < .0001$. gMFI, geometric mean fluorescence intensity; SD, standard deviation.

receptor is highly expressed on Tregs in tumors.^{32,41} Our conclusion is also supported by recent findings, which have demonstrated relevance of 4-1BBL/4-1BB signaling in Treg activation, physiological function, and transcriptomic identity.⁴⁶⁻⁴⁸ Our results pinpoint 4-1BBL as a new protein engaged in the upregulation of suppressive activity and eTreg phenotype in leukemia,

and we show, to our knowledge for the first time, that 4-1BBL signaling may occur via EVs. Such findings may have a diagnostic and therapeutic value, and 4-1BBL expression in leukemic EVs could be considered in liquid biopsy approaches as an early biomarker of leukemia and immunosuppression in CML/AML. Substantial advances in the extracellular vesicles field have been

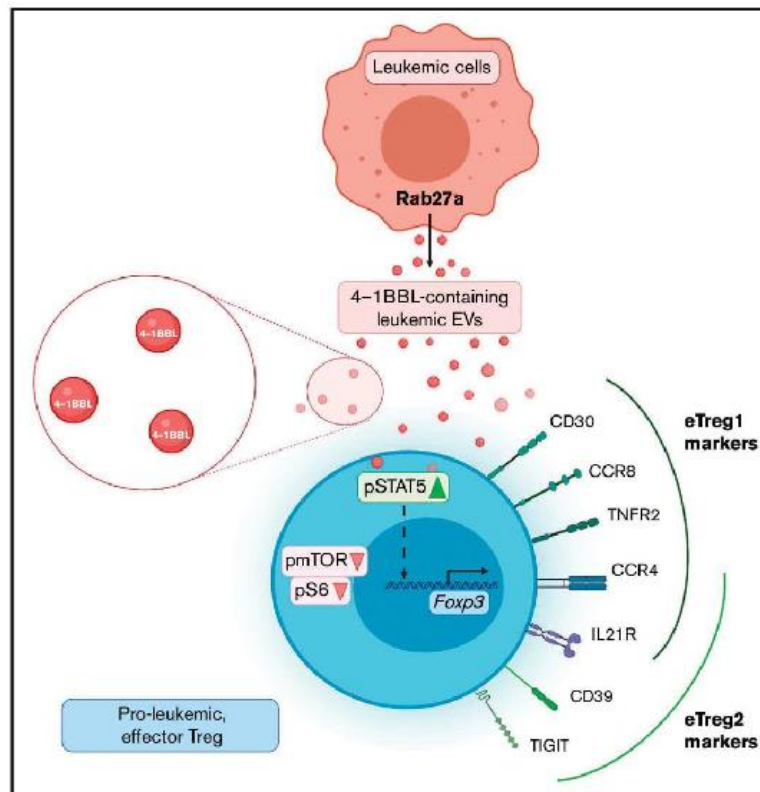


Figure 7. Proposed model of proleukemic eTreg generation, driven by 4-1BBL containing leukemic extracellular vesicles. Leukemic cells secrete Rab27a-dependent EVs that contain 4-1BBL protein. EVs are internalized by Tregs to upregulate STAT5 and attenuate mTOR-S6 signaling. As a result, Fc γ R3 expression is driven and Tregs upregulate specific EV-mediated phenotypic marker signatures (CD39, CCR8, CD30, TNFR2, CCR4, TIGIT, and IL21R), responsible for immunosuppressive, proleukemic function. Leukemic EVs-driven Tregs include 2 distinct eTreg subsets - CD30⁺CCR8⁺TNFR2⁺ eTreg1 and CD39⁺TIGIT⁺ eTreg2. CCR4 and IL21R are expressed by both eTreg1 and eTreg2 subsets.

developed for such approach, in terms of quick and effective phenotyping of EVs in plasma, for instance by EV-cytometry.⁶¹

In conclusion, we discovered a proleukemic, immunosuppressive mechanism, dependent on 4-1BBL-containing EVs derived from CML and AML cells. Leukemic EVs act as drivers of effector subsets and highly suppressive phenotype of Tregs. Our findings demonstrate the rationale to target Rab27a-dependent EV secretion, which may lead to prospective therapeutic applications aimed at attenuating immunosuppression in myeloid neoplasms.

Acknowledgments

The authors would like to acknowledge Grażyna Mosieniak for granting generous access to the Nanosight NS300 nanoparticle analyzer and Ewa Wąsiewicz for collecting clinical samples from leukemic patients. Figures 2A and 7 and supplemental Figure 5B

were created with biorender.com online software. RNA sequencing was performed at Genomics Core Facility (GeneCore) EMBL Heidelberg.

This work was supported by grants from the Polish National Science Centre: 2013/10/E/NZ3/00673 (K.P.), 2018/29/N/NZ3/01754 (J.S.), and 2014/15/B/NZ7/00966 (U.W.). J.S., K.P., and M.B.-O. were supported by the Foundation for Polish Science grant TEAM TECH Core Facility Plus/2017-2/2 (POIR.04.04.00-00-23C2/17-00, cofinanced by the European Union under the European Regional Development Fund). E.M. and J.M. were funded from the Foundation for Polish Science International Research Agendas Programme financed from the Smart Growth Operational 2014-2020 (Grant Agreement No. MAB/2018/6). S.D.B. has been a Marylou Ingram Scholar of the International Society for Advancement of Cytometry (ISAC) for the period 2016-2020.

Authorship

Contribution: J.S., L.T.-K., M.B.-O., A.K., B.V.L., and W.G.-P. performed ex vivo experiments; J.S., L.T.-K., B.V.L., S.C., P.P., and E.K. performed in vivo experiments and analyses; W.D. isolated RNA and prepared libraries for RNA sequencing; E.M. and J.M. analyzed RNA sequencing data; J.S., S.D.B., and A.C. analyzed and discussed flow cytometry data; D.C. performed proteomics; U.W., W.G.-P., and G.B. provided primary material; J.S. and K.P. conceptualized and supervised the project and experiments; J.S. prepared the figures and manuscript draft; J.S., L.T.-K., S.D.B., S.C., E.K., U.W., J.M., T.S., A.C., and K.P. prepared and reviewed final version of manuscript; and all authors have read and agreed to the published version of the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interest.

ORCID profiles: J.S., 0000-0001-8710-2808; L.T.-K., 0000-0002-1732-3397; S.D.B., 0000-0002-3217-9821; W.D., 0000-0002-7637-8480; A.K., 0000-0003-1567-9442; S.C., 0000-0001-9611-531X; P.P., 0000-0003-3354-439X; D.C., 0000-0001-6206-0672; E.K., 0000-0002-8655-851X; W.G.-P., 0000-0002-2971-9575; U.W., 0000-0002-4525-2004; G.B., 0000-0003-3858-8180; J.M., 0000-0002-2091-012X; A.C., 0000-0002-5381-1558; K.P., 0000-0001-6676-5282.

Correspondence: Katarzyna Piwocka, Laboratory of Cytometry, Nencki Institute of Experimental Biology, 3 Pasteur St, 02-093 Warsaw, Poland; e-mail: k.piwocka@nencki.edu.pl.

References

1. Dufva O, Pöönen P, Brück O, et al. Immunogenomic landscape of hematological malignancies [published correction appears in *Cancer Cell*. 2020;38(3):424–428]. *Cancer Cell*. 2020;38(9):380–399.e13.
2. Le Dieu R, Taussig DC, Ramsay AG, et al. Peripheral blood T cells in acute myeloid leukemia (AML) patients at diagnosis have abnormal phenotype and genotype and form defective immune synapses with AML blasts. *Blood*. 2009;114(18):3909–3916.
3. Swatler J, Tuross-Korgul L, Kozłowska E, Piwocka K. Immunosuppressive cell subsets and factors in myeloid leukemias. *Cancers (Basel)*. 2021;13(6):1203.
4. Pyzer AR, Stroopinsky D, Rajabi H, et al. MUC1-mediated induction of myeloid-derived suppressor cells in patients with acute myeloid leukemia. *Blood*. 2017;129(13):1791–1801.
5. Hughes A, Clareson J, Tang C, et al. CML patients with deep molecular responses to TKI have restored immune effectors and decreased PD-1 and immune suppressors. *Blood*. 2017;129(9):1166–1176.
6. Szczepanski MJ, Szajnik M, Czysłowska M, et al. Increased frequency and suppression by regulatory T cells in patients with acute myelogenous leukemia. *Clin Cancer Res*. 2009;15(10):3325–3332.
7. Brück O, Dufva O, Hohtari H, et al. Immune profiles in acute myeloid leukemia bone marrow associate with patient age, T-cell receptor clonality, and survival. *Blood Adv*. 2020;4(2):274–286.
8. Brück O, Blom S, Dufva O, et al. Immune cell contexture in the bone marrow tumor microenvironment impacts therapy response in CML. *Leukemia*. 2018;32(7):1643–1656.
9. Williams P, Basu S, Garcia-Manero G, et al. The distribution of T-cell subsets and the expression of immune checkpoint receptors and ligands in patients with newly diagnosed and relapsed acute myeloid leukemia. *Cancer*. 2019;125(9):1470–1481.
10. Lambie AJ, Lind EF. Targeting the immune microenvironment in acute myeloid leukemia: a focus on T cell immunity. *Front Oncol*. 2018;8:213.
11. Wang R, Feng W, Wang H, et al. Blocking migration of regulatory T cells to leukemic hematopoietic microenvironment delays disease progression in mouse leukemia model. *Cancer Lett*. 2020;469:151–161.
12. Irani YD, Hughes A, Clareson J, et al. Successful treatment-free remission in chronic myeloid leukaemia and its association with reduced immune suppressors and increased natural killer cells. *Br J Haematol*. 2020;191(3):433–441.
13. Tanaka A, Sakaguchi S. Targeting Treg cells in cancer immunotherapy. *Eur J Immunol*. 2019;49(8):1140–1146.
14. Zhou Q, Munger ME, Highfill SL, et al. Program death-1 signaling and regulatory T cells collaborate to resist the function of adoptively transferred cytotoxic T lymphocytes in advanced acute myeloid leukemia. *Blood*. 2010;116(14):2484–2493.
15. Curti A, Pandolfi S, Valzasina B, et al. Modulation of tryptophan catabolism by human leukemic cells results in the conversion of CD25 Δ into CD25 Δ T regulatory cells. *Blood*. 2007;109(7):2871–2877.
16. Zhou Q, Munger ME, Veenstra RG, et al. Coexpression of Tim-3 and PD-1 identifies a CD8 α T-cell exhaustion phenotype in mice with disseminated acute myelogenous leukemia. *Blood*. 2011;117(17):4501–4510.
17. Swatler J, Dudka W, Piwocka K. Isolation and characterization of extracellular vesicles from cell culture conditioned medium for immunological studies. *Curr Protoc Immunol*. 2020;129(1):e96.
18. Marar C, Starich B, Wirtz D. Extracellular vesicles in immunomodulation and tumor progression. *Nat Immunol*. 2021;22(5):560–570.
19. Poggio M, Hu T, Pai C-C, et al. Suppression of exosomal PD-L1 induces systemic anti-tumor immunity and memory. *Cell*. 2019;177(2):414–427.e13.
20. Yin Y, Cai X, Chen X, et al. Tumor-secreted miR-214 induces regulatory T cells: a major link between immune evasion and tumor growth. *Cell Res*. 2014;24(10):1164–1180.
21. Mrizak D, Martin N, Barjon C, et al. Effect of nasopharyngeal carcinoma-derived exosomes on human regulatory T cells. *J Natl Cancer Inst*. 2014;107(1):363.

22. Nehrbas J, Butler JT, Chen D-W, Kurre P. Extracellular vesicles and chemotherapy resistance in the AML microenvironment. *Front Oncol.* 2020;10:90.
23. Raimondo S, Saieva L, Corrado C, et al. Chronic myeloid leukemia-derived exosomes promote tumor growth through an autocrine mechanism. *Cell Commun Signal.* 2015;13(1):8.
24. Mineo M, Garfield SH, Taverna S, et al. Exosomes released by K562 chronic myeloid leukemia cells promote angiogenesis in a Src-dependent fashion. *Angiogenesis.* 2012;15(1):33-45.
25. Huan J, Hornick NI, Goloviznina NA, et al. Coordinate regulation of residual bone marrow function by paracrine trafficking of AML exosomes. *Leukemia.* 2015;29(12):2285-2295.
26. Dumontet E, Pangault C, Roulois D, et al. Extracellular vesicles shed by follicular lymphoma B cells promote polarization of the bone marrow stromal cell niche. *Blood.* 2021;138(1):57-70.
27. Szczepanski MJ, Szajnik M, Welsh A, Whiteside TL, Boyiadzis M. Blast-derived microvesicles in sera from patients with acute myeloid leukemia suppress natural killer cell function via membrane-associated transforming growth factor-beta1. *Haematologica.* 2011;96(9):1302-1309.
28. Swatler J, Dudka W, Bugajski L, Brewinska-Olchowik M, Kozłowska E, Piwocka K. Chronic myeloid leukemia-derived extracellular vesicles increase Foxp3 level and suppressive activity of thymic regulatory T cells. *Eur J Immunol.* 2020;50(4):606-609.
29. Théry C, Witwer KW, Aikawa E, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles.* 2018;7(1):1535750.
30. Van Deun J, Mestdagh P, Agostinis P, et al; EV-TRACK Consortium. EV-TRACK: transparent reporting and centralizing knowledge in extracellular vesicle research. *Nat Methods.* 2017;14(3):228-232.
31. Van Gassen S, Callebaut B, Van Helden MJ, et al. FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry A.* 2015;87(7):636-645.
32. Wing JB, Tanaka A, Sakaguchi S. Human FOXP3⁺ regulatory T cell heterogeneity and function in autoimmunity and cancer. *Immunity.* 2019;50(2):302-316.
33. Ostrowski M, Carmo NB, Krumeich S, et al. Rab27a and Rab27b control different steps of the exosome secretion pathway. *Nat Cell Biol.* 2010;12(1):1-9-30, 1-13.
34. Moesta AK, Li X-Y, Smyth MJ. Targeting CD39 in cancer. *Nat Rev Immunol.* 2020;20(12):739-755.
35. Sosnowska A, Chlebowska-Tuz J, Matryba P, et al. Inhibition of arginase modulates T-cell response in the tumor microenvironment of lung carcinoma. *Oncol Immunology.* 2021;10(1):1956143.
36. Pilanc P, Wojnicki K, Roura A-J, et al. A Novel oral arginase 1/2 inhibitor enhances the antitumor effect of PD-1 inhibition in murine experimental gliomas by altering the immunosuppressive environment. *Front Oncol.* 2021;11:703465.
37. Haiqi H, Yong Z, Yi L. Transcriptional regulation of Foxp3 in regulatory T cells. *Immunobiology.* 2011;216(6):678-685.
38. Lu L, Barbi J, Pan F. The regulation of immune tolerance by FOXP3. *Nat Rev Immunol.* 2017;17(11):703-717.
39. Delgoffe GM, Kole TP, Zheng Y, et al. The mTOR kinase differentially regulates effector and regulatory T cell lineage commitment. *Immunity.* 2009;30(6):832-844.
40. Dominguez-Villar M, Hafler DA. Regulatory T cells in autoimmune disease. *Nat Immunol.* 2018;19(7):665-673.
41. Alvisi G, Brummelman J, Puccio S, et al. IRF4 instructs effector Treg differentiation and immune suppression in human cancer. *J Clin Invest.* 2020;130(6):3137-3150.
42. Plitas G, Konopacki C, Wu K, et al. Regulatory T cells exhibit distinct features in human breast cancer. *Immunity.* 2016;45(5):1122-1134.
43. Azizi E, Carr AJ, Plitas G, et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell.* 2018;174(5):1293-1308.e36.
44. De Simone M, Arrighi A, Rossetti G, et al. Transcriptional landscape of human tissue lymphocytes unveils uniqueness of tumor-infiltrating T regulatory cells. *Immunity.* 2016;45(5):1135-1147.
45. Magnuson AM, Kiner E, Ergun A, et al. Identification and validation of a tumor-infiltrating Treg transcriptional signature conserved across species and tumor types. *Proc Natl Acad Sci USA.* 2018;115(45):E10672-E10681.
46. Ronin E, Lubrano di Ricco M, Vallion R, et al. The NF-κB RelA transcription factor is critical for regulatory T cell activation and stability. *Front Immunol.* 2019;10:2487.
47. Vasanthakumar A, Liao Y, Teh P, et al. The TNF receptor superfamily-NF-κB axis is critical to maintain effector regulatory T cells in lymphoid and non-lymphoid tissues. *Cell Rep.* 2017;20(12):2906-2920.
48. Lubrano di Ricco M, Ronin E, Collares D, et al. Tumor necrosis factor receptor family costimulation increases regulatory T-cell activation and function via NF-κB. *Eur J Immunol.* 2020;50(7):972-985.
49. Johnson JL, Ramadass M, He J, et al. Identification of neutrophil exocytosis inhibitors (Nexinhibs), small molecule inhibitors of neutrophil exocytosis and inflammation: druggability of the small GTPase Rab27a. *J Biol Chem.* 2016;291(50):25965-25982.
50. Tallon C, Hollinger KR, Pal A, et al. Nipping disease in the bud: nSMase2 inhibitors as therapeutics in extracellular vesicle-mediated diseases. *Drug Discov Today.* 2021;26(7):1656-1668.
51. Pric M, Mair F, Erickson J, et al. Extricating human tumor-unique immune alterations from non-malignant tissue inflammation. *Nat. Port.* 2021.

52. Kleijwegt FS, Laban S, Duinkerken G, et al. Critical role for TNF in the induction of human antigen-specific regulatory T cells by tolerogenic dendritic cells. *J Immunol*. 2010;185(3):1412-1418.
53. Lucca LE, Dominguez-Villar M. Modulation of regulatory T cell function and stability by co-inhibitory receptors. *Nat Rev Immunol*. 2020;20(11):680-693.
54. Piitas G, Rudensky AY. Regulatory T cells: differentiation and function. *Cancer Immunol Res*. 2016;4(9):721-725.
55. Hirata Y, Furuhashi K, Ishii H, et al. CD150^{hi} bone marrow Tregs maintain hematopoietic stem cell quiescence and immune privilege via adenosine. *Cell Stem Cell*. 2018;22(3):445-453.e5.
56. Morita K, Okamura T, Inoue M, et al. Egr2 and Egr3 in regulatory T cells cooperatively control systemic autoimmunity through Ltbp3-mediated TGF- β 3 production. *Proc Natl Acad Sci USA*. 2016;113(50):E8131-E8140.
57. Carpenter AC, Wohlfert E, Chopp LB, et al. Control of regulatory T cell differentiation by the transcription factors Thpok and LRF. *J Immunol*. 2017;199(5):1716-1726.
58. Zheng Y, Josefowicz S, Chaudhry A, Peng XP, Forbush K, Rudensky AY. Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate. *Nature*. 2010;463(7282):808-812.
59. La Trobe University. Bundoora, Victoria. ExoCarta: Exosome protein, RNA and lipid database. <http://exocarta.org/index.html>. Accessed 27 July 2021.
60. Lee S-W, Park Y, So T, et al. Identification of regulatory functions for 4-1BB and 4-1BBL in myelopoiesis and the development of dendritic cells. *Nat Immunol*. 2008;9(8):917-926.
61. Rikkert LG, Beekman P, Caro J, et al. Cancer-ID: toward identification of cancer by tumor-derived extracellular vesicles in blood. *Front Oncol*. 2020;10:608.

Downloaded from <https://ashpublications.org/blood/advance-article-abstract/doi/10.1182/blood.2021.009105> by guest on 26 September 2025

Warsaw, 26.09.2023

Dr. hab. Katarzyna Piwocka, Professor of the Nencki Institute
(professional title, name and surname)

STATEMENT

As the corresponding author of the work entitled **4-1BBL-containing leukemic extracellular vesicles promote immunosuppressive effector regulatory T cells** I declare hereby that my own substantive contribution to the preparation, implementation and analysis of this study and presentation of experimental work in the form of above mentioned publication is **15%**. **In this work, together with the first author (my PhD student Julian Swatler), I put forward the research hypothesis, I conceptualized and designed the study. I supervised the experiments, data analysis and interpretation. I coordinated preparation of the first version of the manuscript, figures and worked with other co-authors on the final version. The study were partially financed by National Science Centre Sonata Bis grant in which I was PI, as well as the TEAM-TECH Core Facility Plus/2017-2/2 Grant POIR.04.04.00-00-23C2/17-00 from the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund in which I was PI.**

At the same time, I consent to the submission of the above-mentioned work by **Elyas Mohhamadi** as part of a doctoral dissertation in the form of a thematically coherent collection of articles published in scientific journals.

I hereby declare that an independent and separable part of the above work shows the individual contribution of **Elyas Mohhamadi** when developing the concept, analysis and presentation of the transcriptomics data (Fig. 4).

Podpis jest prawidłowy

Dokument podpisany przez
Katarzyna Piwocka
Data: 2023.09.26 10:54:02 CEST

.....
(co-author signature)

Milano, IT, 25.09.2023

Julian Swatler, PhD

STATEMENT

As the leading author of the work entitled **4-1BBL-containing leukemic extracellular vesicles promote immunosuppressive effector regulatory T cells** I declare hereby that my own substantive contribution to the preparation, implementation and analysis of this study and presentation of experimental work in the form of above mentioned publication is **55%**. **In this work, together with the principal investigator/corresponding author (Professor Katarzyna Piwocka), I put forward the research hypothesis, I conceptualized and designed the study. In the presented study, I performed majority of the experiments, performed data analysis and interpretation. I prepared manuscript figures, wrote the manuscript draft and worked with other co-authors on the final version. The study was partially financed by National Science Centre Preludium grant in which I was PI.**

At the same time, I consent to the submission of the above-mentioned work by **Elyas Mohhamadi** as part of a doctoral dissertation in the form of a thematically coherent collection of articles published in scientific journals.

I hereby declare that an independent and separable part of the above work shows the individual contribution of **Elyas Mohhamadi** when developing the analysis and presentation of the transcriptomics data (Fig. 4).

.....
Julian Swatler
(co-author signature)

Warsaw, Poland, 25.09.2023

Laura Tuross-Korgul, MSc

STATEMENT

As the co-author of the work entitled **4-1BBL-containing leukemic extracellular vesicles promote immunosuppressive effector regulatory T cells** I declare hereby that my own substantive contribution to the preparation, implementation and analysis of this study and presentation of experimental work in the form of above mentioned publication is **10% and comprise: *in vivo* and *ex vivo* experiments the results of which are part of the results presented in Figures 2 and 3, preparation of the final version of the manuscript.**

At the same time, I consent to the submission of the above-mentioned work by **Elyas Mohhamadi** as part of a doctoral dissertation in the form of a thematically coherent collection of articles published in scientific journals.

I hereby declare that an independent and separable part of the above work shows the individual contribution of **Elyas Mohhamadi** when developing the analysis and presentation of the transcriptomics data (Fig. 4).


.....
(co-author signature)

Gdansk, 11th of September 2023

Dr. hab, Jakub Mieczkowski
(professional title, name and surname)

STATEMENT

As the co-author of the work entitled **4-1BBL-containing leukemic extracellular vesicles promote immunosuppressive effector regulatory T cells** I declare hereby that my own substantive contribution to the preparation, implementation and analysis of this study and presentation of experimental work in the form of above mentioned publication is **5% and comprise** supervising the analysis of RNA sequencing data.

At the same time, I consent to the submission of the above-mentioned work by **Elyas Mohhamadi** as part of a doctoral dissertation in the form of a thematically coherent collection of articles published in scientific journals.

I hereby declare that an independent and separable part of the above work shows the individual contribution of **Elyas Mohhamadi** when developing the concept, performing the experimental part, developing and interpretation of the results of this work.


.....
(co-author signature)

RESEARCH ARTICLE

Improvement of the performance of anticancer peptides using a drug repositioning pipeline

Elyas Mohammadi^{1,2,3} | Mojtaba Tahmoorespur¹ | Rui Benfeitas⁴ | Ozlem Altay² | Ali Javadmanesh¹ | Simon Lam⁵ | Adil Mardinoglu^{2,5} | Mohammad Hadi Sekhavati¹

¹ Department of Animal Science, Ferdowsi University of Mashhad, Mashhad, Iran

² Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH – Royal Institute of Technology, Stockholm, Sweden

³ 3P-Medicine Laboratory, Medical University of Gdańsk, Gdańsk, Poland

⁴ National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

⁵ Centre for Host–Microbiome Interactions, Faculty of Dentistry, Oral and Craniofacial Sciences, King's College London, London, UK

Correspondence

Mohammad Hadi Sekhavati, Azadi square, Ferdowsi University of Mashhad, Mashhad, 91779–4B97, Iran.
Email: sekhavati@um.ac.ir

Abstract

The use of anticancer peptides (ACPs) as an alternative/complementary strategy to conventional chemotherapy treatments has been shown to decrease drug resistance and/or severe side effects. However, the efficacy of the positively-charged ACP is inhibited by elevated levels of negatively-charged cell-surface components which trap the peptides and prevent their contact with the cell membrane. Consequently, this decreases ACP-mediated membrane pore formation and cell lysis. Negatively-charged heparan sulphate (HS) and chondroitin sulphate (CS) have been shown to inhibit the cytotoxic effect of ACPs.

In this study, we propose a strategy to promote the broad utilization of ACPs. In this context, we developed a drug repositioning pipeline to analyse transcriptomics data generated for four different cancer cell lines (A549, HEPG2, HT29, and MCF7) treated with hundreds of drugs in the LINCS L1000 project. Based on previous studies identifying genes modulating levels of the glycosaminoglycans (GAGs) HS and CS at the cell surface, our analysis aimed at identifying drugs inhibiting genes correlated with high HS and CS levels. As a result, we identified six chemicals as likely repositionable drugs with the potential to enhance the performance of ACPs. The codes in R and Python programming languages are publicly available in https://github.com/ElyasMo/ACPs_HS_HSPGs_CS.

As a conclusion, these six drugs are highlighted as excellent targets for synergistic studies with ACPs aimed at lowering the costs associated with ACP-treatment.

KEYWORDS

cancer, drug repositioning, heparan sulfate, LINCS L1000, therapeutic peptides

Abbreviations: <https://doi.org/10.1002/biot.202100417> | https://pubs.rsc.org/en/glossary/canonical_smiles, The simplified molecular-input line-entry system (SMILES) is a specification in the form of a line notation for describing the structure of chemical species using short ASCII strings; cell_id, A shorthand CMap identifier number assigned to each cell line used in the L1000 assay; Direction, Shows Up/Down regulation of HC and CS genes while they are affected by drug; inchIKey, The IUPAC International Chemical Identifier (InChI) is a textual identifier for chemical substances, designed to provide a standard way to encode molecular information and facilitate searching for such information in databases and on the web; pert_name, The internal (CMap-designated) name of a drug. By convention, for genetic perturbations CMap uses the HUGO gene symbol.

1 | INTRODUCTION

Cancer is known as the second leading cause of death after cardiovascular disease.^[1] Due to the heterogeneous nature of cancer in the same type of tumour in different individuals (intertumoral heterogeneity) and among cancer cells within the same tumour (intratumoral heterogeneity), finding an accurate and reliable treatment has always been a challenge for scientists.^[2] Current treatment methods are costly and cause adverse side effects.^[1,3] For instance, the oxidative stress-mediated injury of doxorubicin as a conventional chemotherapeutic drug has adverse effects on the kidney,^[4] brain,^[5] and heart.^[6] Additionally, tumours resistant to chemotropic drugs, for example, metastatic breast cancers resistant to taxanes and anthracyclines, are another obstacle impeding development of effective treatment.^[7] Hence, investigation of an alternative method to help address conventional chemotropic drugs' limitations is crucial.

Cation therapeutic peptides with anti-cancerous features are toxic to cancer cells.^[8] Compared with antibodies and small molecules, anti-cancer peptides (ACPs) are more selective and benefit from higher penetration rate and easy modifications.^[9] Two of the significant modes of action of ACPs are apoptosis and necrosis by membrane lysis or pore formation.^[10] Contrary to malignant cells, healthy cells are protected against ACPs due to their membrane's unique features. For instance, a higher level of cholesterol in healthy cells inhibits lytic activity and protects the membrane against the lytic action of ACPs by modifying membrane fluidity.^[11] On the other hand, cancer cells show more membrane fluidity and abundant microvilli, which increase the cell surface area.^[12] In addition to this, negatively-charged components on the surface of malignant cells cause membrane destabilization, cytotoxicity and cancer cell lysis when interacting with ACPs.^[13] This is not an issue for healthy cells since they have electrically neutral cell surfaces.^[14] Finally, it should be noted that the primary force of interaction between ACPs and healthy cell membrane are hydrophobic interactions. In contrast, the forces between ACPs and cancer cells are electrostatic interactions.^[15] As a result, no matter how heterogeneous cancer cells are, ACPs can preferentially eliminate a wide range of malignant cells.^[9]

It has been shown that some components at the surface of the cancer cell membrane have an influence on cell susceptibility to ACPs.^[16,17] ACPs interact with the two glycosaminoglycans (GAGs) heparan sulphate (HS) and chondroitin sulphate (CS), which are present on the surface of most cells. Although negatively-charged molecules at the cell surface of cancer cells increase the selectivity of ACPs, some of them actually inhibit their cytotoxic activity. In other words, HS and CS at the surface of cancer cells sequester ACPs away from the phospholipid bilayer, and thereby impede their ability to induce cytotoxicity.^[17] As such, further investigation on HS and CS and how to modify their aggregation at the cell surface is needed. For instance, the genes that directly or indirectly cause changes in the level of HS and CS at the surface of cancer cells or the drugs that escalate or diminish the amount of these cell-surface components are currently under-studied. Using the aforementioned drugs to study their synergistic effect with ACPs in further studies may decrease

the cost of utilization of ACPs and improve their performance. In addition, according to our developed pipeline, safety-certified chemicals would be used in synergistic studies to find new functions for previously known perturbations. The drug repositioning approach may significantly reduce the required time and investment in drug development.^[18,19]

2 | MATERIAL AND METHODS

2.1 | Parsing the data and data manipulation

We retrieved normalized gene expression profiles of landmark genes and imputed transcripts for four different cancer cell lines, A549, HEPG2, MCF7, and HT-29 (Table 1), treated with either drug or control (DMSO), from the LINCS L1000 database for a total of 12328 genes more than 900 drugs. To this end, normalized gene expression profiles of landmark genes and imputed transcripts were parsed through the Slinky R package (version 1.8.0)^[20] for both control (treated with DMSO) and treatment (treated with various drugs) with the highest standard dose (10 μ m) and longest time points (24 h) (Figure 1A).

Log fold changes (LFC) were computed through the NumPy library (version 1.19.1) in Python 3.7.6. To avoid undefined LFC values due to division by zero or \log_2 transformation of non-positive numbers, one was added to gene expression values for both treatment and control before transformation. As a result, we produced LFCs for four different cancer cell lines across about 12,000 genes and 900 drugs (Figure 1B).

2.2 | Computing gene-gene correlation

LFCs were used to compute gene-gene correlations. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Centre for Advanced Computational Science. Spearman (SP), Pearson (PE), and Kendall tau rank (KE) correlation coefficients were determined using SciPy (version 1.5.2) `scipy.stats.spearmanr`, `scipy.stats.pearsonr`, and `scipy.stats.kendalltau` functions, respectively.^[21] All three correlation methods were applied to the A549 cancer cell line LFCs to determine the best statistical method to calculate gene-gene correlation in the other three cancer cell lines (Figure 1C). Computational optimisation was achieved by omitting correlations with self and correlations reciprocal to those already tested. False discovery rate (FDR) was determined from correlation p-values by `scipy.stats.rankdata`.^[21]

2.3 | Functional analysis

To determine the most meaningful correlation method, correlations were sorted based on each method's FDRs and the top 100 and 500 co-expressed gene pairs were considered for functional analysis, as previously described.^[22] ClusterProfiler R package (version 3.16.1)^[23] was

TABLE 1 Cancer cell lines

Cell id	Primary site	Subtype	Growth pattern	ATCC	Age	Sex
A549	Lung	Non-small cell lung cancer carcinoma	Adherent	CCL-185	58	M
HEPG2	Liver	Hepatocellular carcinoma	Adherent	HB-8065	15	M
HT29	Large intestine	Colorectal adenocarcinoma	Adherent	HTB-38	44	F
MCF7	Breast	Adenocarcinoma	Adherent	HTB-22	69	F

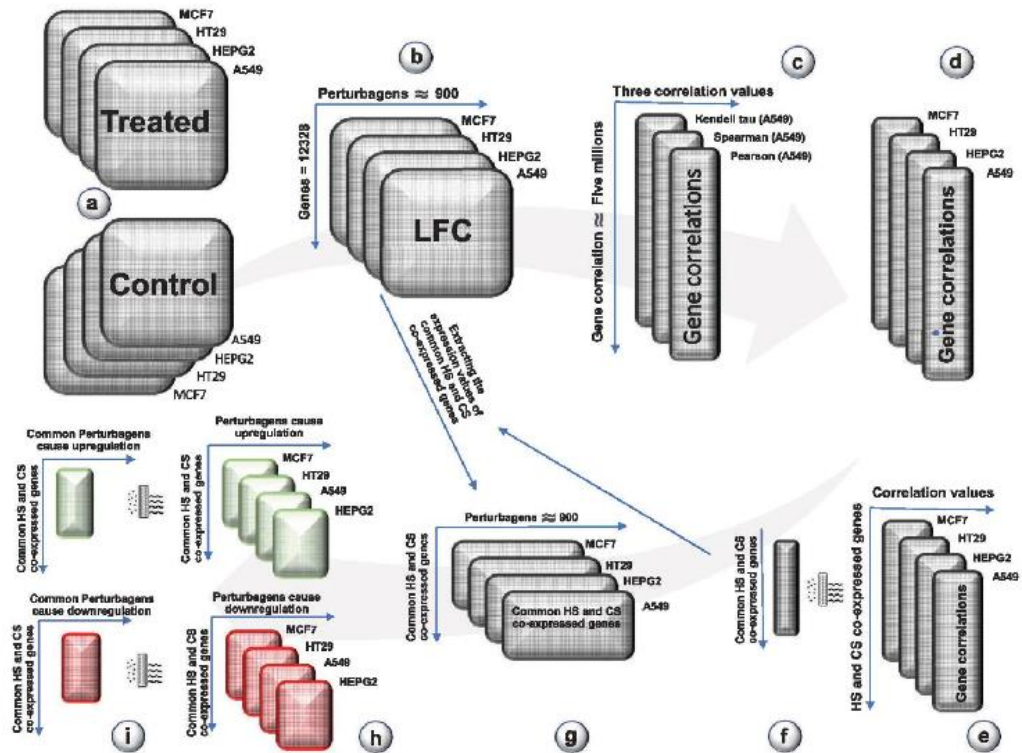


FIGURE 1 LINC1000 provides the opportunity to discover signals from a vast amount of data. (A) Control and drug-treated transcriptomic datasets for four cancer cell lines were retrieved from the LINC1000 database. (B) Log fold changes (LFCs) for expression values were computed for all four cancer cell lines. (C) To find a reference statistical method to compute gene-gene correlations, three methods (Pearson, Spearman, Kendall tau) were evaluated using the A549 LFC matrix. (D) The chosen statistical method to compute gene-gene correlations was applied to all four cancer cell lines. (E) The correlations with experimentally validated HC and CS genes were extracted. (F) As the first filtration step, only common co-expressed genes with experimentally supported HC and CS genes in all four gene-gene correlation data frames (corresponding to four cancer cells) were considered. (G) Retrieved genes from step 6 and their expression values along all perturbations were extracted from four matrices in step 2. (H) The drugs which cause up or downregulations in most of the selected genes were placed in two separated data frames; red: downregulated genes, green: upregulated genes. (I) As the second filtration step, drugs that jointly cause up or down-regulation of most selected genes in all four cancer cells were extracted

used to find enriched terms based on the two sets of significantly co-expressed gene pairs. The correlation method which resulted in more enriched terms in the functional analysis, was chosen and applied to three remaining cancer cell lines (Figure 1D).

2.4 | HS- and CS-related genes

To extract gene-gene correlations related to HS and CS, we first retrieved experimentally validated genes from the literature.^[24,25]

Genes that were highly correlated with genes in this list were considered in the downstream analysis (Figure 1E).

2.5 | Discovering probable common co-expressed genes with HS and CS genes

To diminish the probability of false positives, for each experimentally validated gene, common co-expressed genes across all four cancer cell lines were retrieved (Figure 1F). The top ten (based on FDR) common co-expressed genes for each experimentally validated gene were accepted for downstream analysis. Across all drugs, LFCs of these chosen sets of genes, in addition to the Lab-identified genes, were accepted for downstream analysis. (Figure 1G).

To investigate the pathways related to the selected genes, KEGG and gene ontology (GO) analysis was performed by clusterProfiler.^[23]

2.6 | Discovering common drugs which are likely to cause up/downregulation in HS and CS genes

To monitor the effect of drugs on selected genes' expression in the four different cancer cell lines, heatmaps were generated of LFCs per drug. By inspection of the heatmap, drugs that did not depict a significant gene up- or downregulation effect were deemed non-effective and omitted from further analysis. The remaining drugs were categorized into up and down groups according to the direction of their gene regulation (Figure 1H). Again, to diminish the probability of false positives, only common drugs in all four cancer cell lines which cause downregulation were investigated to be proposed as promising repositionable drugs and a candidate in complementary treatment studies with ACPs against cancer (Figure 1I).

3 | RESULTS

3.1 | Parsing the data and data manipulation

The LINCS L1000 project as a new gene expression profiling method has provided an excellent opportunity to study the mechanism of action of small molecules, functionally annotate genetic variants of disease genes, and inform clinical trials by collecting gene expression profiles for thousands of drugs at a variety of time points, doses, and cell lines.^[26] Taking the massive amount of data produced by LINCS L1000 into account and parsing the data would be computationally intensive. To address this issue, various methods have been developed.^[20,27] In this study, four cancer cell lines, including A549, HEPG2, MCF7, and HT-29 were considered due to their extensive use in ACP studies (Table 1).^[28,29] We determined log fold change (LFC) values for drug compared to control for each gene in each cancer cell line. LFC describes how much expression values change between these two conditions.

TABLE 2 GO and KEGG pathway enrichment analysis to determine the optimal gene-gene correlation method, Pearson (PE), Spearman (SP), or, Kendall tau (KE)

Statistical method	Input	Number of enriched terms	
		GO	KEGG
PE	First 100 pairs	0	3
	First 500 pairs	44	8
SP	First 100 pairs	0	0
	First 500 pairs	5	2
KE	First 100 pairs	5	8
	First 500 pairs	3	2

Abbreviations: PE, Pearson; SP, Spearman; KE, Kendall tau.

3.2 | Functional analysis

To find the best statistical method which discovers the most meaningful co-expression correlations in our datasets, Spearman (SP), Pearson (PE), and Kendall tau (KE) methods were evaluated for the A549. PE outperformed SP and KE based on the number of significantly enriched terms (adjusted p -value < 0.5) (Table 2) (Figure 2).

Concerning the top 100 co-expressed gene pairs, both KEGG and GO pathway analysis depict better KE performance compared with PE and SP. However, looking at the top 500 co-expressed gene pairs, PE depicts significantly more enriched terms (Figure S1A, Table 2) and generally more involved genes in enriched terms (Figure 2) in both KEGG and GO pathway analysis and was chosen as the reference statistical method to compute gene-gene correlations for the remaining cancer cell lines (HEPG2, HT29, and, MCF7).

3.3 | Computing gene-gene correlation

As the reference statistical method, PE was applied to the remaining cancer cell lines (HEPG2, HT29 and MCF7) to calculate gene-gene co-expression correlations. Considering those gene-gene correlations that appeared in all four cell lines and involving genes already known in the literature to be associated with HS and CS (Table 3; first row: experimentally validated genes^[17,24,25]), one and zero gene correlations were found for AGRN and CHST3, respectively. Ten significantly co-expressed genes were introduced for the rest of the laboratory-validated HS and CS genes. Considering that the in silico-driven gene-gene associations are common between all four cancer cell lines, biological correlations between these genes could be expected.

To investigate the pathways related to selected genes (Table 3), we conducted KEGG and GO analysis. GO analysis identified Golgi lumen, which is significant as this is where EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyses the synthesis of HS.^[30] In addition to this, collagen-containing extracellular matrix, where HS and CS are available,^[31] was the highest enriched term (Figure 3A, a and b). On the other hand, KEGG pathway analysis

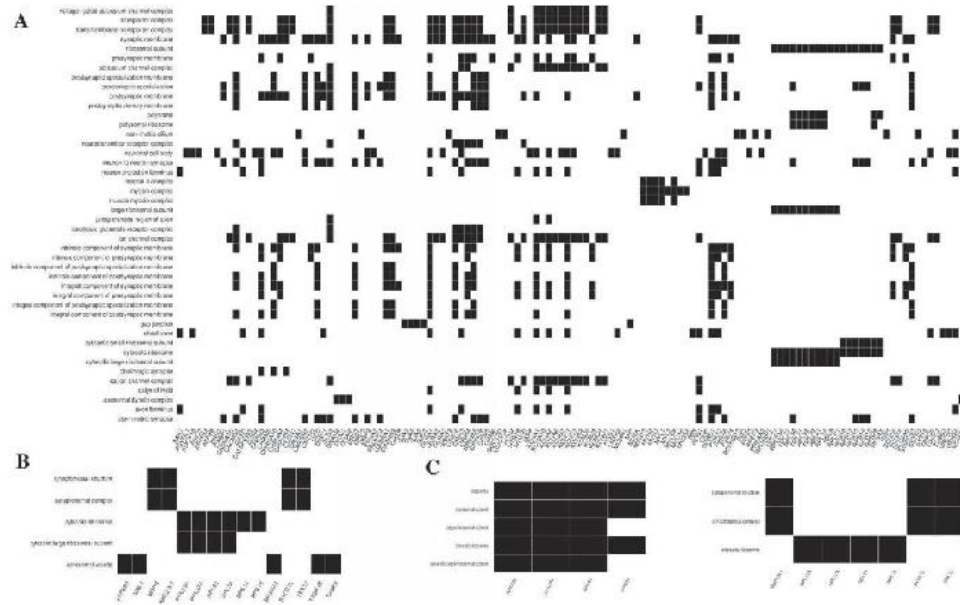


FIGURE 2 Heatplots of GO enriched terms and their corresponding genes to illustrate the performance of different statistical methods in finding meaningful gene-gene co-expression correlations. (A) Pearson correlation (PE) using the top 500 gene pairs. (B) Spearman correlation (SP) using the top 500 gene pairs. (C) Kendall tau rank correlation (KE) using the top 100 (left) and top 500 (right) gene pairs

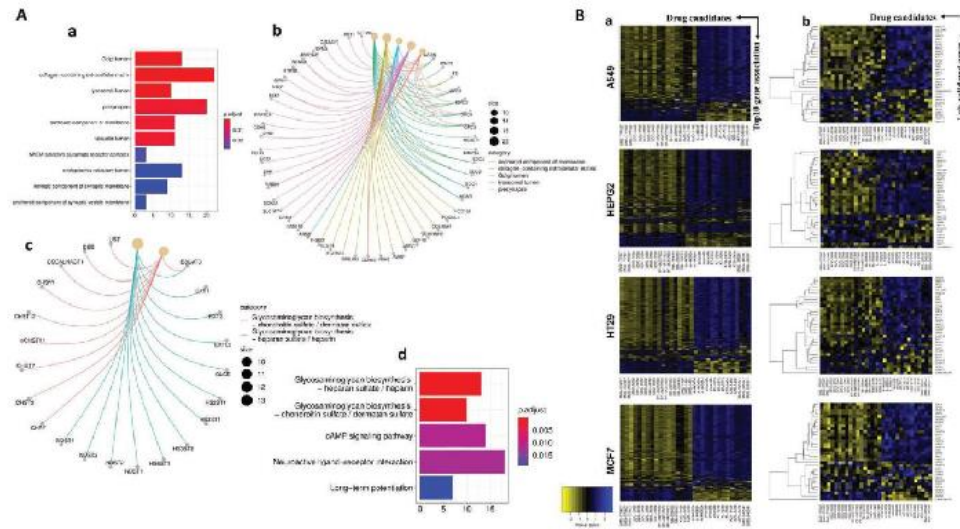


FIGURE 3 Monitoring the gene functions and their expression alterations against drugs. (A) Pathway analysis of selected genes revealed enriched terms related to H5 and CS. a) GO analysis. b) Illustration of genes that are related to GO pathways. c) KEGG analysis. d) Visualisation of genes that are associated with the first two KEGG pathways. (B) There is a consistency between the effect of proposed drugs on all introduced HS and CS genes and only lab-validated HS and CS genes. a) The effect of proposed drugs on all introduced HS and CS genes and b) lab-validated HS and CS genes

TABLE 3 The list of HS and CS top ten common co-expressed genes. Laboratory validated genes and their associated co-expressed genes are presented in the first column and rows respectively

Experimentally validated genes	Predicted co-expressed genes with experimentally validated CS and/or HS genes									
<i>AGRN</i>	ZNF580									
<i>B3GAT3</i>	BSCL2	COX8A	DRAP1	MAP2K2	CD151	LSM4	RBM5	FIBP	GPR87	UBE2M
<i>CHPF</i>	SLC4A3	TRIM46	ADAMTS13	SCNN1D	SLC8A2	CORT	HOXA3	RAB11B	THRB	OBSCN
<i>CHST3</i>										
<i>CHST7</i>	JAKMIP2	SLCO1C1	GRIA1	CYP19A1	CD68	SMARCA1	CHST2	FGR	GP5M2	MMP19
<i>CHST11</i>	UBQLN4	alpk1	IPAR3	ZNF665	sall2	PODNL1	KIAA0141	DAPP1	PLVAP	MRNIP
<i>CHST12</i>	CMTR1	WDR74	POGK	THUMPD2	COA7	lrch4	SMG9			
<i>CHSY1</i>	YTHDF2	ZNF154	BARX1	WNT3	PCDHGB5	KRT83	APOL5	SAMD4B	KALRN	OR2C1
<i>COL18A1</i>	SLC6A12	SERPINF2	UPB1	DNAJA1	F11	HNF4A	PKLR	ASGR1	AGPAT3	F2
<i>CSGALNACT1</i>	AFF1	SLC1A7	CDIP1	GDF10	IL9R	NOCT	INSL4	KLK12	ARTN	IL1B
<i>DSE</i>	GLOD4	FOXL2	TUBA8	CLEC4A	FGD2	PSMG2	TEX14	TLR7	PCDHB13	VWA7
<i>EXT1</i>	ACVRL1	ZNF141	ACKR4	TRIM13	KCNH6	adamts12	PTK2B	SCN2A	PTPRF	PAFAH1B1
<i>EXT2</i>	ZC3H11A	GLOD4	IST1	NUP62	RPL36AL	TXLNA	MRFAP1L1	RAB5A	CHMP7	SNW1
<i>EXTL2</i>	GLMN	SLC25A17	ACADM	ARF1	ZKSCAN3	SLC12A3	INPP4A	SYP	DDX39A	FPGT
<i>GLCE</i>	ALDH1A1	WWP1	SGTA	PCGF1	GUK1	slc35a1	btg3	TRMT1	PIP5K1B	RARRES1
<i>GPC1</i>	C22orf24	NEU3	CDRT1	pde1b	ADRA2B	KLHL1	SAG	RPL36AL	KIAA0391	CRMP1
<i>GPC3</i>	AFP	APOA2	AHSG	DUSP9	SERPIND1	SOAT2	AMBP	ITIH2	emx1	BMP10
<i>GPC4</i>	TREM2	STK19	KDR	sall2	YIPF3	LINC00894	CDH13	KIF17	GHRHR	PCLO
<i>GPC5</i>	OR2J3	CLSPN	PBOV1	DRD1	B3GALT1	GPR85	MYO3A	CDH9	GS1-600G8.3	OR2J2
<i>HS2ST1</i>	DHX15	CLTC	L5M4	ARL4C	NUP62	POP4	KATNB1	ATP1B1	PPP1CA	WDRB2
<i>HS3ST1</i>	ELK1	SEL1L	RBM5	SELL	CHKB	MTFR1	RAB32	ICMT	ADCY7	ME1
<i>HS3ST2</i>	SYN1	MS4A5	IL1RAPL2	CETN1	GJB4	GPR52	FSCN3	OR10H3	scube3	LILRB5
<i>HS6ST1</i>	ERN1	ALX4	GRIN1	AKT2	ARR3	TRIM46	TMEM143	OBSCN	ART1	CACNA1E
<i>NDST1</i>	MMP14	SPTBN4	EPN1	MYL10	QTRT1	SLC12A9	ALG12	arvcf	TP53I11	CDC42EP1
<i>NDST2</i>	HTR1B	RCVRN	CRYGA	SNX29	EMILIN2	PCDHB1	emx1	OR2H1	TRHR	CHRM5
<i>NDST3</i>	WRNIP1	KIAA0408	CNKSRR2	TRHR	GRIN2B	DCC	TMEM19	ALX4	ATP8B3	ZNF549
<i>NDST4</i>	ACKR2	ADAM18	PSG11	NTNG1	HCN4	DCC	GPR50	TACR3	RFX3	TRPC3
<i>SDC2</i>	KCTD20	ARSB	FPR3	EFCAB1	NEU3	INHBC	CPED1	C5RNP3	ADCYAP1	RBBP9
<i>SDC3</i>	SYN3	GFRA4	GRIN2B	CHD2	KCNJ5	CETN1	masp1	DCLK2	GPR173	LINC00894
<i>SDC4</i>	OSBPL2	ATP6V1E1	FGFR2	CALM3	CD9	ACAN	DSTN	DFFA	NIT1	PAPD7
<i>SULF1</i>	THBS2	CDH13	KIAA1462	ACVRL1	GPRC5D	NOX4	IFNAR1	CHKB	SPIN2A	CCL11
<i>UST</i>	DYM	PLSCR2	IQCH	zbtb3	FLRT1	GFM1	ZNF701	F2RL3	MYO7B	DCAF4

revealed glycosaminoglycan biosynthesis of HS and CS as the top two enriched terms (Figure 3A, c and d).

3.4 | Inferring drugs that deregulate HS and CS imputed and validated genes

The expression profile of HS and CS co-expressed genes were extracted from LFCs of all four cancer cells. The heatmap was used to identify the drugs that cause up/down-regulation of these genes. On the left and right side of each plot, the effect of drugs on down and

upregulations of most introduced genes is presented (Figure S1B.a). Figure S1B.b is an abstract of Figure S1B.a, in which ineffective drugs are omitted.

Finally, only the drugs which had the same role (in case of up or downregulation) in all four cancer cell lines were predicted as the most likely agents affecting the expression of HS and CS genes (Figure 3B).

Accordingly, 16 (Table S1) and 6 (Table 4) drugs were discovered, which might upregulate and downregulate the expression of HS and CS genes, respectively. Among the predicted drugs, only those which cause downregulation of HS and CS genes were proposed as repositionable drugs. The effect of these drugs on all introduced genes

TABLE 4 List of drugs predicted to cause down-regulation in HS and CS genes

Pert_iname	Canonical_smiles	Inchi_key	Direction
BMS-777607	<chem>CCOC1CCN(C2CCC(F)C2)C(=O)C1C(=O)Nc1ccc(Oc2ccc(N)c2Cl)c(F)c1</chem>	VNBRGSXVFBYQNN-UHFFFAOYSA-N	Down
GDC-0068	<chem>CC(C)NC[C@@H](C(=O)N1CCN(CC1)c1ncnc2[C@@H](O)C[C@@H](C)c12)c1ccc(Cl)cc1</chem>	GRZXWCHAXNAUHY-AHRSYUTCSA-N	Down
GDC-0068	<chem>C[C@@H]1C[C@H](C2=C1C(=NC=N2)N3CCN(CC3)C(=O)[C@H](CNC(C)C)C4=CC=C(C(=C4)C)O</chem>	GRZXWCHAXNAUHY-NSISKUIASA-N	Down
PF-04217903	<chem>OCCn1cc(zn1)-c1enc2nmn(Cc3ccc4ncccc4c3)c2n1</chem>	PDMUGYOXRHVNMU-UHFFFAOYSA-N	Down
SB-239063	<chem>COc1nccc(n1)c2c(ncn2C3CCC(O)CC3)c4ccc(F)cc4</chem>	ZQUSFAUAYSEREK-UHFFFAOYSA-N	Down
XMD-1499	<chem>COc1ccc1Nc1nc(N)c(s1)C(=O)c1c(Cl)cccc1Cl)N1CCN(C)CC1</chem>	U15RWLSAXIOCND-UHFFFAOYSA-N	Down

Abbreviations: HS, heparan sulphate; CS, chondroitin sulphate.

and only lab-validated HS and CS genes is illustrated in Figure 3B,a and 3B,b, respectively.

4 | DISCUSSION

De novo drug discovery is a costly process that entails lengthy trial periods and low success rates. Contrary to this approach, drug repositioning is a promising method for finding medications for newly-emerged and unknown diseases, such as COVID-19^[19] and orphan and common disorders.^[18] We comprehensively discussed the urgent need for drug repurposing previously.^[18] In addition, drug repurposing using retrieved data from the LINCS L1000 database has been performed previously.^[32] We developed a LINCS L1000-based pipeline to suggest novel functions for previously approved drugs or chemicals under investigation for safety validation. This pipeline aims to introduce chemicals that may complement ACPs to be studied in synergistic investigations against cancer. There are two advantages of following this procedure; first, proposed chemicals may improve the performance of ACPs by decreasing the level of HS and CS at the cell surface and second, diminishing the utilization cost of ACPs by reducing the required dose of consumption. The same efforts with promising outcomes have been done in vitro^[33,34] and in silico studies.^[35]

With regard to the effect of GAGs on performance of therapeutic peptides, Fadnes et al.^[17] clearly showed the negative effect of HS and CS on performance of bovine lactoferrin (LfcinB) and the designer peptide, KW5, as cationic therapeutic peptides. In agreement with this, in our previous study, we showed strong interaction of a cationic chimeric peptide (cLFchimera), derived from camel lactoferrin, with HS and CS by molecular dynamic simulation.^[16]

Our developed pipeline is designed based on HS and CS experimentally validated genes and their predicted gene-gene associations. By proposing in silico-driven co-expressed genes with HS and CS experimentally validated genes, this ambiguity may arise that how we can make sure that discovered relationships are meaningful. In other words, can we infer statistically proposed gene-gene correlations from biological evidence? To address this question, in this section we discuss the role of HS and CS genes in different malignancies in comparison with some of their predicted co-expressed genes. Besides, some

functions of co-expressed genes in protein level are also discussed. As we see in this part, similar to HS and CS genes, many suggested co-expressed genes have a crucial role in oncogenesis, which might be preliminary evidence that the list of gene associations is prepared meaningfully to pave the way for further studies to uncover new functions for known, and novel performances for unknown co-expressed genes. Concerning table 3, the first row includes HS and CS experimentally validated data while columns are likely co-expressed genes.

In terms of HS2T1 and its proposed co-expressed genes, the product of HS2T1, PPP1CA are involved in prostate cancer^[36] and oral squamous cell carcinoma^[37] respectively. PPP1CA is a member of the serine/threonine phosphatase complex and WDR82 is a part of its regulatory subunits.^[38] The STRING database^[39] as a protein-protein-interaction information source, has introduced PPP1CA and WDR82 and also DHX15 and LSM4 as co-expressed genes which interact at the protein level. As a sign to prove why Nup62 is co-expressed with HS2T1 in reality, HIF-1alpha is likely to play a role in progressive prostate cancer (similar to HS2T1) and it has been shown that HIF-1alpha is associated with nuclear pore complex Nup62 protein.^[40]

With regard to the HS6T1 and its proposed co-expressed genes, upregulation of HS6T1 in ovarian cancer cause increased tumour angiogenesis^[41] and mutated AKT2 regulates growth in thyroid carcinomas and colorectal cancer cells.^[42] Interestingly, the first proposed drug in our study, A-443654, is a potent small-molecule inhibitor of Akt serine/threonine kinases which induces Akt Ser-473 phosphorylation in all human cancers and has equal potency against AKT1, AKT2, or AKT3 within cells.^[43]

Upregulation of NDST1 and NDST2 leads to tumour progression in hepatocellular carcinoma,^[44] MMP14, anticipated co-expressed genes with NDST1, is introduced as a pro-metastatic gene product, which contributes to the acquisition of metastatic phenotypes in epithelial ovarian cancer cells.^[45] It is also presented as a prognostic marker in patients with sinonasal and oral malignant melanoma^[46] and prostate cancer.^[47] Besides, EMILIN2 as a likely co-expressed gene with NDST2 is a key extracellular regulator of the Wnt signalling pathway suppressing breast cancer cell growth and migration.^[48]

Downregulation of SULF1 affects suppressing tumour cell proliferation and invasion in malignant mesothelioma breast cancer, pancreatic, ovarian and head and neck cancers.^[49,50] Accordingly, it has been

shown that promoter methylation of *CDH13*, a proposed co-expressed gene with *SULF1*, plays a role in the endometrial carcinoma.^[51] Besides, *NOX4*, the other co-expressed gene with *SULF1*, transmits cell survival signals through the AKT-ASK1 pathway in pancreatic cancer cells.^[52]

Reduction of *SDC4* is correlated with neuroblastoma.^[53] Accordingly, it has been proved that *FGFR2* as a probable co-expressed gene with *SDC4* is associated with intrahepatic cholangiocarcinoma.^[54] Besides, the probable *SDC4* and *ACAN* protein-protein interaction have been proposed in the STRING database.^[39]

Elevation of *GPC1*, *GPC3*, and *GPC5* is associated with many cancers.^[55–57] *NEU3*, as a possible *GPC1*- and *SDC2* co-expressed gene, is a key regulator of the beta1 integrin-recycling pathway and FAK/AKT signalling and demonstrate its crucial role in renal cell carcinomas malignancy.^[58] *SAG* is known as a Kras-cooperating oncogene that promotes lung tumorigenesis^[59] and loss of *CRMP1*, a co-expressed gene with *GPC1*, escalates invasive phenotype of human Glioblastoma.^[60] Also, *AFP*, a *GPC3* co-expressed gene, can promote the proliferation of hepatoma cells via activation of PI3K/Akt signalling pathway^[61] and *A-443654*, the proposed drug in our study, by inducing Akt Ser-473 phosphorylation in all human cancer cell lines may affect this process.^[43] It should be mentioned that *GPC3*, *APOA2*, *SERPIND1*, *ITIH2*, *AHSG*, *AFP*, and *AMBP* form a protein-protein interaction module in STRING.^[39]

In terms of proposed chemicals as repositionable drugs, most of the proposed drugs could be categorized as different groups of inhibitors including protein kinase B (Akt) inhibitors (GDC-0068). It has been shown that *CGP-60474*^[32,62] and *WZ-3105*^[62] could be probable repurposed drugs against endotoxemic process^[32] and multiple cancers, including melanoma, adenocarcinoma, liver carcinoma, and breast, colon, and prostate cancers.^[62] Besides, the extensive therapeutic properties of celestrol in various diseases, including inflammatory diseases and cancer, neurodegenerative disorders and other diseases have been reviewed comprehensively,^[63] however, we only propose the chemicals which cause downregulation of HS and CS genes as repositionable drug candidates to enhance the performance of ACPs. For instance GDC-0068 which its effectiveness has been studied in multiple tumour models.^[63]

In conclusion, targeting HS and CS is a potential strategy to eliminate cancer.^[25,64] As a promising anticancer remedy, ACPs can be used in developing a treatment strategy against cancer due to their considerable advantages, including easy synthesising, high target specificity, selectivity and low toxicity.^[11] We proposed six drugs that diminish HS and CS gene expressions as one of the major obstacles on ACPs treatment. We believe that the synergistic effect of ACPs and these drugs could be a likely strategy (a) to investigate the performance of drugs on changing the HS and CS gene expression and (b) to monitor the complementary effect of ACPs and these chemicals on cancer treatment.

ACKNOWLEDGEMENTS

The authors acknowledge the entire staff of Systems Medicine group at the Science for Life Laboratory, KTH – Royal Institute of Technology.

They also acknowledge support from Science for Life Laboratory, and Swedish National Infrastructure for Computing (SNIC) at UPPMAX for providing assistance in computational infrastructure.

AUTHOR CONTRIBUTIONS


E. M.: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Visualization; Writing-original draft; Writing-review & editing. M. T.: Supervision; Validation. Rui Benfeitas: Conceptualization; Formal analysis; Methodology; Supervision. Ozlem Altay: Conceptualization; Methodology; Supervision. A. J.: Supervision; Validation. Simon Lam: Validation; Visualization; Writing-review & editing. A. M.: Conceptualization; Formal analysis; Supervision. Md. H. S.: Conceptualization; Supervision.

DATA AVAILABILITY STATEMENT

The codes for drug repositioning pipeline in R and Python programming languages are publicly available in https://github.com/ElyasMo/ACPs_HS_HSPGs_CS

ORCID

Ali Javadmanesh  <https://orcid.org/0000-0001-6016-5905>

Mohammad Hadi Sekhavati  <https://orcid.org/0000-0001-8091-3798>

REFERENCES

1. Marqus, S., Pirogova, E., & Piva, T. J. (2017). Evaluation of the use of therapeutic peptides for cancer treatment. *Journal of Biomedical Science*, 24, 21.
2. Burrell, R. A., McGranahan, N., Bartek, J., & Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, 501, 338–345.
3. Mahassni, S. H., & Al-Reemi, R. M. (2013). Apoptosis and necrosis of human breast cancer cells by an aqueous extract of garden cress (*Lepidium sativum*) seeds. *Saudi Journal of Biological Sciences* 20, 131–139.
4. Ponnusamy, L., Mahalingaiah, P. K. S., & Singh, K. P. (2016). Chronic oxidative stress increases resistance to doxorubicin-induced cytotoxicity in renal carcinoma cells potentially through epigenetic mechanism. *Molecular Pharmacology*, 89, 27–41.
5. Joshi, G., Sultana, R., Tangpong, J., Cole, M. P., St Clair, D. K., Vore, M., Estus, S., & Butterfield, D. A. (2005). Free radical mediated oxidative stress and toxic side effects in brain induced by the anti cancer drug adriamycin: insight into chemobrain. *Free Radical Research*, 39, 1147–1154.
6. Zhou, S., Palmeira, C. M., & Wallace, K. B. (2001). Doxorubicin-induced persistent oxidative stress to cardiac myocytes. *Toxicology Letters*, 121, 151–157.
7. Rivera, E., & Gomez, H. (2010). *Breast Cancer Research*, Springer p. 52.
8. Tyagi, A., Tuknait, A., Anand, P., Gupta, S., Sharma, M., Mathur, D., Joshi, A., Singh, S., Gautam, A., & Raghava, G. P. S. (2015). CancerPPD: a database of anticancer peptides and proteins. *Nucleic Acids Research*, 43, D837–D843.
9. Chiangjong, W., Chutipongtanate, S., & Hongeng, S. (2020). Anticancer peptide: Physicochemical property, functional aspect and trend in clinical application (review). *International Journal of Oncology*, 57, 678–696.
10. Hoskin, D. W., & Ramamoorthy, A. (2008). Studies on anticancer activities of antimicrobial peptides. *Biochimica et Biophysica Acta (BBA)- Biomembranes*, 1778, 357–375.
11. Simons, K., & Ikonen, E. (2000). How cells handle cholesterol. *Science*, 290, 1721–1726.

12. Zwaal, R. F., & Schroit, A. J. (1997). Pathophysiologic implications of membrane phospholipid asymmetry in blood cells. *Blood, The Journal of the American Society of Hematology*, 89, 1121–1132.
13. Harris, F., Dennison, S. R., Singh, J., & Phoenix, D. A. (2013). On the selectivity and efficacy of defense peptides with respect to cancer cells. *Medicinal Research Reviews*, 33, 190–234.
14. Schweizer, F. (2009). Cationic amphiphilic peptides with cancer-selective toxicity. *European Journal of Pharmacology*, 625, 190–194.
15. Li, G., Huang, Y., Feng, Q., & Chen, Y. (2014). Tryptophan as a probe to study the anticancer mechanism of action and specificity of α -helical anticancer peptides. *Molecules*, 19, 12224–12241.
16. Tanhaei, A., Jaafari, M. R., Ahmadi, F. S., Vakili-Ghartavol, R., & Sekhnavati, M. H. (2019). Secretory expression of a chimeric peptide in *Lactococcus lactis*: Assessment of its cytotoxic activity and a deep view on its interaction with cell-surface glycosaminoglycans by molecular modeling. *Probiotics and Antimicrobial Proteins*, 11, 1034–1041.
17. Fadnes, B., Rekdal, Ø., & Uhlin-Hansen, L. (2009). The anticancer activity of lytic peptides is inhibited by heparan sulfate on the surface of the tumor cells. *BMC Cancer*, 9, 183.
18. Mohammadi, E., Benfeitas, R., Turkez, H., Boren, J., Nielsen, J., Uhlen, M., & Mardinoglu, A. (2020). Applications of genome-wide screening and systems biology approaches in drug repositioning. *Cancers*, 12, 2694.
19. Altay, O., Mohammadi, E., Lam, S., Turkez, H., Boren, J., Nielsen, J., Uhlen, M., & Mardinoglu, A. (2020). Current status of COVID-19 therapies and drug repositioning applications. *IScience*, 23, 101303.
20. Kort, E. J., & Jovinge, S. (2019). Streamlined analysis of LINCS L1000 data with the slinky package for R. *Bioinformatics*, 35, 3176–3177.
21. Jones, E., Oliphant, T., & Peterson, P. (2001). SciPy: Open source scientific tools for Python.
22. Kumari, S., Nie, J., Chen, H. - S., Ma, H., Stewart, R., Li, X., Lu, M. - Z., Taylor, W. M., & Wei, H. (2012). Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. *PLoS One*, 7, e50411.
23. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*, 16, 284–287.
24. Fernández-Vega, I., García, O., Crespo, A., Castañón, S., Menéndez, P., Astudillo, A., & Quirós, L. M. Specific genes involved in synthesis and editing of heparan sulfate proteoglycans show altered expression patterns in breast cancer. *BMC Cancer*, 2013, 13, 24.
25. Nagarajan, A., Malvi, P., & Wajapeyee, N. (2018). Heparan sulfate and heparan sulfate proteoglycans in cancer initiation and progression. *Frontiers in Endocrinology*, 9, 9.
26. Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., Gould, J., Davis, J. F., Tubelli, A. A., Asiedu, J. K., Lahr, D. L., Hirschman, J. E., Liu, Z., Donahue, M., Julian, B., Khan, M., Wadden, D., Smith, I. C., Lam, D., ... Golub, T. R. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*, 171, 1437–1452.e17. e1417.
27. Duan, Q., Flynn, C., Niepel, M., Hafner, M., Muhlich, J. L., Fernandez, N. F., Rouillard, A. D., Tan, C. M., Chen, E. Y., Golub, T. R., Sorger, P. K., Subramanian, A., & Ma'ayan, A. (2014). LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. *Nucleic Acids Research*, 42, W449–W460.
28. Hendrickson, N., Eitel, A., Jujjavarapu, M., Guthrie, J., Evans, H., & Heyl-Clegg, D. (2017). Anticancer properties of the antimicrobial peptide CDT on A549 lung cancer cells. *The FASEB Journal*, 31, 610.614–610.614.
29. Xie, M., Liu, D., & Yang, Y. (2020). Anti-cancer peptides: classification, mechanism of action, reconstruction and modification. *Open Biology*, 10, 200004.
30. McCormick, C., Duncan, G., Goutsos, K. T., & Tufaro, F. (2000). The putative tumor suppressors EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyzes the synthesis of heparan sulfate. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 668–673.
31. Sarrazin, S., Lamanna, W. C., Esko, J. D. (2011). Heparan sulfate proteoglycans. *Cold Spring Harbor Perspectives in Biology*, 3, a004952.
32. Han, H. - W., Hahn, S., Jeong, H. Y., Jee, J. - H., Nam, M. O., Kim, H. K., Lee, D. H., Lee, S. Y., Choi, D. K., Yu, J. H., Min, S. H., & Yoo, J. (2018). LINCS L1000 dataset-based repositioning of CGP-60474 as a highly potent anti-endotoxemic agent. *Scientific Reports*, 8, 1–9.
33. Zhao, J., Huang, Y., Liu, D., & Chen, Y. (2015). Two hits are better than one: Synergistic anticancer activity of α -helical peptides and doxorubicin/epirubicin. *Oncotarget*, 6, 1769–1778.
34. Mohapatra, S., Saha, A., Mondal, P., Jana, B., Ghosh, S., Biswas, A., & Ghosh, S. (2017). Synergistic anticancer effect of peptide-docetaxel nanoassembly targeted to tubulin: Toward development of dual warhead containing nanomedicine. *Advanced Healthcare Materials*, 6, 1600718.
35. Celebi, R., Bear Don't Walk, O., Movva, R., Alpsy, S., & Dumontier, M. (2019). In-silico prediction of synergistic anti-cancer drug combinations using multi-omics data. *Scientific Reports*, 9, 8949.
36. Ferguson, B. W., & Datta, S. (2011). Role of heparan sulfate 2-O-sulfotransferase in prostate cancer cell proliferation, invasion, and growth factor signaling. *Prostate Cancer*, 2011, 1–14.
37. Hsu, L. - C., Huang, X., Seasholtz, S., Potter, D. M., & Collin, S. M. (2006). Gene amplification and overexpression of protein phosphatase 1 α in oral squamous cell carcinoma cell lines. *Oncogene*, 25, 5517–5526.
38. Lee, J.-H., You, J., Dobrota, E., & Skalik, D. G. (2010). Identification and characterization of a novel human PP1 phosphatase complex. *Journal of Biological Chemistry*, 285, 24466–24476.
39. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47, D607–D613.
40. Bertozzi, D., Iurlaro, R., Sordet, O., Marinello, J., Zaffaroni, N., & Capranico, G. (2011). Characterization of novel antisense HIF-1 α transcripts in human cancers. *Cell Cycle*, 10, 3189–3197.
41. Cole, C. L., Rushton, G., Jayson, G. C., & Avizienyte, E. (2014). Ovarian cancer cell heparan sulfate 6-O-sulfotransferases regulate an angiogenic program induced by heparin-binding epidermal growth factor (EGF)-like growth factor/EGF receptor signaling. *Journal of Biological Chemistry*, 289, 10488–10501.
42. Ericson, K., Gan, C., Cheong, I., Rago, C., Samuels, Y., Velculescu, V. E., Kinzler, K. W., Huso, D. L., Vogelstein, B., & Papadopoulos, N. (2010). Genetic inactivation of AKT1, AKT2, and PDKP1 in human colorectal cancer cells clarifies their roles in tumor growth regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 2598–2603.
43. Han, E. K.-H., Levenson, J. D., Mcgonigal, T., Shah, O. J., Woods, K. W., Hunter, T., Giranda, V. L., & Luo, Y. (2007). Akt inhibitor A-443654 induces rapid Akt Ser-473 phosphorylation independent of mTORC1 inhibition. *Oncogene*, 26, 5655–5661.
44. Tátrai, P., Egedi, K., Somorácz, Á., Van Kuppevelt, T. H., Ten Dam, G., Lyon, M., Deakin, J. A., Kiss, A., Schaff, Z., & Kovalszky, I. (2010). Quantitative and qualitative alterations of heparan sulfate in fibrogenic liver diseases and hepatocellular cancer. *Journal of Histochemistry and Cytochemistry*, 58, 429–441.
45. Nakayama, I., Shibazaki, M., Yashima-Abo, A., Miura, F., Sugiyama, T., Masuda, T., & Maesawa, C. (2013). Loss of HOXD10 expression induced by upregulation of miR-10b accelerates the migration and invasion activities of ovarian cancer cells. *International Journal of Oncology* 43, 63–71.
46. Kondratiev, S., Gnep, D. R., Yakirevich, E., Sabo, E., Annino, D. J., Rebeiz, E., & Laver, N. V. (2008). Expression and prognostic role of

- MMP2, MMP9, MMP13, and MMP14 matrix metalloproteinases in sinonasal and oral malignant melanomas. *Human Pathology*, 39, 337–343.
47. Trudel, D., Fradet, Y., Meyer, F., Harel, F., & Têtu, B. (2008). Membrane-type-1 matrix metalloproteinase, matrix metalloproteinase 2, and tissue inhibitor of matrix proteinase 2 in prostate cancer: identification of patients with poor prognosis by immunohistochemistry. *Human Pathology*, 39, 731–739.
 48. Marastoni, S., Andreuzzi, E., Paulitti, A., Colladel, R., Pellicani, R., Todaro, F., Schiavinato, A., Bonaldo, P., Colombatti, A., & Mongiat, M. (2014). EMILIN2 down-modulates the Wnt signalling pathway and suppresses breast cancer cell growth and migration. *The Journal of Pathology*, 232, 391–404.
 49. Lai, J.-P., Sandhu, D. S., Shire, A. M., & Roberts, L. R. (2008). The tumor suppressor function of human sulfatase 1 (SULF1) in carcinogenesis. *Journal of Gastrointestinal Cancer*, 39, 149–158.
 50. Heidari-Hamedani, G., Vivès, R. R., Seffouh, A., Afratis, N. A., Oosterhof, A., Van Kuppevelt, T. H., Karamanos, N. K., Metintas, M., Hjerpe, A., Dobra, K., & Szatmári, T. (2015). Syndecan-1 alters heparan sulfate composition and signaling pathways in malignant mesothelioma. *Cellular Signalling*, 27, 2054–2067.
 51. Sheng, Y., Wang, H., Liu, D., Zhang, C., Deng, Y., Yang, F., Zhang, T., & Zhang, C. (2016). Methylation of tumor suppressor gene CDH13 and SHP1 promoters and their epigenetic regulation by the UHRF1/PRMT5 complex in endometrial carcinoma. *Gynecologic Oncology*, 140, 145–151.
 52. Mochizuki, T., Furuta, S., Mitsushita, J., Shang, W. H., Ito, M., Yokoo, Y., Yamaura, M., Ishizone, S., Nakayama, J., Konagai, A., Hirose, K., Kiyosawa, K., & Kamata, T. (2006). Inhibition of NADPH oxidase 4 activates apoptosis via the AKT/apoptosis signal-regulating kinase 1 pathway in pancreatic cancer PANC-1 cells. *Oncogene*, 25, 3699–3707.
 53. Knelson, E. H., Gaviglio, A. L., Nee, J. C., Starr, M. D., Nixon, A. B., Marcus, S. G., & Blobe, G. C. (2014). Stromal heparan sulfate differentiates neuroblasts to suppress neuroblastoma growth. *Journal of Clinical Investigation*, 124, 3016–3031.
 54. Sia, D., Losic, B., Moicini, A., Cabellos, L., Hao, K., Reville, K., Bonal, D., Miltiadous, O., Zhang, Z., Hoshida, Y., Cornella, H., Castillo-Martin, M., Pinyol, R., Kasai, Y., Roayaie, S., Thung, S. N., Fuster, J., Schwartz, M. E., Waxman, S., ... Llovet, J. M. (2015). Massive parallel sequencing uncovers actionable FGFR2-PPHLN1 fusion and ARAF mutations in intrahepatic cholangiocarcinoma. *Nature Communications*, 6, 1–11.
 55. Huang, G., Ge, G., Izzi, V., Greenspan, D. S. (2017). $\alpha 3$ Chains of type V collagen regulate breast tumour growth via glypican-1. *Nature Communications*, 8, 1–17.
 56. Tanaka, M., Ishikawa, S., Ushiku, T., Morikawa, T., Isagawa, T., Yamagishi, M., Yamamoto, H., Katoh, H., Takeshita, K., Arita, J., Sakamoto, Y., Hasegawa, K., Kokudo, N., & Fukuyama, M. (2017). EVI1 modulates oncogenic role of GPC1 in pancreatic carcinogenesis. *Oncotarget*, 8, 99552.
 57. Feng, M., Gao, W., Wang, R., Chen, W., Man, Y.-G., Figg, W. D., Wang, X. W., Dimitrov, D. S., & Ho, M. (2013). Therapeutically targeting glypican-3 via a conformation-specific single-domain antibody in hepatocellular carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, 110, E1083–E1091.
 58. Tringali, C., Lupo, B., Silvestri, I., Papini, N., Anastasia, L., Tettamanti, G., & Venerando, B. (2012). The plasma membrane sialidase NEU3 regulates the malignancy of renal carcinoma cells by controlling $\beta 1$ integrin internalization and recycling. *Journal of Biological Chemistry*, 287, 42835–42845.
 59. Li, H., Tan, M., Jia, L., Wei, D., Zhao, Y., Chen, G., Xu, J., Zhao, L., Thomas, D., Beer, D. G., & Sun, Y. (2014). Inactivation of SAG/RBX2 E3 ubiquitin ligase suppresses Kras G12D-driven lung tumorigenesis. *Journal of Clinical Investigation*, 124, 835–846.
 60. Mukherjee, J., Desouza, L. V., Micallef, J., Karim, Z., Croul, S., Siu, K. W. M., & Guha, A. (2014). Loss of collapsin response mediator Protein1, as detected by iTRAQ analysis, promotes invasion of human gliomas expressing mutant EGFRvIII. *Cancer Research*, 69, 8545–8554.
 61. Zheng, L., Gong, W., Liang, P., Huang, X., You, N., Han, K. Q., Li, Y. M., & Li, J. (2014). Effects of AFP-activated PI3K/Akt signaling pathway on cell proliferation of liver cancer. *Tumor Biology*, 35, 4095–4099.
 62. Taguchi, Y.-H. (2019). Drug candidate identification based on gene expression of treated cells using tensor decomposition-based unsupervised feature extraction for large-scale data. *BMC Bioinformatics*, 19, 27–42.
 63. Cascão, R., Fonseca, J. E., & Moita, L. F. (2017). Celastrol: A spectrum of treatment opportunities in chronic diseases. *Frontiers in Medicine*, 4, 69.
 64. Hammond, E., Khurana, A., Shridhar, V., & Dredge, K. (2014). The role of heparanase and sulfatases in the modification of heparan sulfate proteoglycans within the tumor microenvironment and opportunities for novel cancer therapeutics. *Frontiers in Oncology*, 4, 195.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Mohammadi, E., Tahmoorespur, M., Benfeitas, R., Altay, O., Javadmanesh, A., Lam, S., Mardinoglu, A., & Sekhavati, M. H. (2021). Improvement of the performance of anticancer peptides using a drug repositioning pipeline. *Biotechnol. J.*, e2100417.
<https://doi.org/10.1002/biot.202100417>

Stockholm, on 21th of August 2023

Dr. Ozlem Altay
(professional title, name and surname)

STATEMENT

As the co-author of the work entitled **Improvement of the performance of anticancer peptides using a drug repositioning pipeline** I declare hereby that my own substantive contribution to the preparation, implementation and analysis of this study and presentation of experimental work in the form of above mentioned publication is **10% and comprise conceptualization and methodology of the analysis.**

At the same time, I consent to the submission of the above-mentioned work by **Elyas Mohhamadi** as part of a doctoral dissertation in the form of a thematically coherent collection of articles published in scientific journals.

I hereby declare that an independent and separable part of the above work shows the individual contribution of **Elyas Mohhamadi** when developing the concept, performing the experimental part, developing and interpretation of the results of this work.


.....
(co-author signature)

Stockholm, on 21th of August 2023

Dr. Rui Benfeitas

(professional title, name and surname)

STATEMENT

As the co-author of the work entitled **Improvement of the performance of anticancer peptides using a drug repositioning pipeline** I declare hereby that my own substantive contribution to the preparation, implementation and analysis of this study and presentation of experimental work in the form of above mentioned publication is **10% and comprise conceptualization and methodology of the analysis.**

At the same time, I consent to the submission of the above-mentioned work by **Elyas Mohhamadi** as part of a doctoral dissertation in the form of a thematically coherent collection of articles published in scientific journals.

I hereby declare that an independent and separable part of the above work shows the individual contribution of **Elyas Mohhamadi** when developing the concept, performing the experimental part, developing and interpretation of the results of this work.



.....
(co-author signature)

Mashhad, on 26th of August 2023


Dr. Mohammadi Hadi Sekhavati
(professional title, name and surname)

STATEMENT

As the co-author of the work entitled **Improvement of the performance of anticancer peptides using a drug repositioning pipeline** I declare hereby that my own substantive contribution to the preparation, implementation and analysis of this study and presentation of experimental work in the form of above mentioned publication is **15% and comprise conceptualization and methodology of the analysis and supervision.**

At the same time, I consent to the submission of the above-mentioned work by **Elyas Mohhamadi** as part of a doctoral dissertation in the form of a thematically coherent collection of articles published in scientific journals.

I hereby declare that an independent and separable part of the above work shows the individual contribution of **Elyas Mohhamadi** when developing the concept, performing the experimental part, developing and interpretation of the results of this work.

.....

(co-author signature)

Cambridge, on 21th of August 2023

Dr. Simon Lam
(professional title, name and surname)

STATEMENT

As the co-author of the work entitled **Improvement of the performance of anticancer peptides using a drug repositioning pipeline** I declare hereby that my own substantive contribution to the preparation, implementation and analysis of this study and presentation of experimental work in the form of above mentioned publication is **10% and comprise conceptualization and revision.**

At the same time, I consent to the submission of the above-mentioned work by **Elyas Mohhamadi** as part of a doctoral dissertation in the form of a thematically coherent collection of articles published in scientific journals.

I hereby declare that an independent and separable part of the above work shows the individual contribution of **Elyas Mohhamadi** when developing the concept, performing the experimental part, developing and interpretation of the results of this work.


.....
(co-author signature)