

**Paweł Segit** 

## Analiza profili transkryptomicznych pojedynczych komórek mieloidalnych w mikrośrodowisku glejaków

Praca doktorska wykonana w Pracowni Neurobiologii Molekularnej Instytutu Biologii Doświadczalnej im. M. Nenckiego Polskiej Akademii Nauk

PROMOTORZY: Prof. dr hab. Bożena Kamińska-Kaczmarek Dr hab. Jakub Mieczkowski



NARODOWE CENTRUM NAUKI



Warszawa, 2024

#### Podziękowania

Chciałbym podziękować:

Prof. dr hab. Bożenie Kamińskiej-Kaczmarek – za umożliwienie mi realizacji projektu doktorskiego w Pracowni Neurobiologii Molekularnej i za wszelką udzieloną pomoc w trakcie pracy

Dr. hab. Jakubowi Mieczkowskiemu – za wprowadzenie w świat bioinformatyki, ciekawe dyskusje i wszelką udzieloną pomoc w trakcie pracy

Dr Natalii Ochockiej i dr. Kacprowi Walentynowiczowi – za wspólną pracę nad identyfikacją populacji komórek tworzących mikrośrodowisko glejaków

Dr. Kamilowi Wojnickiemu, dr. Salwadorowi Cyranowskiemu, dr. Julianowi Swatlerowi i mgr. inż. Karolowi Jackowi – za współpracę przy realizacji projektów

Dr. hab. inż. Bartoszowi Wojtasiowi, dr Paulinie Szadkowskiej i dr Bartłomiejowi Gielniewskiemu – za sekwencjonowanie danych i dyskusje o NGS

Wszystkim Koleżankom i Kolegom z Pracowni Neurobiologii Molekularnej – za stworzenie dobrej atmosfery pracy i za wszystkie dyskusje o biologii

Rodzinie, Przyjaciołom, Znajomym – za wsparcie i cierpliwość

Badania, których wyniki zostały przedstawione w niniejszej rozprawie finansowane były przez: Narodowe Centrum Nauki: OPUS 14 2017/27/B/NZ3/01605 Fundację na rzecz Nauki Polskiej: TEAM TECH CORE FACILITY "Stworzenie opartej o sekwencjonowanie nowej generacji platformy do kompleksowej diagnostyki i spersonalizowanej terapii w neuroonkologii"

#### SPIS TREŚCI

1	STRI	STRESZCZENIE			
2	SUMMARY				
3	WS7	ΈΡ	10		
	3.1	Badania transkryptomu pojedynczych komórek – aktualne metody i wyzwania w			
	pozyskiwaniu i analizie danych				
	3.2	Metody o niskiej wydajności (ang. low throughput)	11		
	3.3	Metody wysokoprzepustowe (ang. high throughput) pomiaru RNA i białek w pojedynczej			
	komórce				
	3.3.1	Metody analizy ekspresji genów w pojedynczych komórkach (single-cell RNA-seq)	11		
	3.3.2	Cytometria przepływowa	13		
	3.3.3	Cytomeria masowa	15		
	5.5.4	Strategie analizy i wizualizacja wielowymiarowych uanych	10		
	3.4	Glejaki i ich mikrośrodowisko	18		
	3.4.1	Glejaki	18		
	3.4.2	Glejaki z punktu widzenia wysokoprzepustowych badań pojedynczych komórek	19		
	3.4.3	Rola układu odpornościowego w glejakach	21		
	3.	4.3.1 Układ odpornościowy w mózgu	21		
	3.	4.3.2 Komorki układ odpornosciowego w mikrosrodowisku glejaka	22		
	3. br	4.3.3 Komorki układu odpornościowego w mikrosrodowisku giejaka w wysokoprzepustowych udaniach pojedwaczych komórek	22		
	50		25		
4	CEL	PRACY	27		
5	MET	ОДҮ	28		
	5.1	Konwersja plików BCL do formatu FASTQ	28		
	5.2 Mapowanie odczytów do transkryptomu referencyjnego i stworzenie macierzy zliczeń				
	odczyto	ów zmapowanych do poszczególnych genów w pojedynczych komórkach	29		
	5.3	Analiza jakości danych w macierzy ekspresji	30		
	5.4	Normalizacja danych	32		
	5.5	Integracja wielu zbiorów danych	32		
	5.6	Redukcja wymiarów	34		
	5.7	Grupowanie komórek	35		
	5.8	Analiza różnicowej ekspresji genów	37		
6	WYNIKI				
	6.1	Analiza profili transkryptomicznych pojedynczych komórek CD11b+ pochodzących z			
	mikrośrodowiska mysiego glejaka GL261				
	6.1.1 Przygotowanie i normalizacja danych uzyskanych z scRNAseq		39		
	6.1.2	Identyfikacja subpopulacji komórek wśród komórek CD11b+	42		
	6.1.3 Ocena wpływu rozdzielczości grupowania na liczbę grup komórek i przypisanie komórek do				
	poszo	zególnych subpopulacji	43		
	6.1.4	Analiza protili transkryptomicznych subpopulacji komórek mikrogleju, makrofagów i BAM	47		

	6.1.5 Analiza wpływu z komórkami z mózgów	operacji pozorowanej na profile transkryptomiczne komórek CD11b+ w por myszy naiwnych	ównaniu 52			
	<ul> <li>6.2 Analiza profili tra mikrośrodowiska ludzkie</li> <li>6.2.1 Przygotowanie</li> <li>6.2.2 Integracja zbior</li> <li>6.2.3 Ocena wpływu poszczególnych subpopu</li> <li>6.2.4 Identyfikacja w wielopostaciowego</li> </ul>	inskryptomicznych pojedynczych komórek CD11b+ pochodzących z go glejaka wielopostaciowego i normalizacja danych uzyskanych z scRNA-seq ów danych z niezależnych próbek GBM rozdzielczości grupowania na liczbę grup komórek i przypisanie komórek do Ilacji ybranych subpopulacji komórek CD11b+ w mikrośrodowisku glejaka	54 60 64			
7	7 DYSKUSJA		72			
	7.1 Identyfikacja pop	oulacji komórek CD11b+ w mikrośrodowisku mysiego glejaka GL261	72			
	7.2 Zróżnicowanie p wielopostaciowego	opulacji komórek CD11b+ w mikrośrodowisku ludzkiego glejaka	75			
	7.3 Ograniczenia ana 7.3.1 Ograniczenia te	alizy profili transkryptomicznych pojedynczych komórek scRNA-seq				
	7.3.1.1 Niska iloś 7.3.1.2 Błędy i ni 7.3.1.3 Efekt seri	ć mRNA z pojedynczych komórek erównomierność amplifikacji i/próbki ( <i>ang. batch effect</i> )				
	7.3.2Ograniczenia bi7.3.2.1Zróżnicov7.3.2.2Odczytyw7.3.2.3Zależności	ologiczne vane profile transkryptomiczne komórek będących częścią jednej populacji ranie krótkich fragmentów transkryptów i przestrzenne	80 80 81 81			
8	3 WNIOSKI		83			
9	Ə LISTA PUBLIKACJI Z (	JDZIAŁEM AUTORA ROZPRAWY				
1(	10 BIBLIOGRAFIA	BIBLIOGRAFIA				

#### 1 STRESZCZENIE

Mikrośrodowisko nowotworu to heterogenna struktura, w której komórki tego samego typu, ale znajdujące się w innym obszarze guza, mogą mieć różny fenotyp i pełnić odmienne funkcje. Rozwój technik transkryptomiki i proteomiki pojedynczych komórek umożliwił badanie złożoności mikrośrodowiska w niedostępnej wcześniej rozdzielczości. Poznanie, jak różne komórki w guzie wspierają nowotwór pozwala lepiej zrozumieć mechanizmy jego progresji i racjonalnie projektować celowane techniki terapii.

Złośliwe glejaki to najczęściej występujące nowotwory ośrodkowego układu nerwowego, a najbardziej agresywnym wśród nich jest glejak wielopostaciowy. Mikrośrodowisko glejaków nie było szczegółowo zbadane pod kątem heterogenności rezydentnych i naciekających komórek układu odpornościowego ze względu na brak selektywnych markerów różnicujących te typy komórek. Wcześniejsze badania transkryptomiczne RNA-seq komórek sortowanych za pomocą przeciwciał wykazały, że profile transkryptomiczne jednocześnie wskazywały na fenotyp immunosupresyjny i typowy dla odpowiedzi zapalnej. Identyfikacja i charakterystyka subpopulacji komórek mieloidalnych w mikrośrodowisku wymagała zastosowania metody, która pozwoliłaby analizować pojedyncze komórki. Z tego powodu prowadzone badania rozszerzono i wykorzystano metodę transkryptomiki pojedynczych komórek (*ang. single-cell RNA-seq*, *scRNA-seq*).

Celem niniejszej rozprawy było opracowanie zoptymalizowanej metodyki analizy i wizualizacji danych scRNA-seq z komórek mieloidalnych w mikrośrodowisku doświadczalnych glejaków oraz w próbkach z guzów glejaka wielopostaciowego. Wykorzystując dostępne metody przeprowadzono optymalizację parametrów analizy danych scRNA-seq w glejakach GL261 w 14 dniu po implantacji. Porównano również profile transkryptomiczne komórek mieloidalnych w mózgach myszy naiwnych i po operacji pozorowanej. Ponadto przeanalizowano dane scRNA-seq z komórek mieloidalnych z mikrośrodowiska ludzkich glejaków, a także dokonano oceny wpływu wartości parametrów wykorzystywanych metod na uzyskiwane wyniki analiz i ich wizualizacje.

Wykazano, że uzasadnione było wykorzystanie zwierząt naiwnych w wykonanym doświadczeniu. Wykonano pierwszą opublikowaną analizę, w której połączono i bezpośrednio porównano dane scRNA-seq uzyskane z komórek CD11b+ z mikrośrodowiska mysiego

glejaka GL261 (a także z mózgów zwierząt naiwnych) pochodzące od samic i samców. Pozwoliło to wyodrębnić zróżnicowane subpopulacje komórek CD11b+, które wcześniej nie były możliwe od rozróżnienia w związku z brakiem znanych strategii sortowania. Ponadto wykonano analizę profili transkryptomicznych komórek mikrogleju i makrofagów z mikrośrodowiska glejaka wielopostaciowego, pochodzących z próbek od pięciu pacjentów. Zidentyfikowano geny o statystycznie istotnie podwyższonej ekspresji w poszczególnych podgrupach tych komórek, co pozwoliło na wyróżnienie ich podtypów funkcjonalnych. Uzyskane wyniki pozwalają na konstruowanie nowych hipotez badawczych weryfikujących funkcje zidentyfikowanych subpopulacji komórek mieloidalnych i ich powiązanie z właściwościami mikrośrodowiska złośliwych glejaków.

Uzyskane dane scRNA-seq, wykorzystane w opublikowanych analizach, zostały zdeponowane w publicznych repozytoriach i są dostępne dla środowiska naukowego do dalszych analiz.

#### 2 SUMMARY

The tumor microenvironment is a heterogenous structure where cells of the same type, but located in different areas of the tumor, can exhibit different phenotypes and perform distinct functions. Development of the single-cell transcriptomics and proteomics has enabled to study the complexity of the microenvironment at a previously inaccessible resolution. Understanding how different cells in the tumor support its development allows for a better grasp of the mechanisms behind tumor progression and the rational design of targeted therapeutic techniques.

Malignant gliomas are the most common tumors of the central nervous system and glioblastoma multiforme is the most aggressive among them. The microenvironment of gliomas has not been thoroughly investigated considering the heterogeneity of resident and infiltrating immune cells (due to the lack of selective markers that differentiate these cell types). Previous RNA-seq transcriptomic studies of antibody-sorted cells indicated that transcriptomic profiles simultaneously exhibited an immunosuppressive phenotype and a one typical for an inflammatory response. Identifying and characterizing subpopulations of myeloid cells in the microenvironment required a method capable of analyzing single cells. For this reason, ongoing research was expanded to utilize single-cell RNA-seq (scRNA-seq).

The aim of this dissertation was to develop an optimized methodology for analyzing and visualizing scRNA-seq data from myeloid cells in the microenvironment of experimental gliomas and from myeloid cells obtained from samples of glioblastoma multiforme tumors. Using established methods, parameters for scRNA-seq data analysis of the GL261 gliomas were optimized for data collected 14 days post-implantation. Furthermore, transcriptomic profiles of myeloid cells in the brains of naïve mice and those after sham surgery were compared. Additionally, scRNA-seq data from myeloid cells in human GBM microenvironment were analyzed, assessing the impact of parameter values used in methods on the obtained results and their visualizations.

It was shown that the use of naïve animals in this experiment was justified. This was the first published analysis, combining and directly comparing scRNA-seq data obtained from CD11b+ cells in the microenvironment of mouse GL261 gliomas (as well as from naïve mouse brains)

derived from both female and male mice. This allowed for the identification of multiple CD11b+ cell subpopulations that had previously been impossible to distinguish due to a lack of known sorting strategies. Additionally, transcriptomic profiles of microglia and macrophages from the microenvironment of glioblastoma multiforme were analyzed using samples collected intraoperatively from five patients. Genes with significantly upregulated expression in specific subgroups of these cells were identified, allowing for the characterization of their functional subtypes. The obtained results enable the construction of new research hypotheses verifying the functions of identified myeloid cell subpopulations and their connection to the properties of malignant glioma microenvironments.

The scRNA-seq data used in published analyses was deposited in public repositories and is available for further analysis by the scientific community.

#### 3 WSTĘP

### 3.1 Badania transkryptomu pojedynczych komórek – aktualne metody i wyzwania w pozyskiwaniu i analizie danych

Jednym ze sposobów poznania funkcji komórek w organizmach były od wielu lat badania poziomu ekspresji genów. Z obecnego w komórce mRNA mogą powstać białka, które pełnią określone funkcje i wpływają na fenotyp komórki. Istnieje szereg technik, takich jak hybrydyzacje do mikromacierzy DNA (Schena *et al.*, 1995) lub sekwencjonowanie RNA (RNA-seq) (Mortazavi *et al.*, 2008; Wang, Gerstein and Snyder, 2009) pozwalających ocenić ilościowo lub jakościowo zmiany ekspresji od pojedynczych do dziesiątek tysięcy genów. Wymienione metody pozwalają na badanie transkryptomu wszystkich komórek w próbce (*ang. bulk*), bez możliwości rozróżnienia profili transkryptomicznych w specyficznych podgrupach komórek występujących w próbce.

W 2009 roku Tang i in. w pracy "mRNA-Seq whole-transcriptome analysis of a single cell" (Tang *et al.*, 2009) zaprezentowali wyniki ilościowej analizy transkryptomu pojedynczych mysich blastomerów, a także przedstawili procedurę, która pozwoliła na uzyskanie odpowiedniej ilości mRNA, aby możliwe było skorzystanie z dostępnych w tamtym czasie metod sekwencjonowania następnej generacji (NGS). Podczas analizy otrzymanych danych zbadali ekspresję o 75% większej liczby genów w porównaniu do wyników otrzymywanych przy wykorzystaniu technik mikromacierzowych, ale przede wszystkim w przypadku części genów stwierdzili różnorodność izoform transkryptów występujących w pojedynczej komórce.

Od tego momentu nastąpił rozwój dziesiątek metod pozwalających na badania transkryptomiczne, genomiczne, a także epigenomiczne pojedynczych komórek wchodzących w skład populacji tysięcy, a nawet milionów komórek. Ze względu na liczbę komórek, które (realnie) można zbadać w pojedynczym doświadczeniu powstałe metody można podzielić na te o niskiej wydajności (*ang. low throughput*) i metody wysokoprzepustowe (*ang. high throughput*). W szczególności upowszechnienie się metod wysokoprzepustowych umożliwiających prowadzenie analiz w niedostępnej wcześniej rozdzielczości i skali doprowadziło do odkrycia nieznanych

subpopulacji komórek o istotnych rolach w procesach fizjologicznych i patologicznych (Villani *et al.*, 2017; Sade-Feldman *et al.*, 2018).

#### 3.2 Metody o niskiej wydajności (ang. *low throughput*)

W początkowej fazie rozwoju badań transkryptomiki pojedynczych komórek rozwijano metody o niskiej wydajności, pozwalające w jednym eksperymencie dokonać pomiarów (np. ekspresji genów) równolegle w kilkudziesięciu komórkach. Metody te opierają się na pipetowaniu pojedynczych komórek na płytki wielodołkowe. W poszczególnych dołkach płytki umieszczane są odczynniki pozwalające na przeprowadzenie określonej reakcji (np. hybrydyzacji cząsteczek mRNA do starterów zawierających olido(dT)). Następnie otrzymany materiał jest amplifikowany przy użyciu metody aplifikacji w czasie rzeczywistym - PCR (*ang. polymerase chain reaction*), aby po zakończeniu procesu uzyskać ilość odpowiednią do wykonania sekwencjonowania NGS. Do najpopularniejszych metod tego rodzaju do analizy transkryptomicznej należą Smart-seq2 (Picelli *et al.*, 2014) i CEL-seq2 (Hashimshony *et al.*, 2016). Niewątpliwa zaleta tych metod to wysoka czułość, a także możliwość odczytu długich fragmentów mRNA pozwalająca m.in. na badanie różnorodności izoform transkryptów występujących w jednej komórce. Natomiast głównym ich ograniczeniem jest w praktyce niewielka liczba komórek, które można w ten sposób zbadać.

### 3.3 Metody wysokoprzepustowe (*ang. high throughput*) pomiaru RNA i białek w pojedynczej komórce

#### 3.3.1 Metody analizy ekspresji genów w pojedynczych komórkach (single-cell RNA-seq)

W 2015 zaprezentowana została metoda Drop-seq (Macosko *et al.*, 2015), która pozwoliła w jednym eksperymencie zmierzyć ekspresję genów w tysiącach pojedynczych komórek. W przeciwieństwie do metod o niskiej wydajności nie opierała się ona na pipetowaniu komórek na płytki wielodołkowe, a na wykorzystaniu układu mikroprzepływowego do wyodrębnienia pojedynczych komórek z zawiesiny zawierającej dziesiątki tysięcy komórek.

Układy mikroprzepływowe (Qin *et al.*, 1998) to urządzenia laboratoryjne, w których przepływ cieczy kontrolowany jest z wysoką precyzją i odbywa się w mikrokanałach (o rozmiarach poniżej 1 mm). W mikroskali podczas przepływu cieczy siły powierzchniowe dominują nad siłami objętościowymi. Jednym z możliwych zastosowań takich układów jest podział cieczy na krople, zawieszone w niemieszającym się płynie (np. oleju). Takie krople mogą stać się odizolowanym środowiskiem przebiegu reakcji chemicznych, podobnym do pojedynczego dołka na płytce wielodołkowej. Właśnie do takiego celu autorzy metody Drop-seq wykorzystali układy mikroprzepływowe (Macosko *et al.*, 2015). Komórki, znajdujące się początkowo w wodnej zawiesinie, zostały przepompowane przez układ mikroprzepływowy, w którym ich przepływ był mieszany z przepływem oleju. W efekcie pojedyncze komórki zostały zamknięte w kroplach o nanolitrowej objętości.

Przeprowadzenie analizy ekspresji genów w pojedynczych komórkach wymagało, aby podczas sekwencjonowania była możliwość odczytania transkryptów znajdujących się w komórkach. Konieczne było aby w kropli, w której znalazła się pojedyncza komórka znalazły się również odczynniki mogące dołączyć się do obecnych w komórce transkryptów mRNA. W tym celu dodano trzeci strumień z kulkami żelowymi zawierającymi potrzebne odczynniki. Podstawowym wyzwaniem takiego podejścia było to, aby po zakończeniu reakcji w obrębie kropli zachować informację o tym, że dana grupa transkryptów pochodzi z tej samej kropli, która w założeniu powinna zawierać jedną komórkę. Autorzy zaproponowali, aby w ramach rozwiązania wykorzystać znaczniki (*ang. barcode*) dołączone do starterów oligo(dT) składające się z unikalnej sekwencji 20 nukleotydów. Znaczniki te miały dwie części: znacznik komórki o długości 12 nukleotydów. Zestaw syntezowany był w taki sposób, aby dla jednej mikrocząsteczki z ponad 10<sup>8</sup> starterami oligo(dT) znacznik komórki był identyczny dla każdego startera, a jednocześnie, aby każdy starter miał również dołączoną unikalną sekwencję 8 nukleotydów (UMI), tworząc razem unikalny 20 nukleotydowy znacznik (Macosko *et al.*, 2015).

W wyniku działania układu mikroprzepływowego w metodzie Drop-seq w pojedynczej kropli powinna się znaleźć jedna komórka i jedna mikrocząsteczka ze starterami oligo(dT) z dołączonymi znacznikami. Po zakończeniu procedury otrzymywana jest mieszanina transkryptów

mRNA pochodzących ze wszystkich komórek, które zostały zamknięte w kroplach, ale każdy z tych transkryptów ma dołączony unikalny znacznik. W kolejnym kroku przygotowywana jest biblioteka do sekwencjonowania (RNA-seq) i wykonywane sekwencjonowanie następnej generacji.

W wyniku sekwencjonowania otrzymywane są dane zawierające informację o sekwencji odczytanego transkryptu i znaczniku, który był dołączony do startera oligo(dT). Na etapie analizy bioinformatycznej, wykorzystując informację zawartą w znacznikach, można odczyty podzielić na grupy pochodzące z pojedynczych komórek (posiadające ten sam znacznik komórki), a dzięki UMI dodatkowo usunąć zduplikowane odczyty będące efektem amplifikacji materiału metodą PCR przed sekwencjonowaniem.

Metody wysokoprzepustowe wykorzystujące układy mikroprzepływowe pozwalają aktualnie na pomiar ekspresji genów w dziesiątkach, a nawet setkach tysięcy komórek w ramach pojedynczego eksperymentu. Daje to możliwość przeprowadzania analiz heterogennych populacji komórek w nieosiągalnej wcześniej rozdzielczości, badając typy obecnych w nich subpopulacji, a także fenotyp poszczególnych subpopulacji.

W rozpowszechnieniu wysokoprzepustowych metod analizy pojedynczych komórek duży udział ma firma 10x Genomics, która stworzyła komercyjny system (obejmujący kontroler mikroprzepływowy i zestawy gotowych odczynników przeznaczonych do różnego rodzaju eksperymentów) oferowany na całym świecie, dobrze udokumentowany i znacznie skracający czas wdrożenia metody w laboratorium (https://www.10xgenomics.com/products/single-cellgene-expression). Obecnie na rynku oferowanych jest jeszcze kilka alternatywnych systemów (z których część ma zdecydowanie węższy, wyspecjalizowany zakres działania), jednak aktualnie to rozwiązania stworzone bądź rozwinięte przez 10x Genomics zdominowały dziedzinę badań pojedynczych komórek.

#### 3.3.2 Cytometria przepływowa

Cytometria przepływowa to technika wykorzystywana intensywnie w immunologii, biologii nowotworów, mikrobiologii i monitorowaniu chorób zakaźnych, do charakterystyki heterogennej populacji komórek na podstawie analizy wielu cech (tj. rozmiar, kształt lub intensywność fluorescencji) mierzonych na poziomie pojedynczej komórki. Przepływając przez cytometr pojedyncze komórki oświetlane są spolaryzowanymi wiązkami światła pochodzącymi z wbudowanych laserów, a następnie dokonywany jest pomiar rozproszenia wiązki światła i poziomu fluorescencji. Każda komórka jest analizowana pod kątem rozproszenia światła widzialnego i jednego lub wielu parametrów fluorescencji. Rozproszenie światła widzialnego jest mierzone w dwóch różnych kierunkach, w kierunku do przodu (*ang. Forward Scatter lub FSC*), co wskazuje względny rozmiar komórki oraz pod kątem 90° (*ang. Side Scatter lub SSC*), który wskazuje na wewnętrzną złożoność lub ziarnistość komórki. Rozpraszanie światła jest niezależne od fluorescencji (McKinnon, 2018). Próbki są przygotowywane do pomiaru fluorescencji poprzez transfekcję i ekspresję białek fluorescencyjnych (np. Green Fluorescent Protein, GFP), barwienie barwnikami fluorescencyjnymi (np. jodek propidyny) lub barwienie fluorescencyjnie sprzężonymi przeciwciałami (np. CD3 z fluoresceiną).

Wykorzystując przeciwciała skoniugowane z barwnikami fluorescencyjnymi (fluorochromami) można oznaczyć wybrane epitopy i zmierzyć poziom fluorescencji dołączonych barwników (McKinnon, 2018). Po zakończeniu procedury, dla każdej komórki przepływającej przez cytometr (a właściwie dla każdego wykrytego przepływającego obiektu (event), mogącego być też jedynie zanieczyszczeniem) otrzymywany jest zestaw wartości mierzonych parametrów. Na tej podstawie możliwe jest następnie przeprowadzenie analizy wielowymiarowej i scharakteryzowanie badanej populacji komórek (m.in. statystycznie oceniając poziom ekspresji białek powierzchniowych, z którymi powinny były związać się wykorzystane przeciwciała). Otrzymane wyniki mogą posłużyć do wyodrębnienia subpopulacji komórek o zbliżonych cechach.

Jednym z większych wyzwań cytometrii przepływowej jest nakładanie się widm emisji poszczególnych fluorochromów, a w efekcie problem z wiarygodną oceną poziomu badanych sygnałów. W celu ograniczenia tego zjawiska stosowane są metody kompensacji (eliminacji detekcji pokrywających się zakresów widm), jednak mimo to, w tradycyjnych cytometrach w trakcie pojedynczego eksperymentu możliwy jest jednoczesny pomiar poziomu fluorescencji jedynie kilku do kilkunastu barwników (Barteneva, Fasler-Kan and Vorobjev, 2012).

Cytometria przepływowa umożliwia również podział (sortowanie) zawiesiny komórek na subpopulacje w oparciu o wartości mierzonych parametrów. W ten sposób z heterogennej zawiesiny komórek (uzyskanej np. po dysocjacji pobranej tkanki) można wyodrębnić wybraną subpopulację (np. komórki charakteryzujące się ekspresją wybranego białka powierzchniowego) i następnie przeprowadzić na niej dodatkowe eksperymenty (Cossarizza *et al.*, 2019).

#### 3.3.3 Cytomeria masowa

Cytometria masowa (*ang. Cytometry by time of flight,* CyTOF) łączy w sobie cechy cytometrii przepływowej i spektrometrii masowej. W tej metodzie pomiaru cech pojedynczych komórek wybrane przeciwciała, zamiast barwników fluorescencyjnych, koniuguje się z izotopami metali (głównie z grupy lantanowców) (Brodie and Tosevski, 2017). W związku z tym, w cytometrach tego typu lasery i elementy optyczne zostały zastąpione elementami występującymi w spektrometrach mas. Każdy z wykorzystywanych metali ma ściśle określoną masę cząsteczkową (ich widma masowe się nie nakładają w przeciwieństwie do widm barwników fluorescencyjnych). Dzięki temu podczas jednego eksperymentu możliwy jest równoległy pomiar ponad 50 wybranych parametrów komórkowych.

Przed rozpoczęciem eksperymentu komórki znakowane są przeciwciałami skoniugowanymi z izotopami metali. Tak przygotowane przechodzą przez tzw. nebulizator, aby każda komórka znalazła się w pojedynczej kropli roztworu. Następnie, wykorzystując plazmę, pojedyncze krople zostają odparowane, zatomizowane, a jony zostają zjonizowane. W kolejnym kroku część jonów (o niskiej masie) jest usuwana, a pozostałe kierowane są do analizatora masy TOF (ang. Time Of Flight). Na podstawie masy następuje rozdział jonów i przekierowanie ich do detektora. Po zakończeniu procedury dla każdej komórki otrzymywana jest ilość poszczególnych izotopów metali zmierzona przez detektor. Dane te, podobnie jak w przypadku cytometrii przepływowej, wykorzystywane są do przeprowadzenia analizy wielowymiarowej badanej populacji komórek. Z uwagi na możliwość jednoczesnego pomiaru ponad 50 parametrów analiza ta jest bardziej złożona, jednak z drugiej strony może pozwolić na lepsze scharakteryzowanie wyodrębnionych subpopulacji (Baharlou et al., 2019). Metoda ta została wykorzystana m.in. w badaniach charakteryzujących populacje odpornościowe w mózgu naiwnych myszy za pomocą 44 markerów powierzchniowych komórek. Porównując skład komórek odpornościowych i profile komórkowe w mózgu i krwi, scharakteryzowano wcześniej nieopisane podzbiory komórek CD8 T, limfocyty B, NK (natural killer, naturalni zabójcy) i komórki dendrytyczne (DCs) w mózgu. Za

pomocą cytometrii przepływowej pokazano zróżnicowaną dystrybucję populacji układu odpornościowego między oponami mózgowymi, splotem naczyniówkowym i parenchymą mózgu (Korin *et al.*, 2017; Korin, Dubovik and Rolls, 2018).

Pomimo niewątpliwych zalet cytometria masowa ma też istotną wadę - w przeciwieństwie do cytometrii przepływowej nie pozwala na podział (sortowanie) zawiesiny komórek na subpopulacje na podstawie wartości mierzonych parametrów. Wynika to wprost z przebiegu procedury, podczas której krople roztworu zwierające pojedyncze komórki są odparowywane (Bandura *et al.*, 2009; Groborz and Drąg, 2021).

#### 3.3.4 Strategie analizy i wizualizacja wielowymiarowych danych

Wykorzystanie metod analizy ekspresji genów w pojedynczych komórkach wiąże się z uzyskaniem wysoce wielowymiarowych danych (liczba wymiarów równa jest liczbie analizowanych genów i najczęściej wynosi od kilku do kilkudziesięciu tysięcy). Z tego powodu jednym z głównych kroków wstępnego przetwarzania danych jest redukcja wymiarów. Istnieje wiele metod redukcji wymiarów ogólnego przeznaczenia, z których w analizach ekspresji genów w pojedynczych komórkach najczęściej obecnie wykorzystuje się analizę składowych głównych (*ang. principal component analysis*, PCA). Zestaw wybranych składowych głównych może tworzyć dane wejściowe na potrzeby grupowania komórek lub ich wizualizacji. Algorytm t-SNE (*ang. t-Distributed Stochastic Neighbor Embedding*) i algorytm UMAP (*ang. Uniform Manifold Approximation and Projection*) wykorzystywane są do wizualizacji zbioru komórek w przestrzeni (najczęściej) dwuwymiarowej.

Analiza składowych głównych (PCA) (Jollife and Cadima, 2016) to metoda, w której oryginalny wielowymiarowy zbiór poddawany jest transformacji i uzyskiwany jest zbiór, w którym każdej obserwacji (w tym przypadku: każdej pojedynczej komórce) przypisywany jest zestaw zmiennych zwanych składowymi głównymi, w liczbie równej liczbie wymiarów (w tym przypadku: genów). Składowe główne są uszeregowane w taki sposób, że każda kolejna wyjaśnia nie więcej wariancji w oryginalnym zbiorze danych niż składowa poprzednia. Z tego względu w rzeczywistości w analizach ekspresji genów w pojedynczych komórkach wykorzystuje się najczęściej kilkanaście do kilkudziesięciu pierwszych składowych głównych jako wystarczającą

reprezentację oryginalnego zbioru. Wymiary oryginalnego zbioru danych wyrażane są jako kombinacje liniowe składowych głównych. Liczba składowych głównych wykorzystana w poszczególnych analizach przedstawionych w niniejszej pracy została umieszczona w sekcji Wyniki.

t-SNE (Van Der Maaten and Hinton, 2008) to nieliniowa metoda redukcji wymiarów najczęściej wykorzystywana do wizualizacji danych wielowymiarowych w przestrzeni dwu lub trójwymiarowej. Jest to rozwinięcie metody Stochastic Neighbor Embedding (Hinton and Roweis, 2002), w którym zastosowano zmienioną funkcję kosztu (z uproszczonym gradientem) i rozkład t-Studenta (zamiast rozkładu Gaussa) przy wyliczaniu podobieństwa pomiędzy punktami w niskowymiarowej przestrzeni. W t-SNE na początku dla każdego z punktów wykonywana jest identyfikacja najbliższych sąsiadów w przestrzeni wysokowymiarowej. Korzystając z rozkładu Gaussa i wyliczonych odległości między punktami określane jest później prawdopodobieństwo, że dane dwa punkty sąsiadują ze sobą. Przejście do przestrzeni niskowymiarowej rozpoczyna się od losowego rozmieszczenia punktów w przestrzeni dwu lub trójwymiarowej. Następnie rzutując punkty na jeden wymiar, korzystając z rozkładu t-Studenta i miary podobieństwa między punktami w przestrzeni jednowymiarowej określane jest prawdopodobieństwo, że punkty sąsiadują ze sobą. W ten sposób uzyskuje się dwa zestawy prawdopodobieństw: jeden w przestrzeni wysokowymiarowej i jeden w przestrzeni niskowymiarowej. W kolejnym kroku, iteracyjnie, modyfikuje się położenie punktów w przestrzeni niskowymiarowej, tak aby jak najlepiej odwzorować zbiór prawdopodobieństw z przestrzeni wysokowymiarowej w przestrzeni niskowymiarowej (minimalizując wartość dywergencji Kullbacka-Lieblera).

UMAP (Uniform Manifold Approximation and Projection) (McInnes *et al.*, 2018) to nieliniowa metoda redukcji wymiarów również najczęściej wykorzystywana do wizualizacji danych wielowymiarowych w przestrzeni dwu i trójwymiarowej. Algorytm został stworzony na bazie trzech założeń dotyczących analizowanych danych: dane mają rozkład jednorodny w rozmaitości Riemanna, metryka Riemanna jest lokalnie stała (lub może być przybliżona jako taka) i rozmaitość jest lokalnie spójna. W pierwszym kroku konstruowany jest graf w przestrzeni wysokowymiarowej (topologiczna reprezentacja danych), a następnie przechodząc do przestrzeni niskowymiarowej wykonywana jest optymalizacja topologicznej reprezentacji w tej przestrzeni, tak aby zminimalizować entropię krzyżową pomiędzy reprezentacjami (tą w przestrzeni wysokowymiarowej i niskowymiarowej).

Na takich dwuwymiarowych wykresach przedstawiane są na poziomie pojedynczych komórek m.in. wyniki grupowania, wartości ekspresji pojedynczych genów lub reprezentacji ekspresji zdefiniowanych zestawów genów. Wizualizacje są pomocne w lepszym zrozumieniu charakterystyki analizowanego zbioru komórek, jednak warto podkreślić, że przypisanie komórek do grup i identyfikacja genów o podwyższonej ekspresji w poszczególnych grupach powinny odbywać się przy wykorzystaniu przeznaczonych do tego celu metod statystycznych.

#### 3.4 Glejaki i ich mikrośrodowisko

#### 3.4.1 Glejaki

Glejaki to nowotwory ośrodkowego układu nerwowego (OUN), które powstają z neuralnych komórek macierzystych lub komórek progenitorowych astrocytów lub oligodendrocytów. Rozlane glejaki stanowią do 80% pierwotnych złośliwych nowotworów mózgu u osób dorosłych. W 2016 roku Światowa Organizacja Zdrowia (WHO) przedstawiła czwartą wersję klasyfikacji nowotworów OUN (Louis et al., 2016). Opracowanie to miało na celu stworzenie ogólnoświatowych wytycznych wykorzystywanych przy określaniu typu i stopnia złośliwości nowotworu, przede wszystkim w oparciu o cechy histopatologiczne. Glejaki podzielono na: łagodne (stopnia I - WHO grade I), o niskiej złośliwości (stopnia II - WHO grade II), rozlane złośliwe (stopnia III - WHO grade III), a także glejaki wielopostaciowe, wysoce złośliwe i inwazyjne (stopnia IV - WHO grade IV). Klasyfikacja ta została zaktualizowana w 2021 roku (Park et al., 2023), gdy zaproponowano nowy system oceny uwzględniający zarówno cechy histopatologiczne nowotworu jak i zidentyfikowane zmiany genetyczne. Glejaki podzielono na pięć typów, a klasyfikacja uwzględnia dodatkowo m.in. występowanie mutacji w genach IDH1/IDH2 (kodujących dehydrogenazę izocytrynianową 1 i 2) i obecność kodelecji 1p/19q, które okazały się pozytywnie wpływać na rokowanie pacjentów w przypadku części diagnozowanych nowotworów.

#### 3.4.2 Glejaki z punktu widzenia wysokoprzepustowych badań pojedynczych komórek

Badania transkryptomu pojedynczych komórek (scRNA-seg) glejaków stopnia IV (łac. glioblastoma, GBM) wykazały duże zróżnicowanie genetyczne tych nowotworów, a także występowanie subpopulacji komórek będących w określonym stanie funkcjonalnym (takim jak hipoksja lub odpowiedź immunologiczna). W badaniu za pomocą Smart-seg przeanalizowano 430 komórek z 5 GBM (z prawidłowym wt/HD1). Mieszanina komórek z guzów pierwotnych była pozbawiona komórek odpornościowych (CD45<sup>+</sup>). Analizy profili transkryptomicznych pozwoliły zaobserwować różnice w ekspresji genów o znanej i istotnej roli w procesach nowotworzenia, m.in. EGFR, PDGFRA, PDGFA, FGFR1, FGF1, NOTCH2 i JAG2 (Patel et al., 2014). Ponadto, ekspresja genów kodujących czynniki transkrypcyjne, których obecność jest charakterystyczna dla komórek macierzystych wykazała istnienie komórek z fenotypem komórek macierzystych glejaków (ang. glioma stem cells, GSCs). Wcześniejsze badania transkryptomiczne całych guzów (ang. bulk RNAseq) oparte na profilach ekspresji ujawniły trzy podtypy glejaka wielopostaciowego (GBM): proneuralny (TCGA-PN), klasyczny (TCGA-CL) i mezenchymalny (TCGA-MES) (Verhaak et al., 2010; Wang et al., 2017). Zidentyfikowano "sygnature genowa" (ang. stemness signature), czyli zestaw genów o wysokiej i specyficznej ekspresji, która była najsilniejsza w komórkach podtypu proneuralnego i klasycznego GBM, a słabiej reprezentowana w podtypie mezenchymalnym (Patel et al., 2014).

Zintegrowane analizy ekspresji genów w pojedynczych komórkach z 28 GBM, a także danych genomicznych i transkryptomicznych odczytanych z mieszaniny komórek 401 takich nowotworów wykazały, że komórki GBM (w obrębie jednego guza) występują w czterech stanach, przyjmując fenotyp zbliżony do: neuralnych komórek progenitorowych (*ang. neural progenitor-like*, NPC-like), komórek progenitorowych oligodendrocytów (ang. *oligodendrocyte-progenitor-like*, OPC-like), astrocytów (ang. *astrocyte-like*, AC-like), komórek mezenchymalnych (ang. *mesenchymal-like*, MES-like) (Neftel *et al.*, 2019). Proporcje komórek o określonym fenotypie nie były stałe, a stan, w którym znajdowała się większość komórek w danym guzie był determinowany przez zmiany genetyczne: amplifikacje w genach takich jak *CDK4*, *PDGFRA*, *EGFR* i mutacje w *NF1*. Opublikowane dane sugerują dodatkowo, że proporcje te mają istotny wpływ na skład i funkcjonowanie mikrośrodowiska nowotworów. Programy ekspresji genów

charakteryzujących stany podobne do NPC, OPC i AC są zakotwiczone w programie rozwoju mózgu, a częstość występowania tych stanów w guzie zależy od profilu genetycznych takich jak amplifikacje/mutacje w genach *CDK4, PDGFRA i EGFR* (Neftel *et al.*, 2019). W przeciwieństwie do tego, stan podobny do MES (MES-like) ma niewielkie podobieństwo do typów komórek wykrywanych w fizjologicznie zdrowym ludzkim mózgu i jest tylko częściowo związany ze zmianami genetycznymi glejaka. Wcześniejsze prace wykazały, że podtyp TCGA-MES jest skorelowany z liczebnością makrofagów i że mutacje lub delecje *NF1*, częste w guzach TCGA-MES, zwiększają rekrutację makrofagów (Wang *et al.*, 2017). Autorzy sugerowali, że podtyp TCGA-MES jest częściowo związany ze zwiększoną liczebnością makrofagów w guzie, co może również wyjaśniać jego dominację w sytuacjach nawrotowych (Ozawa *et al.*, 2014). Pomimo tych obserwacji, nie stwierdzono, co powoduje pojawienie się stanu nowotworowego podobnego do MES w glejaku.

Aby ustalić, czy istnieje związek przyczynowy między zwiększoną obecnością makrofagów, a stanem podobnym do MES w GBM, autorzy wykorzystali mysie modele GBM indukowane wzmożoną ekspresją onkogenu HrasG12V w połączeniu z wyciszeniem ekspresji genu kodującego białko supresorowe p53. Ustalili, że z niewielkimi różnicami w porównaniu z ludzkim GBM, cztery stany komórkowe zostały odtworzone w modelach mysich glejaków, a makrofagi były wzbogacone w środowisku komórek glejaków podobnych do MES (MES-like). Wykazali, że po zastosowaniu liposomów z klodronianem w celu wyeliminowania makrofagów, w nowotworach występuje zarówno zmniejszona liczba makrofagów, jak i komórek podobnych do MES. Na podstawie tych wyników autorzy wnioskowali, że to makrofagi wywołują stan podobny do MES w GBM. Co więcej, zidentyfikowali oni wydzielane przez makrofagi białko onkostatynę M (OSM) jako krytyczny czynnik wywołujący stan podobny do MES. OSM wiąże się z receptorami OSMR (ang. oncostatin M receptor) i LIFR (ang. leukemia inhibitory factor recepotor) w kompleksie z białkiem GP130 (ang. glycoprotein 130) na komórkach GBM, indukuje sygnalizację wewnatrzkomórkową z czynnikiem STAT3 (ang. signal transducer and activator of transcription 3). To prowadzi do przejścia komórek glejaka do stanu podobnego do MES, co z kolei indukuje programy mezenchymalne w samych makrofagach na zasadzie pętli zwrotnej. Stan

podobny do MES (MES-like) związany był ze zwiększoną liczebnością i cytotoksycznością limfocytów T (Hara *et al.,* 2021).

Zbadano różnice w ogólnych profilach scRNA-seq 16 próbek pacjentów z glejakiem z prawidłowym lub zmutowanym IDH1 (14 226 profili transkryptomicznych pojedynczych komórek). Wykazano, że oba typy guzów mają podobną hierarchię rozwojową i ścieżki różnicowania glejowego. Wraz ze wzrostem stopnia złośliwości guza stwierdzono zwiększoną proliferację komórek złośliwych, większą pulę niezróżnicowanych komórek glejaka i wyższą wartość sygnatury wskazującej na większą liczbę makrofagów w porównaniu z sygnaturą mikrogleju (Venteicher *et al.*, 2017).

#### 3.4.3 Rola układu odpornościowego w glejakach

#### 3.4.3.1 Układ odpornościowy w mózgu

Mózg jest bardzo dobrze odseparowany od pozostałych organów dzięki obecności bariery krew-mózg. Bariera ta chroni ośrodkowy układ nerwowy (OUN) przed niewłaściwą reakcją komórek układu odpornościowego krążących w krwi obwodowej i w efekcie wystąpieniem niepożądanego zapalenia. W OUN występuje różnorodna populacja komórek mieloidalnych, w skład której wchodzą przede wszystkim mikroglej (tkankowo specyficzne, rezydentne komórki mieloidalne biorące udział w odpowiedzi zapalnej), a także makrofagi pochodzące z przestrzeni okołonaczyniowych Virchowa-Robina, splotu naczyniówkowego i opon mózgowych (*ang. CNS border-associated macrophages*, BAMs) (Kierdorf *et al.*, 2019; Prinz *et al.*, 2021). Dodatkowo wykrywane są w OUN również komórki dendrytyczne i komórki NK (natural killers) choć ich liczba jest niewielka (0.5%). Rezydentne komórki mieloidalne w mózgu w stanie spoczynkowym wykazują stałą ekspresję białek należących do głównego układu zgodności tkankowej klasy I i II (MHC I i MHC II) i są zaliczane do profesjonalnych komórek prezentujących antygen (*ang. antigen-presenting cell*, APC). W stanie patologicznym komórki układu odpornościowego krążące w krwi obwodowej mogą migrować do OUN, w szczególności gdy uszkodzeniu ulegnie bariera krew-mózg. Wtedy w OUN mogą pojawić się monocyty, a także różnorodna populacja limfocytów.

Podczas infekcji OUN mikroglej fagocytuje zainfekowane komórki i prezentuje antygeny. Prowadzi to do aktywacji limfocytów T i wytwarzania przez nie cytokin prozapalnych interferonu (IFN)-γ i czynnika nekrozy nowotworu (TNF). Cytokiny te stymulują mikroglej do pochłaniania synaps długo po zakażeniu. Dodatkowo astrocyty i neurony przesyłają sygnał do mikrogleju za pośrednictwem różnych białek. Mikroglej ma zdolność prezentacji antygenu limfocytom T CD8+ i limfocytom T CD4+, a komórki dendrytyczne, pomimo swojej niewielkiej liczby, mogą aktywować limfocyty T CD4+ (Borst, Dumas and Prinz, 2021; Ochocka and Kaminska, 2021).

#### 3.4.3.2 Komórki układ odpornościowego w mikrośrodowisku glejaka

Mikrośrodowisko nowotworu jest różnorodne i składa się m.in. z komórek nowotworowych, astrocytów, komórek śródbłonka, a także komórek układu odpornościowego: mikrogleju, makrofagów, mieloidalnych komórek supresyjnych (myeloid-derived suppressor cells, MDSC), granulocytów, monocytów i limfocytów T i B (Gieryng et al., 2017; Ochocka & Kaminska, 2021). Poza różnorodnymi populacjami komórek, mikrośrodowisko składa się również z elementów i białek tworzących macierz zewnątrzkomórkową (*ang. extracellualr matrix*, ECM), wypełniającą przestrzeń pomiędzy komórkami i tworzącą połączenia z białkami na powierzchni komórek. Parenchyma mózgu ma unikalny skład, gdyż składa się głównie hialuronianu i pozbawiona jest sztywnych barier białkowych zbudowanych z kolagenu, fibronektyny i lamininy. Integryny i receptor hialuronianu CD44 uczestniczą w adhezji glejaka do ECM. Oddziałują również z proteazami wydzielanymi podczas progresji glejaka, które degradują ECM, umożliwiając komórkom nowotworowym rozprzestrzenianie się i dyfuzyjną infiltrację mózgu. Podczas procesu inwazji indukowane są również aktywatory plazminogenu, metaloproteinazy macierzy (MMP) i lizosomalne peptydazy cysteinowe (katepsyny) (Bellail *et al.*, 2004).

W wielu badaniach doświadczalnych i klinicznych wykazano, że mikrośrodowisko glejaków odgrywa ważną rolę w rozwoju nowotworu, ogranicza odpowiedzi układu odpornościowego na obecność komórek nowotworowych, a także ma wpływ na skuteczność terapii i rokowanie chorego (Gieryng *et al.*, 2017; Quail and Joyce, 2017; Kaminska, Ochocka and Segit, 2021). W przypadku niektórych glejaków wielopostaciowych aż 30% masy nowotworu stanowią komórki układu odpornościowego, które nagromadziły się w guzie. Zdecydowaną

większość tej populacji komórek stanowią mikroglej i makrofagi (*ang. glioma-associated microglia and macrophages*, GAMs), a ich liczba odwrotnie koreluje z rokowaniem chorego (Gieryng *et al.*, 2017). Badania w doświadczalnym modelu glejaka (komórki mysiego glejaka GL261 wszczepione do mózgu myszy) wykazały, że podczas wzrostu nowotworu jako pierwsze do obszaru guza migrują komórki mikrogleju, a dopiero po pewnym czasie pojawiają się w nim makrofagi pochodzące z krwi obwodowej (Gabrusiewicz *et al.*, 2011). Zwłaszcza rozróżnienie mikrogleju od napływowych monocytów/makrofagów pośród GAMs było kłopotliwe, bo większość używanych powszechnie przeciwciał rozpoznaje białka takie jak Cd11b, CD68, CX3CR1, F4/80, Iba1, które występują zarówno na powierzchni mikrogleju, jak i makrofagów (Kierdorf *et al.*, 2019)

Mając na uwadze wysoką heterogenność guzów, badanie glejaków i ich mikrośrodowiska nie może ograniczać się do wykorzystania tradycyjnych metod doświadczalnych, w których materiał (np. mRNA) otrzymywany jest z mieszaniny wielu komórek (czasem podzielonej na subpopulacje w oparciu o wartość kilku parametrów). Uzyskiwane w takiej analizie wyniki opierają się na uśrednionej wartości badanych parametrów we wszystkich komórkach znajdujących się w mieszaninie. Przy dużym zróżnicowaniu subpopulacji komórek w populacji, może to doprowadzić do wysuwania niewłaściwych wniosków i hipotez odnoszących się do nieistniejących w rzeczywistości komórek o uśrednionych cechach, sugerujących jednoczesne realizowanie wykluczających się funkcji. Z tego powodu wysoce wskazane jest, aby analizę mikrośrodowiska glejaków prowadzić wykorzystując metody wysokoprzepustowe umożliwiające badanie pojedynczych komórek, w których odczytanie informacji o odpowiednio dużej liczbie komórek pozwala zidentyfikować również relatywnie mniej liczne subpopulacje o bardzo istotnych rolach (np. regulatorowe limfocyty T).

### 3.4.3.3 Komórki układu odpornościowego w mikrośrodowisku glejaka w wysokoprzepustowych badaniach pojedynczych komórek

W ostatnich latach opublikowano wyniki badań pojedynczych komórek tworzących mikrośrodowisko różnych rodzajów glejaków w próbkach klinicznych (Darmanis *et al.*, 2017; Müller *et al.*, 2017; Venteicher *et al.*, 2017; Neftel *et al.*, 2019; Sankowski *et al.*, 2019; Friebel *et* 

al., 2020; Hara et al., 2021; Mathewson et al., 2021; Pombo Antunes et al., 2021; Zhang et al., 2021) Sekwencjonowanie transkryptomu pojedynczych komórek pozwoliło na bardziej szczegółową analizę różnic występujących pomiędzy komórkami mikrogleju, a makrofagami pochodzącymi z krwi obwodowej, które przemigrowały do guza. Müller et al. (2017) zaproponowali, że komórki mikrogleju można wyodrębnić na podstawie wysokiego poziomu ekspresji genu P2RY12, a komórki makrofagów na podstawie wysokiego poziomu ekspresji genu CD49d. Zauważyli również, że poziom HLA-DR (ang. human leukocyte antygen-DR) jest wyższy w populacji komórek zidentyfikowanych jako makrofagi niż w komórkach mikrogleju. Różnice w profilach transkryptomicznych pozwoliły zaobserwować, że komórki mikrogleju lokalizują się głównie na obrzeżach, a komórki makrofagów migrują w głąb guza i akumulują się w jego centrum (Darmanis et al., 2017; Müller et al., 2017). Dodatkowo, komórki na obrzeżach guza charakteryzowały się wyższym poziomem ekspresji genów kodujących prozapalną interleukinę 1B (IL1B), cytokin z rodziny CCL (ang. C-C Motif Chemokine Ligand): CCL2, CCL3, CCL4, TNF (ang. tumor necrosis factor), a także czynnika wzrostu kolonii makrofagów (colony-stimulating factor, CSF1) i jego receptora CSF1R. Komórki znajdujące się w centrum guza wykazywały wyższy poziom ekspresji genu VEGFA (ang. vascular endothelial growth factor) uczestniczącego w procesie angiogenezy, genu HIF1A (ang. hypoxia inducible factor 1 alpha) ulegającego ekspresji w środowisku niedotlenienia/hipoksji, a także genu kodującego antagonistę receptora interleukiny 1 (ang. interleukin 1 receptor antagonist, IL1RN) o działaniu przeciwzapalnym. Geny kodujące białka będące immunologicznymi punktami kontrolnymi (ang. immune-checkpoint) - CD274 (PD-L1, ang. programmed death ligand 1), PDCD1LG2 (PD-L2), CD80, CD86 (receptor CTLA4, ang. cytotoxic T-lymphocyte antigen 4) ulegały ekspresji w komórkach w całym obszarze guza, z nieznacznie podwyższonym poziomem na jego (Darmanis et al., 2017).

Wysokoprzepustowe metody badania pojedynczych komórek pozwoliły nie tylko na rozróżnienie typów komórek występujących w mikrośrodowisku, ale także na analizę funkcji poszczególnych subpopulacji. Na podstawie sekwencjonowania pojedynczych komórek próbek klinicznych GBM i próbek "niezmienionej" (*ang. normal-appearing*) tkanki mózgu pobranych od pacjentów z epilepsją (potraktowanych jako próbki kontrolne) Sankowski et al. (2019) wykazali, że w środowisku glejaka komórki mieloidalne wykazywały niższy poziom ekspresji genów charakterystycznych dla mikrogleju w stanie homeostazy, a z drugiej strony miały podwyższoną ekspresję genów powiązanych z określonymi programami funkcjonalnymi. Autorzy zidentyfikowali subpopulacje komórek o podwyższonej ekspresji genów indukowanych przez interferon (IFI27, IFITM3), powiązanych z metabolizmem lipidów (LPL, APOE, TREM2), a także genów kodujących białka kompleksów MHCI i MHCII, i ulegających ekspresji w środowisku hipoksji (VEGFA). Co zastanawiające, w analizowanych próbkach wśród komórek mieloidalnych nie zidentyfikowali komórek monocytów/makrofagów pochodzących z krwi obwodowej, a jedynie komórki mikrogleju. Korzystając z metody CyTOF potwierdzili obecność białek HLA-DR, TREM2 i APOE na komórkach, na których wykryto obecność P2RY12 i TMEM119 (zidentyfikowanych jako komórki mikrogleju), natomiast nie zaobserwowali obecności białek kodowanych przez geny indukowane przez interferon. W kolejnym badaniu mikrośrodowiska GBM (Klemm et al., 2020) autorzy, na podstawie analizy danych RNA-seq komórek CD49d+ (zidentyfikowanych jako komórki monocytów/makrofagów pochodzące z krwi obwodowej) i CD49- (zidentyfikowanych jako komórki mikrogleju) autorzy wykazali, że ekspresja genów powiązanych z odpowiedzią na interferon I występuje głównie w komórkach CD49d+. W pracy Antunes et al. (2021) została przedstawiona bardziej szczegółowa charakterystyka subpopulacji komórek mieloidalnych. Na podstawie profili transkryptomicznych pojedynczych komórek autorzy, podobnie jak (Sankowski et al., 2019), wyróżnili trzy główne programy funkcjonalne prezentowane przez te komórki: program odpowiedzi na interferon (powiązana z ekspresją genów STAT1, IFIT2, ISG15 i CXCL10), program fagocytozy połączonej z metabolizmem lipidów (powiązana z ekspresją genów GPNMB, LGALS3, FABP5 i CD9), a także program odpowiedzi na środowisko hipoksji (powiązany z ekspresją genów BNIP3, ADAM8, MIF i HILPDA). Programy te zidentyfikowano zarówno w komórkach z próbek pochodzących z guzów pierwotnych, jak i z próbek pobranych ze wznów GBM, a także w komórkach z doświadczalnego modelu glejaka, w którym komórki GL261 zostały wszczepione do mózgu myszy C57BL6. W tej pracy autorzy zidentyfikowali wśród badanych komórek subpopulację mikrogleju, a także kilka subpopulacji monocytów/makrofagów. Geny powiązane z programami odpowiedzi na interferon i odpowiedzi środowisko hipoksji ulegały ekspresji na wyższym poziomie w komórkach na monocytów/makrofagów niż w komórkach mikrogleju. Dodatkowo, profile transkryptomiczne

subpopulacji monocytów/makrofagów mogą sugerować, że z krwi obwodowej do obszaru rozwoju guza napływają monocyty, które dopiero w mikrośrodowisku nowotworu różnicują w dojrzałe makrofagi, jednocześnie zmieniając swój fenotyp.

Komórki, w których ekspresji ulegają geny powiązane z odpowiedzią na interferon i geny powiązane z fagocytozą/metabolizmem lipidów mogą pełnić odmienne role. Na podstawie wyników badań scRNA-seq i CyTOF można przypuszczać, że komórki monocytów o silnym fenotypie przeciwnowotworowym migrują do mikrośrodowiska guza, w którym to (pod wpływem czynników znajdujących się w tym mikrośrodowisku) różnicują w makrofagi o fenotypie immunosupresyjnym, wspierającym rozwój nowotworu.

#### 4 CEL PRACY

Mikrośrodowisko doświadczalnych glejaków nie było zbyt szczegółowo zbadane pod kątem heterogenności rezydentnych i naciekających komórek układu odpornościowego. Wcześniejsze badania na komórkach CD11b+ sortowanych z półkuli z guzem w różnym czasie po implantacji glejaka wskazywały wczesną akumulację mikrogleju (komórki CD11b+CD45<sup>low</sup>) w 7 dniu po implantacji komórek glejaka i obecność komórek mieloidalnych z obwodu (komórki CD11b+CD45<sup>high</sup>) 15 dnia po operacji (Gabrusiewicz *et al.*, 2011). Badania transkryptomiczne RNA-seq wskazywały profile transkryptomiczne, które jednocześnie prezentowały części typowe dla odpowiedzi zarówno zapalnej jak i immunosupresyjnej. Aby odpowiedzieć na pytanie, jakie komórki i o jakich funkcjach nagromadzają się w mikrośrodowisku doświadczalnych glejaków, wprowadzono metodykę transkryptomiki pojedynczych komórek (Ochocka, Segit et al., 2021; Ochocka et al., 2023). Wprowadzenie tych metod wymagało opracowania podejścia do analiz danych scRNA-seq dopasowanego do złożonych układów doświadczalnych.

Celem niniejszej rozprawy było opracowanie zoptymalizowanej metodyki analizy i wizualizacji danych scRNA-seq z komórek mieloidalnych w mikrośrodowisku doświadczalnych glejaków oraz w próbkach z guzów glejaka wielopostaciowego.

Cele szczegółowe obejmowały:

- 1. Opracowanie zoptymalizowanej metodyki analizy i wizualizacji danych scRNA-seq z komórek mieloidalnych w mikrośrodowisku doświadczalnych glejaków.
- 2. Ocenę wpływu wartości parametrów wykorzystywanych metod dla uzyskania optymalnych wyników analiz i wizualizacji danych scRNA-seq.
- 3. Zbadanie w danych scRNA-seq różnic w profilach transkryptomicznych komórek mieloidalnych w mózgach myszy naiwnych i po operacji pozorowanej.
- Zbadanie w danych scRNA-seq heterogenności komórek mikrogleju i makrofagów w mikrośrodowisku ludzkich guzów glejaka wielopostaciowego.

#### 5 METODY

#### 5.1 Konwersja plików BCL do formatu FASTQ

Uzyskano wyniki sekwencjonowania scRNA-seq bibliotek przygotowanych za pomocą urządzenia Chromium Controller i zestawu odczynników 10X Genomics Single-Cell Gene Expression v2 lub Single-Cell Gene Expression v3 (w przypadku danych scRNA-seq naiwne versus operacja pozorowana) (Ochocka, Segit et al., 2021). Wynik sekwencjonowania następnej generacji na urządzeniach produkowanych przez firmę Illumina zapisywany jest w plikach BCL (base call). Są to pliki binarne o następującej strukturze:

- Bajty 0 do 3 zawierają liczbę klastrów na powierzchni komory przepływowej (N) w formacie little endian 32-bit
- Bajty 4 do (N+3), gdzie N to liczba klastrów, zawierają informację o odczytanych zasadach i jakości odczytu. Bajty mają poniższą strukturę:
  - Bity 0 do 1 kodują odczytaną zasadę (A: 00, C: 01, G: 10, T: 11)
  - Bity 2-7 kodują wartość odpowiadającą jakości odczytu
  - Kiedy wszystkie 8 bitów to "0" oznacza to, że dany klaster nie został odczytany ("no-call")

(https://emea.support.illumina.com/content/dam/illumina-

support/documents/documentation/software\_documentation/bcl2fastq/bcl2fastq2-v2-20-software-guide-15051736-03.pdf)

Następnie te surowe dane konwertowane były do formatu FASTQ za pomocą oprogramowania bcl2fastq dostarczanego przez firmę Illumina. Podczas konwersji, dla każdego klastra który przejdzie ustalone reguły filtrowania, tworzony jest jeden wpis w pliku FASTQ (w przypadku odczytywania DNA z jednej strony (*ang. single-end*) tworzony jest jeden plik FASTQ, w przypadku odczytywania DNA z dwóch stron (*ang. paired-end*) tworzone są dwa pliki FASTQ). Proces konwersji obejmuje demultipleksowanie odczytów (tzn. przypisanie odczytów do właściwych próbek, na podstawie zdefiniowanych sekwencji indeksowych dodawanych podczas przygotowania biblioteki do sekwencjonowania, unikalnych dla poszczególnych próbek.

Dodatkowo (w zależności od wybranych ustawień) oprogramowanie bcl2fastą może w trakcie działania przyciąć sekwencje adapterowe i usunąć unikalne identyfikatory molekularne (UMI). Po zakończeniu konwersji wyniki zapisywane są w plikach tekstowych FASTQ, wykorzystywanych jako pliki wejściowe w wielu narzędziach analizy danych NGS.

W przypadku danych z pojedynczych komórek (scRNA-seq) uzyskanych z próbek przygotowanych w systemie 10x Genomics, do konwersji plików BCL do plików FASTQ zalecane było wykorzystanie oprogramowania Cell Ranger mkfastq. Jest to nakładka na oprogramowanie bcl2fastq, która ma domyślnie ustawione wartości parametrów bcl2fastq zgodnie z wytycznymi 10x Genomics. Konwersja plików BCL do plików FASTQ wykorzystanych w analizach przedstawionych w niniejszej pracy została wykonana korzystając z oprogramowania Cell Ranger mkfastq w wersji 3.0.1 i domyślnych wartościach parametrów.

# 5.2 Mapowanie odczytów do transkryptomu referencyjnego i stworzenie macierzy zliczeń odczytów zmapowanych do poszczególnych genów w pojedynczych komórkach

Macierze ekspresji genów w pojedynczych komórkach, to macierz w której w kolumny odpowiadają pojedynczym komórkom, a wiersze poszczególnym genom (lub na odwrót), natomiast wartości w macierzy to liczba unikalnych zliczeń zmapowanych do konkretnego genu w konkretnej komórce.

W celu uzyskania takich macierzy dla danych scRNA-seq uzyskanych z próbek przygotowanych w systemie 10x Genomics producent rekomenduje wykorzystanie oprogramowania Cell Ranger count (https://www.10xgenomics.com/support/software/cellranger/7.2/algorithms-overview/cr-gex-algorithm). Oprogramowanie to przyjmuje pliki FASTQ jako dane wejściowe, a następnie wykonuje mapowanie pojedynczych odczytów do transkryptomu referencyjnego (za pomocą zewnętrznego oprogramowania STAR z wartościami parametrów ustawionymi zgodnie z rekomendacjami 10x Genomics). W kolejnych krokach wyniki są filtrowane, a unikalne odczyty zmapowane do konkretnego genu i te przypisane do jednego znacznika są sumowane. W efekcie powstaje (w surowej formie) macierz ekspresji genów we wszystkich odczytanych znacznikach. Następnie, na podstawie wyników algorytmu identyfikacji komórek, znaczniki są filtrowane i w macierzy zostają tylko te, które (najprawdopodobniej) reprezentują pojedynczą komórkę, która została zamknięta w kropli podczas przygotowywania próbki w systemie 10x Genomics. W praktyce macierz ta jest dodatkowo filtrowana na późniejszym etapie analizy polegającym na ocenie jakości komórek (m.in. na podstawie liczby genów, do których zostały zmapowane odczyty dla danego znacznika lub na podstawie procentowej zawartości odczytów zmapowane od genów kodujących białka mitochondrialne). Mapowanie odczytów do transkryptomu referencyjnego i stworzenie macierzy zliczeń wykorzystanych w analizach przedstawionych w niniejszej pracy zostało wykonane z wykorzystaniem oprogramowania Cell Ranger count w wersji 3.0.1. W przypadku danych pochodzących z mikrośrodowiska mysiego glejaka GL261 wybrano transkryptom referencyjny w wersji mm10-3.0.0, a w przypadku danych pochodzących z mikrośrodowiska ludzkiego glejaka wielopostaciowego wybrano transkryptom referencyjny w wersji GRCh38-3.0.0 i domyślnych wartościach pozostałych parametrów.

#### 5.3 Analiza jakości danych w macierzy ekspresji

Jakość komórek ma kluczowe znaczenie dla jakości uzyskanych danych, a w konsekwencji również dla jakości wyników analiz. Z tego względu w materiałach 10x Genomics wielokrotnie zwracana jest uwaga na właściwe przygotowanie próbki w laboratorium (https://cdn.10xgenomics.com/image/upload/v1686678481/support-

documents/CG00053\_Handbook\_CellPreparation\_SingleCellProtocols\_Rev\_D.pdf). Niezależnie od tego, po zakończeniu sekwencjonowania i uzyskaniu (przefiltrowanej) macierzy ekspresji genów w pojedynczych komórkach wykonuje się dodatkowe procedury kontroli jakości danych. Do najczęściej stosowanych należą:

- sprawdzenie liczby unikalnych odczytów przypisanych do poszczególnych znaczników,
- sprawdzenie liczby genów, do których zostały przypisane odczyty w obrębie jednego znacznika,

 sprawdzenie jaki procent unikalnych odczytów został zmapowany do genów kodujących białka mitochondrialne.

Ma to na celu wykrycie przede wszystkim znaczników, które przez algorytm w oprogramowaniu Cell Ranger zostały zaklasyfikowane do zbioru komórek, a w rzeczywistości mogły być komórkami obumierającymi (z nieszczelną błoną komórkową, przez którą część mRNA wypłynęła z cytoplazmy na zewnątrz jeszcze przed zamknięciem komórki w kropli podczas procesu enkapsulacji).

Nie ma uniwersalnych wartości tych parametrów, na podstawie których należałoby filtrować macierz ekspresji. Z tego względu w przypadku każdego zbioru danych należy sprawdzić jaki jest rozkład tych wartości, a następnie, uwzględniając dodatkowo biologiczną wiedzę o komórkach, które znajdują się w zbiorze, wyznaczyć wartości progowe każdego z parametrów. Dzięki temu można odfiltrować znaczniki odpowiadające najpewniej komórkom obumierającym lub kroplom, w których nie znalazła się komórka, a jedynie relatywnie duża ilość mRNA wolno pływającego mRNA. Inny przypadek, gdy wskazane jest odfiltrowanie znacznika, to sytuacja w której w jednej kropli znalazły się przynajmniej dwie komórki. W materiałach 10x Genomics (https://kb.10xgenomics.com/hc/en-us/articles/360001378811-What-is-the-maximum-

number-of-cells-that-can-be-profiled) zawarto informację o szacowanym procencie znaczników, które mogą reprezentować takie zdarzenia (multiplety) w zależności od liczby komórek, które nałożono na płytkę i zastosowanego protokołu. Potencjalne multiplety to znaczniki, w których liczba odczytów odstaje w górę od pozostałych znaczników lub znaczniki, w których identyfikowana jest jednoczesna ekspresja genów kodujących białka, których obecność jest charakterystyczna dla różnych typów komórek.

Filtrowanie danych w macierzy ekspresji zostało wykonane z wykorzystaniem pakietu Seurat v3 i funkcji subset. W przypadku danych pochodzących z mikrośrodowiska mysiego glejaka GL261 w zbiorze pozostawiono znaczniki, w których liczba wykrytych genów była większa niż 200 i mniejsza niż 3000, a także w których poniżej 5% odczytów zostało zmapowanych do genów kodujących białka mitochondrialne. W przypadku danych pochodzących z mikrośrodowiska ludzkiego glejaka wielopostaciowego w zbiorze pozostawiono znaczniki, w których liczba wykrytych genów była większa niż 200, a także w których poniżej 20% odczytów zostało zmapowanych do genów kodujących białka mitochondrialne i wartość log10GenesPerUmi była większa niż 0.8. Znaczniki pozostałe w macierzach ekspresji po filtrowaniu były na dalszych etapach analiz traktowane jako reprezentujące pojedyncze komórki.

#### 5.4 Normalizacja danych

Zbiór danych może być bardzo zróżnicowany pod kątem typów komórek, które w nim występują. Dodatkowo efektywna głębokość sekwencjonowania komórek z tej samej populacji może być inna i nie ma możliwości ujednolicenia jej na etapie sekwencjonowania. Z tego względu jednym z etapów wstępnego przygotowania danych do analizy jest ich normalizacja, umożliwiająca porównywanie ekspresji konkretnych genów pomiędzy różnymi komórkami. W analizach opisanych w niniejszej pracy wykorzystano zaimplementowaną w pakiecie Seurat v3 (Stuart *et al.*, 2019)metodę normalizacji ekspresji w każdej komórce w odniesieniu do całkowitej liczby odczytów przypisanych do tej komórki (tj. normalizacja względem wielkości biblioteki). Tak uzyskane wartości były mnożone przez domyślny współczynnik skalujący (10000) i poddawane transformacji logarytmicznej ln(10000\*"znormalizowana wartość ekspresji"+1). Ta procedura została wykonana dla wszystkich próbek, które wykorzystano w analizach przedstawionych w niniejszej pracy. Na dalszych etapach analiz wykorzystywane były również dane wystandaryzowane. Standaryzacja polegała na poddaniu transformacji z-score macierzy znormalizowane jekspresji, uzyskanej w poprzednim kroku.

#### 5.5 Integracja wielu zbiorów danych

Analiza danych pochodzących z wielu próbek związana jest najczęściej z potrzebą połączenia wielu zbiorów danych (macierzy ekspresji) w jedną macierz zawierającą komórki ze wszystkich analizowanych próbek (z zachowaniem informacji o tym, z której próbki pochodzi każda komórka). Najprostsze podejście do tego zadania polega na złączeniu wszystkich macierzy w jedną macierz. Nie zawsze jest to jednak rozwiązanie akceptowalne, ponieważ na dalszych etapach analizy może okazać się, że populacje komórek uzyskane w wyniku grupowania są

zdominowane przez komórki pochodzące z konkretnych próbek. Sugeruje to zwykle, że wpływ przygotowania lub pochodzenia próbki na profile transkryptomiczne komórek w tej próbce jest silniejszy niż różnice w ekspresji genów pomiędzy biologicznie zdefiniowanymi populacjami komórek. Jeżeli pewna zdefiniowana populacja komórek znajduje się w wielu analizowanych próbkach, to w wyniku grupowania komórki z tej populacji, pochodzące ze wszystkich próbek, powinny zostać przypisane do tej samej grupy.

Występowanie niepożądanej zmienności pomiędzy próbkami (np. związanej z procedurą przygotowania próbki) może jednak wymagać zastosowania bardziej zaawansowanych metod integracji zbiorów, polegających na wykonaniu korekcji wartości ekspresji genów w poszczególnych komórkach. W analizach przedstawionych w niniejszej pracy wykorzystano metodę zaimplementowaną w pakiecie Seurat (Stuart et al., 2019). Metoda ta w obejmuje stworzenie rankingu genów według wartości wystandaryzowanej wariancji w poszczególnych zbiorach danych, a następnie wybór tych które znalazły się wśród grupy 2000 genów o największej wystandaryzowanej wariancji w jak największej liczbie integrowanych zbiorów. Dalsze kroki integracji wykonano wykorzystując 2000 genów wybranych w ten sposób. Następnie w sposób nienadzorowany identyfikowane są kotwice (ang. anchors) pomiędzy integrowanymi zbiorami. Kotwica jest to para komórek (po jednej z dwóch analizowanych bezpośrednio zbiorów), które według predykcji pochodzą z jednej populacji komórek. Poszukiwanie kotwic rozpoczyna się od wykonania redukcji wymiarów wykorzystując analizę kanonicznej korelacji (CCA) i normalizację L2. Następnie, analizując znormalizowane L2 wektory kanonicznej korelacji (CCV) identyfikowanych jest k najbliższych sąsiadów (kNN) dla każdej komórki w każdej parze integrowanych zbiorów. Wyniki wykorzystywane są do identyfikacji wspólnych najbliższych sąsiadów (ang. mutual nearest neighbors), czyli par komórek, po jednej z każdego z analizowanych zbiorów, które wzajemnie znajdują się wśród swoich k najbliższych sąsiadów. W wykonanych analizach wartość k była równa 5. Zbiór kotwic to zbiór zidentyfikowanych wspólnych najbliższych sąsiadów, reprezentujących w założeniu pary komórek pochodzące z jednej populacji, a obecne w obu analizowanych zbiorach. Następnie kotwice są poddawane ocenie jakości i wyznaczane są dla każdej z nich wagi wykorzystywane następnie podczas integracji dwóch zbiorów danych. Oryginalne wartości ekspresji genów poddawane są korekcji uwzględniającej zaobserwowane różnice w profilach transkryptomicznych komórek tworzących poszczególne kotwice, a także wyznaczone wagi dla każdej z kotwic. Tak uzyskane wartości mogą być następnie połączone w jedną macierz (skorygowanej) ekspresji i traktowane w dalszych analizach jak zwykła macierz ekspresji dla pojedynczej próbki, po normalizacji. Integracja danych analizowanych w niniejszej pracy została wykonana korzystając z pakietu Seurat v3. W przypadku danych pochodzących z mikrośrodowiska mysiego glejaka GL261 integrowano próbki będące powtórzeniami biologicznymi w obrębie każdej z badanych grup (n=4). Kotwice zidentyfikowano korzystając z funkcji FindIntegrationAnchors (parametr dims=1:30), a integrację wykonano funkcją IntegrateData (parametr dims=1:30). W przypadku danych pochodzących z mikrośrodowiska ludzkiego glejaka wielopostaciowego zintegrowano dane pochodzące z wszystkich pięciu analizowanych próbek. Kotwice zidentyfikowano korzystając z funkcji IntegrateData (parametr dims=1:30), a integrację wykonano funkcją IntegrationAnchors (parametr dims=1:30), a integracją z funkcji z funkcji z glejaka wielopostaciowego zintegrowano korzystając z funkcji pięciu analizowanych próbek. Kotwice zidentyfikowano korzystając z funkcji IntegrateData (parametr dims=1:30), a integrację wykonano funkcją IntegrateData

W przypadku więcej niż dwóch zbiorów danych integrację przeprowadza się etapami, gdzie na każdym etapie integrowane są dokładnie dwa zbiory danych (z których każdy może być wynikiem integracji wykonanej na jednym z poprzednich etapów). Przed rozpoczęciem sekwencji integracji obliczane są odległości pomiędzy wszystkimi możliwymi parami integrowanych zbiorów, a następnie korzystając z klastrowania hierarchicznego wyznaczany jest schemat, kolejne etapy, integracji (w pakiecie Seurat zostało to zaimplementowane jako część funkcji IntegrateData). Po zakończeniu wszystkich etapów powstaje jedna macierz (skorygowanej) ekspresji, zawierająca komórki ze wszystkich integrowanych zbiorów.

#### 5.6 Redukcja wymiarów

Macierze ekspresji genów w pojedynczych komórkach to dane wielowymiarowe, gdzie liczba wymiarów równa jest liczbie analizowanych genów i najczęściej jest to od kilku do kilkudziesięciu tysięcy wymiarów. Jednym z kroków wstępnego przetwarzania takich danych jest redukcja wymiarów, służąca transformacji danych w taki sposób, aby zachować możliwie dużą część informacji zawartych w oryginalnym zbiorze danych, ale zapisanych w znacznie mniejszej liczbie wymiarów. W analizach opisanych w niniejszej pracy zastosowano metody najczęściej obecnie wykorzystywane w analizie ekspresji genów w pojedynczych komórkach: analizę składowych głównych (PCA), algorytm t-SNE (*ang. t-Distributed Stochastic Neighbor Embedding*) i algorytm UMAP (*ang. Uniform Manifold Approximation and Projection*).

Przy wizualizacji danych transkryptomicznych z pojedynczych komórek powszechnie spotykane podejście to wykonanie pierwszej redukcji wymiarów korzystając z PCA, a następnie wizualizacja danych korzystająca z przestrzeni t-SNE/UMAP obliczonej dla wybranej liczby składowych głównych. W analizach przedstawionych w niniejszej pracy w przypadku danych pochodzących z mikrośrodowiska mysiego glejaka GL261 wyniki PCA uzyskano korzystając w pakiecie Seurat v3 z funkcji RunPCA (parametr dims=1:50), a następnie przejście do przestrzeni dwuwymiarowych t-SNE i UMAP wykonano korzystając z pierwszych 30 składowych głównych i funkcji, odpowiednio, RunTSNE (parametr dims=1:30) i RunUMAP (parametr dims=1:30, parametr n.neighbors=35). W przypadku danych pochodzących z mikrośrodowiska ludzkiego glejaka wielopostaciowego wyniki PCA uzyskano korzystając w pakiecie Seurat v3 z funkcji RunPCA (parametr dims=1:50, a następnie przejście do przestrzeni dwuwymiarowej UMAP wykonano korzystając z pierwszych 40 składowych głównych i funkcji RunUMAP (parametr dims=1:40, n.neighbors=30).

#### 5.7 Grupowanie komórek

Analiza ekspresji genów w pojedynczych komórkach wykonywana jest najczęściej na zbiorach zawierających dane z kilku do kilkuset tysięcy komórek. Górnym ograniczeniem liczby analizowanych komórek jest dostępność danych i mocy obliczeniowej potrzebnej do ich przetworzenia. W przypadku dużych projektów, takich jak przygotowanie atlasów komórek poszczególnych tkanek lub narządów, są to czasem miliony komórek. Analiza profili transkryptomicznych na poziomie pojedynczych komórek (obserwacji) nie ma najczęściej uzasadnienia i wskazane jest połączenie komórek w grupy (*ang. clusters*) zawierające komórki o zbliżonych profilach. Umożliwia to następnie charakterystykę poszczególnych grup i porównywanie ich z innymi grupami w zbiorze. W zależności od tego jak bardzo heterogenny jest

analizowany zbiór danych i jak bardzo szczegółowa jest wykonywana analiza komórki przypisywane są w większości przypadków do od kilku do kilkudziesięciu grup.

Grupowanie jest zadaniem nienadzorowanego przypisania elementów zbioru (obserwacji) do grup. W przeciwieństwie do zadania klasyfikacji pod nadzorem, nie bazuje na przykładach obserwacji przypisanych do konkretnych kategorii (klas). Grupowanie ma na celu wykonanie takiego podziału zbioru na grupy, aby obserwacje znajdujące się w jednej grupie były możliwie do siebie podobne. W przypadku analizy pojedynczych komórek, komórki w jednej grupie powinny reprezentować ten sam typ/stan. W analizach przedstawionych w niniejszej pracy wykorzystano algorytmy zaimplementowane w pakiecie Seurat (Stuart et al., 2019) opierające się na wcześniej zaproponowanych rozwiązaniach (Macosko et al., 2015). Jest to podejście wykorzystujące model grafowy. W pierwszym kroku tworzony jest graf k najbliższych sąsiadów (KNN) na podstawie odległości euklidesowej pomiędzy komórkami policzonej w przestrzeni PCA. Wagi krawędzi między wierzchołkami grafu (komórkami) są modyfikowane korzystając z współczynnika Jaccarda jako miary podobieństwa pomiędzy sąsiedztwami poszczególnych par komórek. Następnie, korzystając z technik optymalizacji modularności (domyślnie z metody Louvaina) komórki przypisywane są do grup. Liczba grup, do których zostaną przypisane komórki nie jest określana przed rozpoczęciem grupowania, a zależy od wartości parametru rozdzielczości (resolution) – czym większa wartość tego parametru tym spodziewana jest większa liczba grup (klastrów) komórek. W przypadku analizy ekspresji genów w pojedynczych komórkach nie ma przyjętej optymalnej wartości rozdzielczości (lub liczby grup). Najczęściej wymagane jest wielokrotne wykonanie grupowania, przy różnych wartościach parametrów, porównanie wyników i zestawienie ich z aktualną wiedzą biologiczną. Otrzymane w ten sposób populacje komórek można następnie charakteryzować i porównywać analizując geny o statystycznie istotnie podwyższonej ekspresji w poszczególnych populacjach.

W analizach przedstawionych w niniejszej pracy w przypadku danych pochodzących z mikrośrodowiska mysiego glejaka GL261 wyszukiwanie najbliższych sąsiadów zostało wykonane korzystając z funkcji FindNeighbors (parametr dims=1:30), a grupowanie korzystając z funkcji FindClusters (wybór wartości parametru resolution przestawiono w części Wyniki). W przypadku danych pochodzących z mikrośrodowiska ludzkiego glejaka wielopostaciowego wyszukiwanie
najbliższych sąsiadów zostało wykonane korzystając z funkcji FindNeighbors (parametr dims=1:40), a grupowanie korzystając z funkcji FindClusters (w tym przypadku wybór wartości parametru resolution również przestawiono w części Wyniki).

#### 5.8 Analiza różnicowej ekspresji genów

Kolejnym etapem analizy ekspresji genów w pojedynczych komórkach, po uzyskaniu przypisania komórek do grup w wyniku grupowania, jest identyfikacja genów pozwalających scharakteryzować poszczególne populacje. Geny te powinny mieć statycznie istotnie zmienioną ekspresję w opisywanej populacji w porównaniu do pozostałych komórek w zbiorze, a z drugiej strony kluczowa jest właściwa interpretacja funkcji/wpływu ekspresji tych genów na komórkę. Nie wszystkie geny są równie istotne uwzględniając wiedzę biologiczną, a fakt ten nie jest uwzględniany w trakcie przetwarzania danych (np. poziom ekspresji powszechnie znanego genu markerowy pewnego typu komórek będzie miał na etapie obliczeniowym taką samą wagę jak poziom ekspresji każdego innego genu z zbiorze). Z tego względu konieczna jest merytoryczna weryfikacja uzyskiwanych wyników, oparta na aktualnej wiedzy biologicznej.

W pakiecie Seurat (Stuart *et al.*, 2019) domyślna metoda analizy różnicowej genów pomiędzy klastrami wykorzystuje test sumy rang Wilcoxona (znany również jako test Manna-Whitneya). Najważniejsze parametry jakie należy określić to minimalny procent komórek z ekspresją testowanego genu w przynajmniej jednej z grup (min.pct) i minimalna akceptowana średnia zmiana krotności ekspresji pomiędzy klastrami (logfc.threshold). Analiza różnicowej ekspresji może zostać wykonana pomiędzy dwoma wybranymi klastrami lub porównując komórki z konkretnego klastra do wszystkich pozostałych komórek. Druga z tych opcji służy identyfikacji genów o najbardziej statystycznie istotnie podwyższonej ekspresji w poszczególnych klastrach w porównaniu do pozostałych komórek w zbiorze. W połączeniu z interpretacją funkcji poszczególnych genów, opartą na aktualnej wiedzy biologicznej, informacja ta jest kluczowa dla właściwego opisu populacji komórek zidentyfikowanych w analizowanym zbiorze, a także np. do zidentyfikowania transkryptomicznych różnic pomiędzy populacjami pochodzącymi z próbek należących do różnych warunków eksperymentalnych.

W analizach przedstawionych w niniejszej pracy w przypadku danych pochodzących z mikrośrodowiska mysiego glejaka GL261 w celu identyfikacji genów o statystycznie istotnie podwyższonej ekspresji wykorzystano w pakiecie Seurat v3 funkcję FindAllMarkers, która wykonuje porównanie profili transkryptomicznych komórek przypisanych do konkretnego klastrów, do profili transkryptomicznych pozostałych komórek w zbiorze. Procedura wykonywana jest iteracyjnie dla wszystkich zidentyfikowanych klastrów będących wynikiem grupowania. W funkcji FindAllMarkers ustawiono parametry assay="RNA", min.pct=0.25, only.pos=T i logfc.threshold=0.25. W przypadku danych pochodzących z mikrośrodowiska ludzkiego glejaka wielopostaciowego również wykorzystano funkcję FindAllMarkers (parametr assay="RNA", min.pct=0.5, only.pos=T i logfc.threshold=0.25).

### 6 WYNIKI

6.1 Analiza profili transkryptomicznych pojedynczych komórek CD11b+ pochodzących z mikrośrodowiska mysiego glejaka GL261

#### 6.1.1 Przygotowanie i normalizacja danych uzyskanych z scRNAseq

Mimo licznych badań transkryptomicznych komórek CD11b+ izolowanych z mózgu z glejakiem uzyskiwano niejednoznaczne wyniki (co do fenotypu) sugerujące występowanie subpopulacji komórek mieloidalnych. Nie znano jednak białek powierzchniowych, które mogłyby zostać wykorzystane do wysortowania subpopulacji i późniejszej analizy ich profili transkryptomicznych. W celu uzyskania bardziej szczegółowej charakterystyki subpopulacji tych komórek w mikrośrodowisku glejaka wdrożono technologie sekwencjonowania RNA 3' komórek Cell pojedynczych (Single Gene Expression v2 (https://www.10xgenomics.com/support/single-cell-gene-expression)) wsród populacji CD11b+, wysortowanych z mózgu kontrolnego lub z półkuli z guzem w 14 dniu po implantacji komórek glejaka GL261 do mózgu myszy C57BL/6. w modelu syngenicznym. Do stwozrenmia bibliotek wykorzystano urządzenie mikroprzepływowe Chromium Controller (https://www.10xgenomics.com/instruments/chromium-controller). Szczegółowy opis procedury implantacji komórek glejaka i izolowania komórek CD11b+ z mózgu znajduje się w pracy (Ochocka, Segit et al., 2021) oraz w rozprawie doktorskiej dr Natalii Ochockiej-Lewickiej (2022). Analizowano także, czy zmiany te różnią się w zależności od płci wszczepiając komórki glejaka myszom różnej płci.

W doświadczeniu zbadano 4 grupy: samce i samice naiwne oraz samce i samice, którym zaimplantowano komórki GL261. Parametry próbek umieszczanych na płytce urządzenia zostały dobrane w taki sposób, aby uzyskać dane z 5000 komórek dla każdej z próbek. W przypadku zastosowanego protokołu scRNA-seq określenie liczby komórek, dla których odczytano profile transkryptomiczne możliwe jest dopiero na etapie analiz bioinformatycznych danych będących efektem sekwencjonowania i zostało przestawione w dalszej części pracy. Dla każdej z grup eksperymentalnych wykonano dwa biologiczne powtórzenia.

Surowe dane zostały zdemultipleksowane, a następnie otrzymane odczyty zmapowano do mysiego transkryptomu referencyjnego GRCm38 (mm10) i zliczono liczbę odczytów zmapowanych do poszczególnych genów dla poszczególnych znaczników (*ang. barcode*). W ten sposób dla każdej próbki otrzymano dane do macierzy ekspresji genów w pojedynczych "komórkach". Łącznie zidentyfikowano 41059 komórek, dla których uzyskano średnio 36412 odczytów na komórkę (Tabela 6.1)

Tabela 6.1 Liczba zidentyfikowanych komórek i średnia liczba odczytów na komórkę w poszczególnych próbkach (dane przetworzone i wstępnie przefiltrowane oprogramowaniem Cell Ranger)

	CD11b+ ze zwierząt naiwnych				CD11b+ ze zwierząt 14 dni po implantacji komórek					
					GL261					
Próbka	F1	F2	M1	M2	F1	F2	M1	M2	Średnia	Suma
Liczba	5223	4870	4873	5301	5802	5579	4402	5009	5150	41059
zidentyfikowanych										
komórek										
Średnia liczba	42512	33630	35228	37195	31190	31680	43450	31842	36412	
odczytów na										
komórkę										

Analiza profili transkryptomicznych została przeprowadzona w środowisku R (https://www.r-project.org/) wykorzystując funkcje pakietu Seurat v3 (Stuart et al., 2019). Pierwszym etapem kontroli jakości było sprawdzenie liczby odczytów przypisanych do poszczególnych znaczników, aby zidentyfikować i odfiltrować dane pochodzące z kropli najprawdopodobniej nie zawierających żywych komórek, a jedynie wolne mRNA. Na podstawie uzyskanych wyników odrzucono z dalszej analizy wszystkie znaczniki, dla których liczba unikalnych genów, do których zostały zmapowane odczyty była poniżej 200 (jako najprawdopodobniej nie reprezentujące żywych komórek). Dodatkowo, aby odfiltrować potencjalne multiplety (krople, w których znalazła się więcej niż jedna komórka) odrzucono z dalszej analizy komórki, w których liczba unikalnych genów, do których zostały zmapowane odczyty była powyżej 3000. W celu identyfikacji komórek obumierających (z nieszczelną błoną komórkową, przez którą część mRNA wypłynęła z cytoplazmy na zewnątrz, jeszcze przed zamknięciem komórki w kropli podczas procesu enkapsulacji) przeanalizowano procentową zawartość UMI (*ang. unique molecular identifier*) zmapowanych do genów kodujących białka mitochondrialne (Ryc. 6.1). Komórka była odrzucana z dalszej analizy, gdy wartość ta przekraczała 5%. Po zastosowaniu opisanego filtrowania w zbiorze danych pozostało 40401 znaczników, które na kolejnych etapach analizowano jako 40401 pojedynczych komórek. W analizowanych próbkach wartości w macierzy zliczeń zostały znormalizowane (w każdej komórce niezależnie) w odniesieniu do całkowitej liczby zliczeń przypisanych do tej komórki, korzystając z metody zaimplementowanej w pakiecie Seurat. (tj. normalizacja względem wielkości biblioteki). Tak uzyskane wartości były mnożone przez domyślny współczynnik skalujący (10000) i poddawane transformacji logarytmicznej ln(10000\*"znormalizowana wartość ekspresji"+1).



Ryc. 6.1 Wykresy rozrzutu ilustrujące liczbę odczytów przypisanych do poszczególnych znaczników (oś X) i procent tych odczytów, które zostały zmapowane do genów kodujących białka mitochondrialne (oś Y).

#### 6.1.2 Identyfikacja subpopulacji komórek wśród komórek CD11b+

W analizowanym zbiorze znalazły się komórki pochodzące z dwóch powtórzeń biologicznych dla każdej z czterech grup doświadczalnych. W celu zidentyfikowania i scharakteryzowania subpopulacji komórek występujących w poszczególnych grupach dane z powtórzeń biologicznych w obrębie grupy zostały zintegrowane zgodnie z procedurą dostępną w pakiecie Seurat. W obrębie każdej z grup wybrano 2000 genów o największej zmienności ekspresji, a także 2000 komórek będących wzajemnymi najbliższymi sąsiadami (*ang. mutual nearest neighbors*) wśród komórek pochodzących z dwóch powtórzeń. W uzyskanych w ten sposób zbiorach danych wyrugowano zmienność ekspresji genów pochodzącą z biologicznie nieinteresujących źródeł, tj. związaną z różną liczbą odczytów w poszczególnych komórkach, procentową zawartością UMI zmapowanych do genów kodujących białka mitochondrialne i fazą cyklu komórkowego. Oszacowanie fazy cyklu komórkowego, w której znajdowała się każda z komórek wykonano na podstawie ekspresji genów markerowych (Kowalczyk et al., 2015) i https://satijalab.org/seurat/articles/cell cycle vignette.

Dla każdej z grup doświadczalnych profil transkryptomiczny obejmował ponad 10 tysięcy genów. W związku z tym wykonano redukcję wymiarów za pomocą metody analizy składowych głównych (PCA, *ang. principal component analysis*) wykorzystując w dalszej analizie 30 pierwszych składowych głównych (PC). Wybór 30 pierwszych PC jest domyślnym podejściem sugerowanym w zastosowanym potoku analizy danych w Seurat. Ponadto wykorzystano też metodę JackStraw zaimplementowaną w Seurat do określenia liczby znaczących składowych głównych i porównano nowo uzyskane grupowanie z grupowaniem otrzymanym przy wyborze 30 składowych głównych. Nie stwierdzono znaczącej różnicy więc dla uproszczenia i spójności liczby składowych głównych pomiędzy grupami doświadczalnymi w prezentowanej analizie zachowano domyślną liczbę PC.

Następnie, korzystając z algorytmu grupowania, w sposób nienadzorowany podzielono komórki na subpopulacje i zwizualizowano je na przestrzeni dwuwymiarowej korzystając z metody t-SNE (Ryc. 6.2). Aby zidentyfikować rodzaj komórek w klastrach wybrano geny o statystycznie istotnie podwyższonej ekspresji. Geny o statystycznie istotnie zwiększonej ekspresji zidentyfikowano dla każdej z grup doświadczalnych za pomocą funkcji FindAllMarkers z pakietu Seurat.

## 6.1.3 Ocena wpływu rozdzielczości grupowania na liczbę grup komórek i przypisanie komórek do poszczególnych subpopulacji

Analizy danych scRNA-seq wymagają każdorazowo wyboru odpowiednich metod przetwarzania danych i doboru wartości parametrów tych metod. Nie jest to jednak zagadnienie typowo obliczeniowe, ponieważ głównym celem analiz jest biologiczna wartość wyników. Z tego względu brakuje jednoznacznych reguł do ustalenia optymalnych wartości parametrów i należy je wyznaczyć w każdym zestawie danych. Grupowanie komórek to jeden z najważniejszych etapów analizy, ponieważ jego wynik jest bazą do eksperckiej (opierającej się na aktualnej wiedzy biologicznej) charakterystyki populacji komórek obecnych w zbiorze. Dlatego tak ważne jest, aby dobrane wartości parametrów (dostosowane do analizowanego zbioru danych) pozwoliły uzyskać wynik przekładający się na biologicznie istotne podgrupy (*ang. clusters*).



Ryc. 6.2 Wynik grupowania komórek w zbiorach odpowiadających poszczególnych grupom doświadczalnym.

Przygotowując zbiór do szczegółowej charakterystyki przeanalizowano, jak wartość parametru rozdzielczości grupowania wpływała na liczbę grup komórek i przypisanie komórek do poszczególnych subpopulacji w każdej z grup doświadczalnych. Wskazane jest podejście iteracyjne, w którym parametry grupowania modyfikowane są na podstawie oceny wyników uwzględniającej wiedzę biologiczną. Żaden zestaw wartości parametrów nie gwarantuje jednak, że wszystkie biologicznie prawdziwe populacje komórek są właściwie rozdzielone. Jest to związane z heterogennością sygnatur transkryptomicznych w większości systemów biologicznych, a także technicznymi ograniczeniami wykorzystywanych metod. W przypadku gdy uzyskane podziały zostaną uznane (na podstawie wiedzy biologicznej) za nieistotne, uzasadnione może być połączenie klastrów na potrzeby identyfikacji wybranych typów komórek. Podział niektórych klastrów na podgrupy komórek może mieć źródło w zmienności pochodzenia technicznego, a nie biologicznego, co negatywnie wpływa na charakterystykę populacji obecnych w zbiorze – w takiej sytuacji również wskazane może być połączenie klastrów. W przypadku zbiorów danych zawierających dużą liczbę typów komórek zdarza się, że zwiększanie rozdzielczości grupowania (i kolejne podziały klastrów) ma uzasadnienie dla części populacji (np. tych najbardziej interesujących z punktu widzenia danej analizy), natomiast nie wnosi korzyści dla pozostałych, których identyfikacja na ogólnym poziomie będzie wystarczająca. W takiej sytuacji również uzasadnione może być zastosowanie większej wartości rozdzielczości, a następnie eksperckie połączenie części klastrów. W alternatywnym podejściu szczególnie interesujące populacje komórek można wydzielić ze zbioru do osobnej analizy i bardziej szczegółowej charakterystyki.

Efekty zmian rozdzielczości w analizowanych zbiorach zwizualizowano (wykorzystując pakiet clustree (Zappia & Oshlack, 2018)) na Ryc. 6.3 w formie grafu analogicznego do drzew filogenetycznych. Następnie zastosowano podejście iteracyjne i wybrano rozdzielczość 0.3, która pozwaliła zidentyfikować klastry, których charakterystyka była zgodna z wiedzą biologiczną. Po analizie genów, których ekspresja była statystycznie istotnie podwyższona w wyodrębnionych subpopulacjach komórek, zaproponowano, że subpopulacje reprezentują typy komórek przedstawione na Ryc 6.4. Do dalszych analiz wybrane zostały komórki zidentyfikowane jako mikroglej, monocyty/makrofagi i BAM.



Ryc. 6.3 Wyniki grupowania komórek w każdej z czterech grup doświadczalnych w zależności od wartości parametru rozdzielczości. Przedstawiono wpływ zmiany rozdzielczości na liczbę klastrów i zmianę przypisania komórek pomiędzy klastrami. Linia czerwona wskazuje wybraną w analizie rozdzielczość 0.3.



Ryc. 6.4 Zidentyfikowane populacje komórek w zbiorach odpowiadających poszczególnym grupom doświadczalnym (BAM, ang. CNS border associated macrophages)

Należy podkreślić, że w celu zminimalizowania wariancji związanej z przebiegiem procedur laboratoryjnych wszystkie biblioteki scRNA-seq zostały przygotowane razem. Próbki zsekwencjonowano w czterech seriach, na tym samym urządzeniu, aby uzyskać odpowiednią głębokość sekwencjonowania. W celu oceny czy uzyskane grupy nie są zdominowane przez pojedyncze powtórzenie porównano udział komórek z poszczególnych próbek w każdej z subpopulacji. Wyniki przedstawiono na Ryc. 6.5 i na tej podstawie stwierdzono, że analizowane dane w większości przypadków są spójne pomiędzy biologicznymi powtórzeniami, w szczególności w obrębie populacji mikrogleju, makrofagów i BAM, które zostały wybrane do szczegółowej analizy.



Ryc. 6.5 Procentowy udział komórek z obu powtórzeń w klastrach, w każdej grupie doświadczalnej. Szerokość każdego słupka odpowiada względnej wielkości klastra.

6.1.4 Analiza profili transkryptomicznych subpopulacji komórek mikrogleju, makrofagów i BAM

W niniejszej rozprawie skupiono się na zmianach profili transkrypcyjnych komórek mikrogleju, makrofagów i BAM pod wpływem obecności nowotworu. Na kolejnym etapie analizy macierze ekspresji wyselekcjonowanych z wszystkich próbek, z czterech grup doświadczalnych zostały połączone w jedną macierz ekspresji (zachowując informację o próbce, z której pochodziła każda komórka). Wartości ekspresji genów zostały znormalizowane w taki sam sposób jak w pierwszej części analizy, a następnie (na podstawie wyników transformacji stabilizującej wariancję "vst") wybrano 2000 genów o największej zmienności ekspresji pomiędzy wszystkimi komórkami w zbiorze. Dodatkowo wykonano skalowanie (standaryzację) wraz z wyrugowaniem zmienności mającej biologicznie nieinteresujące źródło (tj. związanej z efektem liczby UMI w komórce, efektem procentowej zawartości genów mitochondrialnych lub efektem oszacowanej

fazy cyklu komórkowego G1 vs S/G2). W kolejnym kroku, analogicznie do pierwszej części analizy, przeprowadzono analizę składowych głównych i wybrano 30 pierwszych składowych głównych. Redukcję wymiarów wykonano korzystając algorytmu UMAP (*ang. Uniform Manifold Approximation and Projection*). Następnie stosując iteracyjne podejście opisane w części 6.1.3 w pierwszej kolejności zidentyfikowano klastry mikrogleju, monocytów/makrofagów, a także komórek BAM (Ryc. 6.6). Na podstawie wyników grupowania uzyskanych przy wykorzystaniu różnych wartości parametru rozdzielczości (Ryc. 6.7) stwierdzono wysoką stabilność przypisania komórek do tych klastrów. Dodatkowo, co uznano za szczególnie interesujące, komórki mikrogleju już przy bardzo niskiej rozdzielczości (0.07) zostały podzielone na dwie główne populacje (Ryc. 6.8), które przy zwiększaniu rozdzielczości ulegały podziałom na subpopulacje. Na Ryc. 6.9 zaprezentowano mapę ciepła przedstawiającą geny o statystycznie istotnej i najbardziej podwyższonej ekspresji w tych czterech głównych grupach. Po zakończeniu oceny wyników grupowania, uwzględniającej geny o statystycznie istotnie podwyższonej ekspresji w poszczególnych klastrach i wiedzę biologiczną, do dalszych prac wybrano wartość parametru rozdzielczości równą 0.6 (na Ryc. 6.10) przedstawiono wynik grupowania dla tej rozdzielczości).



Ryc. 6.6 Identyfikacja klastrów komórek homeostatycznego mikrogleju, aktywowanego mikrogleju, monocytów/makrofagów i BAM w połączonym zbiorze danych.



Ryc. 6.7 Wyniki grupowania komórek mikrogleju, monocytów/makrofagów i BAM pochodzących ze wszystkich próbek i połączonych w jeden zbiór. Przedstawiono wpływ zmiany rozdzielczości na liczbę klastrów i zmianę przypisania komórek pomiędzy klastrami. Zaznaczono wybraną w analizie rozdzielczość 0.6.



Ryc. 6.8 Identyfikacja klastrów komórek homeostatycznego mikrogleju, aktywowanego mikrogleju, monocytów/makrofagów i BAM w połączonym zbiorze danych.



Ryc. 6.9 Mapa ciepła przedstawiająca ekspresję genów o statystycznie istotnie i najbardziej podwyższonej ekspresji w populacjach komórek mikrogleju, monocytów/makrofagów i BAM.



Ryc. 6.10 Wynik grupowania komórek mikrogleju, monocytów/makrofagów i BAM pochodzących ze wszystkich próbek i połączonych w jeden zbiór przy zastosowaniu wartości parametru rozdzielczości 0.6, przedstawiony na dwuwymiarowej przestrzeni UMAP.

W kolejnym kroku stwierdzono, że w przypadku części subpopulacji komórki pochodziły w zdecydowanej większości (lub w całości) z próbek z grupy z nowotworem. Na podstawie analizy genów, których ekspresja była statystycznie istotnie podwyższona w tych populacjach stwierdzono, że sytuacja ta występuje nie tylko w przypadku subpopulacji makrofagów napływowych z krwi obwodowej, które wraz z rozwojem nowotworu zostały zrekrutowane do mikrośrodowiska, ale również w przypadku komórek mikrogleju. W przypadku komórek BAM nie pojawiły się wyraźnie odmienne populacje wśród komórek CD11b+ z mózgu myszy naiwnych i z guzem. Na Ryc. 6.11 przedstawiono pochodzenie (grupę doświadczalną) każdej z komórek.



Ryc. 6.11 Wykres przedstawiający warunek doświadczalny z którego pochodziły komórki.

Przechodząc do bardziej szczegółowej analizy wykazano, że występuje dysproporcja w liczbie komórek pochodzących od samic/samców w części klastrów zdominowanych przez komórki z próbek z grupy z nowotworem. Wyniki te sugerowały obecność zależnych od płci różnic w mechanizmie odpowiedzi na obecność nowotworu. Wyniki te zostały potwierdzone w badaniach biochemicznych pokazujących wyższą ekspresję wybranych genów kodujących białka kompleksu MHCII w mikrogleju mysim pochodzącym od samców, wystawionym na działania czynników wydzielanych przez komórki glejaka. Szczegółowa charakterystyka biologiczna zidentyfikowanych subpopulacji oraz potwierdzenie wykrytych zmian transkryptomicznych nie jest przedmiotem niniejszej rozprawy, a szczegółowe wyniki zostały opublikowane w publikacji (Ochocka, Segit et al., 2021) i pracy doktorskiej dr Natalii Ochockiej-Lewickiej.

Podsumowując, skutecznie przetworzono, przygotowano i wykonano analizę porównawczą zbiorów danych scRNA-seq pochodzących z czterech grup doświadczalnych. Była to pierwsza opublikowana analiza, w której połączono i bezpośrednio porównano dane scRNAseq uzyskane dla komórek CD11b+ z mikrośrodowiska mysiego glejaka (a także z próbek kontrolnych) pochodzące zarówno od samców jak i od samic. W efekcie pozwoliło to wyodrębnić zróżnicowane profile transkrypcyjne w populacjach komórek charakteryzujących się ekspresją CD11b+ (w szczególności w subpopulacjach komórek mieloidalnych), które wcześniej nie były możliwe do odróżnienia (ze względu na obecność podobnych białek powierzchniowych i w związku z tym brakiem dostępnych strategii sortowania).

## 6.1.5 Analiza wpływu operacji pozorowanej na profile transkryptomiczne komórek CD11b+ w porównaniu z komórkami z mózgów myszy naiwnych.

W przedstawionych powyżej analizach profile transkrypcyjne komórek CD11b+ wysortowanych z mózgu z glejakiem porównano z profilami transkrypcyjnymi komórek CD11b+ wysortowanych z mózgów myszy naiwnych. Takie podejście nie uwzględniało faktu, że wszczepienie komórek glejaka odbywa się poprzez iniekcję komórek do parenchymy mózgu, co może powodować przejściową aktywację komórek mikrogleju. Z tego powodu przeprowadzono dodatkowe doświadczenie, w którym również wykorzystano technikę scRNA-seq i porównano profile transkryptomiczne komórek wyizolowanych w obu warunkach.

Macierze ekspresji dla próbek pochodzących od zwierząt naiwnych i próbek pochodzących od zwierząt poddanych operacji pozorowanej zostały połączone w jedną macierz ekspresji (z zachowaniem informacji o przypisaniu znacznika do konkretnej próbki). Połączony zbiór danych został następnie przetworzony w sposób analogiczny do wcześniej opisanych analiz.



Ryc. 6.12. Porównanie profili transkrypcyjnych w komórkach CD11b+ izolowanych z mózgu zwierząt naiwnych i z mózgu zwierząt poddanych operacji pozorowanej. (a) UMAP wykazuje równomierne rozmieszczenie komórek wyizolowanych z mózgu naiwnych i pozornie operowanych myszy. (b) Wyniki grupowania wykonanego na połączonym zbiorze danych. Wszystkie uzyskane klastry były obecne zarówno w komórkach CD11b+ z mózgu naiwnych myszy, jak i w komórkach pochodzących z mózgów zwierząt poddanych operacji pozorowanej (c) Odsetek komórek pochodzących od zwierząt naiwnych i zwierząt poddanych operacji pozorowanej był porównywalny we wszystkich uzyskanych klastrach. (d) Poziom ekspresji genów wykazujących zwiększoną ekspresję u zwierząt z wszczepionym nowotworem porównany w próbkach pochodzących od zwierząt naiwnych i w próbkach pochodzących od zwierząt poddanych operacji pozorowanej.

Wykonano redukcję wymiarów korzystając z metody UMAP i grupowanie komórek przy parametrze rozdzielczości o wartości 0.3. Uzyskane wyniki przedstawiono na Ryc. 6.12. Stwierdzono, że zawartość komórek pochodzących z obu warunków w poszczególnych klastrach jest na zbliżonym poziomie, co świadczy o wysokim podobieństwie profili transkryptomicznych subpopulacji komórek pochodzących z próbek od zwierząt naiwnych i z próbek od zwierząt poddanych operacji pozorowanej. Wszystkie zidentyfikowane klastry były obecne w obu grupach doświadczalnych. Dodatkowo wykonano porównanie pomiędzy warunkami ekspresji genów wskazujących na aktywację mikrogleju (geny kompleksu MHCII) i genów prozapalnych (Ccl12, Stat1, Ifitm23). Geny wybrano na podstawie wcześniej opisanych analiz wykonanych na połączonym zbiorze komórek mikrogleju, makrofagów i BAM pochodzących od zwierząt naiwnych i zwierząt z nowotworem. Porównanie ekspresji tych genów, zaprezentowane na Ryc. ##, nie wykazało istotnych różnic. Na podstawie otrzymanych wyników stwierdzono, że 14 dni po operacji pozorowanej nie wykryto istotnych różnic w profilach transkryptomicznych komórek CD11b+ pochodzących z mózgów zwierząt naiwnych i pochodzących z mózgów zwierząt 14 dni po operacji pozorowanej. W związku z tym, wykorzystanie w doświadczeniu zwierząt naiwnych było uzasadnione.

# 6.2 Analiza profili transkryptomicznych pojedynczych komórek CD11b+ pochodzących z mikrośrodowiska ludzkiego glejaka wielopostaciowego

W celu szczegółowej charakterystyki subpopulacji komórek CD11b+ w mikrośrodowisku ludzkiego glejaka wielopostaciowego zastosowano technikę sekwencjonowania RNA pojedynczych komórek pochodzących z próbek zbieranych śródoperacyjnie od pacjentów poddawanych resekcji guza. Uzyskanie wysokiej jakości danych scRNA-seq z próbek pochodzących od pacjentów jest zdecydowanie trudniejsze, m.in. ze względu na brak możliwości wykonania perfuzji. Dodatkowo, próbki te naturalnie charakteryzują się wyższą heterogennością niż próbki pozyskiwane w doświadczeniach w modelu mysim. Mając jednak na uwadze kluczową rolę mikrośrodowiska w patologii glejaka, a także to jak istotne są informacje pozyskane z danych od pacjentów, postanowiono zebrać próbki i przeanalizować profile transkryptomiczne zidentyfikowanych subpopulacji komórek.

#### 6.2.1 Przygotowanie i normalizacja danych uzyskanych z scRNA-seq

Próbki guza zostały pobrane od 5 pacjentów (4 mężczyzn i 1 kobiety) ze zdiagnozowanym glejakiem wielopostaciowym G4 (IDH prawidłowy). W każdym z pięciu przypadków, fragment guza został wyodrębniony podczas operacji resekcji guza, a następnie po przetransportowaniu do laboratorium przeprowadzono dysocjację tkanki. Po dysocjacji, korzystając z cytometru przepływowego, wysortowano populację komórek żywych, a z niej populację żywych komórek

CD11b+. Z tak wyselekcjonowanej populacji uzyskano dane scRNA-seq zgodnie z protokołem 10x Genomics Single Cell Gene Expression 3' v2 (https://www.10xgenomics.com/support/single-cellgene-expression), wykorzystując urządzenie mikroprzepływowe Chromium Controller (https://www.10xgenomics.com/instruments/chromium-controller). Parametry próbek umieszczanych na płytkach urządzenia zostały dobrane w taki sposób, aby uzyskać dane z 5000 komórek dla każdej z próbek. W kolejnym kroku utworzono biblioteki do sekwencjonowania zgodne z wytycznymi 10x Genomics (https://www.10xgenomics.com/support/single-cell-geneexpression/documentation/steps/sequencing/sequencing-requirements-for-single-cell-3) i NGS 1500 wykonano sekwencjonowanie na urządzeniu Illumina HiSeq (https://support.illumina.com/sequencing/sequencing instruments/hiseq 1500.html). Szczegółowy opis procedur doświadczalnych znajduje się w rozprawie dr. Kacpra Walentynowicza (2019).

Surowe dane będące wynikiem sekwencjonowania (pliki BCL) zostały demultipleksowane za pomocą oprogramowania CellRanger (v3.0.1) mkfastą (https://support.10xgenomics.com/single-cell-gene-

expression/software/pipelines/latest/using/mkfastq) i bcl2fastq v2.20.0.422 (https://support.illumina.com/sequencing/sequencing\_software/bcl2fastq-conversionsoftware/downloads.html). Po demultipleksowaniu dane zostały zapisane w formacie FASTQ (https://knowledge.illumina.com/software/general/software-general-reference\_materiallist/000002211). Następnie otrzymane odczyty zmapowano do ludzkiego transkryptomu GRCh38 (Ensembl 98) (https://www.10xgenomics.com/support/software/cell-ranger/downloads) i zliczono liczbę odczytów zmapowanych do poszczególnych genów dla poszczególnych znaczników (ang. barcode). W ten sposób dla każdej próbki otrzymano dane do macierzy ekspresji genów przypisanych do pojedynczych znaczników. Algorytm oprogramowania CellRanger zidentyfikował 24486 znaczniki, które z wysokim prawdopodobieństwem znalazły się w kropli z żywą komórką. Dla tych znaczników uzyskano średnio 51727 odczytów na komórkę

(Tabela 6.2)

Tabela 6.2 Liczba zidentyfikowanych komórek i średnia liczba odczytów na komórkę w poszczególnych próbkach (dane przetworzone i wstępnie przefiltrowane oprogramowaniem CellRanger)

	Próbka	Próbka	Próbka	Próbka	Próbka	Średnia	Suma
	hg-gam-	hg-gam-	hg-gam-	hg-gam-	hg-gam-		
	01	02	03	04	05		
Liczba	3533	6604	3077	5926	5346	4897	24486
zidentyfikowanych							
komórek							
Średnia liczba	105404	30937	58602	26625	37065	51727	
odczytów na							
komórkę							

Jakość materiału po enkapsulacji (w każdej z próbek) oceniono także na podstawie liczby UMI przypisanych do poszczególnych znaczników. Na wykresach na Ryc. 6.13 na osi OX zostały uszeregowane znaczniki, od tego o największej liczbie przypisanych UMI, do tego o najmniejszej liczbie przypisanych UMI (skala logarytmiczna). Na osi OY znajduje się liczba UMI przypisanych do znacznika (skala logarytmiczna). W przypadku gdy jakość materiału po enkapsulacji jest prawidłowa występuje wyraźna różnica pomiędzy liczbą UMI przypisanych do znaczników, które po enkapsulacji znalazły się w kroplach z żywymi komórkami, a liczbą UMI przypisanych do znaczników, które po enkapsulacji najprawdopodobniej trafiły do kropel zwierających wolno pływające w zawiesinie cząsteczki mRNA (źródłem tych cząsteczek były umierające komórki). Na wykresie można w takiej sytuacji zaobserwować załamanie krzywej (kolano) rozgraniczające te dwie grupy znaczników. W przypadku wszystkich pięciu próbek wykresy te sugerowały, że jakość materiału po enkapsulacji umożliwiła kontynuowanie analiz.



Ryc. 6.13 Liczba UMI przypisana do poszczególnych znaczników (znaczniki posortowane zgodnie z malejącą liczbą przypisanych UMI).

Analiza profili transkryptomicznych została przeprowadzona w środowisku R (https://www.r-project.org/) wykorzystując funkcje pakietu Seurat v3 (Stuart et al., 2019). Kontrolę jakości rozpoczęto od sprawdzenia liczby odczytów przypisanych do poszczególnych znaczników, aby zidentyfikować i odfiltrować dane pochodzące z kropli najprawdopodobniej zawierających jedynie wolne mRNA, a nie zawierających żywych komórek. Na podstawie wykonanych analiz, przyjęto, że znaczniki dla których liczba zidentyfikowanych genów była poniżej 200 najprawdopodobniej nie reprezentują żywych komórek i odrzucono je ze zbioru (próg odcięcia sugerowany w podejściu autorów pakietu Seurat). Wykresy rozrzutu dla poszczególnych znaczników przedstawiono na Ryc. 6.14





W kolejnym kroku sprawdzono jaką część wszystkich genów, do których zmapowano odczyty przypisane do danego znacznika, stanowiły geny mitochondrialne. Podobnie jak w przypadku analiz poprzednich zbiorów danych, miało to na celu identyfikację komórek obumierających, z nieszczelną błoną komórkową, przez którą część mRNA wypłynęła z cytoplazmy przed enkapsulacją komórki. Wykresy rozrzutu dla poszczególnych znaczników przedstawiono na Ryc. 6.15. W przypadku analizy zbiorów komórek mysich odfiltrowano znaczniki, dla których ponad 5% odczytów zostało zmapowanych do genów mitochondrialnych, natomiast w przypadku tego zbioru takie podejście skutkowałoby usunięciem ze zbioru nawet

15% wszystkich znaczników (hg\_gam\_03). W przypadku komórek ludzkich tworzących wiele różnych tkanek procent odczytów zmapowanych do genów mitochondrialnych jest istotnie wyższy niż 5%, a najwyższy został zaobserwowany w makrofagach pochodzenia monocytarnego (mediana powyżej 10%) (Osorio and Cai, 2021). W związku z tym zastosowano mniej restrykcyjne podejście i odfiltrowano znaczniki, dla których ponad 20% odczytów zostało zmapowanych do genów mitochondrialnych i w ten sposób zmniejszyć ryzyko wykluczenia ze zbioru żywych komórek. Takie podejście wymaga jednak, aby na dalszych etapach analizy sprawdzić, do których klastrów/populacji komórek zostaną przyporządkowane komórki, dla których ten procent jest na wysokim poziomie.



Ryc. 6.15. Wykresy rozrzutu ilustrujące liczbę odczytów przypisanych do poszczególnych znaczników (oś X) i procent tych odczytów, które zostały zmapowane do genów mitochondrialnych (oś Y).

W kontroli jakości wykonano również ramach ocene złożoności profili transkryptomicznych znaczników/komórek obliczając dla każdego znacznika zlogarytmowaną liczbę unikalnych genów, do których zostały zmapowane odczyty podzieloną przez zlogarytmowaną liczbę UMI przypisanych do tego znacznika. Parametr ten służy do oceny uzyskanej głębokości sekwencjonowania i wykrywania ewentualnego zanieczyszczenia próbki komórkami o niskiej złożoności profili transkryptomicznych, np. czerwonymi krwinkami. Wykresy rozrzutu dla poszczególnych znaczników przedstawiono na Ryc. 6.16. Zgodnie z zaleceniami, ze zbioru odfiltrowano znaczniki, dla których wartość tego parametru była niższa niż 0.8 (https://hbctraining.github.io/scRNA-seq/lessons/04 SC quality control.html).



Ryc. 6.16 Wykresy rozrzutu ilustrujące dla pojedynczych znaczników zlogarytmowaną liczbę unikalnych genów, do których zostały zmapowane odczyty podzieloną przez zlogarytmowaną liczbę UMI przypisanych do tego znacznika.

W efekcie zastosowania opisanych kryteriów filtrowania w zbiorze pozostało 24057 znaczników, które na kolejnych etapach analizy potraktowano jako 24057 pojedynczych komórek.

Dla każdej z próbek, po filtrowaniu, otrzymano macierz ekspresji zawierającą zliczenia odczytów zmapowanych do poszczególnych genów w pojedynczych komórkach. Każda z tych macierzy została niezależnie znormalizowana uwzględniając całkowitą liczbę zliczeń w poszczególnych komórkach. Następnie każdą z macierzy przemnożono przez współczynnik skalujący, a wartości zlogarytmowano. W celu oceny czy przed kontynuacją analiz wskazane będzie wyrugowanie efektu cyklu komórkowego z profili transkryptomicznych pojedynczych komórek oszacowano fazę cyklu komórkowego, w której znajdowała się każda z komórek (na podstawie ekspresji genów markerowych cyklu komórkowego (Kowalczyk *et al.*, 2015) i https://satijalab.org/seurat/articles/cell\_cycle\_vignette). Faza cyklu komórkowego może determinować wyniki grupowania komórek, a jednocześnie być biologicznie nieinteresującym źródłem wariancji ekspresji genów. Gdy taka sytuacja ma miejsce, identyfikacja poszczególnych populacji jest utrudniona, ponieważ komórki w tej samej fazie cyklu komórkowego grupują się razem pomimo tego, że są komórkami różnego typu. W analizowanym zbiorze danych takiej sytuacji nie zaobserwowano i w związku z tym na kolejnych etapach nie wyrugowano efektu cyklu komórkowego.

#### 6.2.2 Integracja zbiorów danych z niezależnych próbek GBM.

Analizowany zbiór danych zawierał po jednej macierzy ekspresji dla każdej z pięciu próbek. Na potrzeby dalszych analiz, w szczególności identyfikacji i charakterystyki typów/stanów komórek postanowiono stworzyć jeden zbiór zawierający komórki z wszystkich próbek. Można było to osiągnąć stosując jedno z dwóch podejść: łącząc macierze ekspresji bez wykonywania transformacji danych wejściowych albo zastosować metodę integracji danych mającej na celu wyrugowanie/minimalizację efektu próbki.

W pierwszej kolejności komórki ze wszystkich próbek zostały połączone w jedną macierz ekspresji (zachowując informację o próbce, z której pochodziła każda komórka). Wartości ekspresji genów dla poszczególnych komórek zostały znormalizowane w taki sam sposób jak podczas oceny efektu cyklu komórkowego. Następnie na podstawie wyników transformacji stabilizującej wariancję ("vst") wybrano 2000 genów o największej zmienności ekspresji pomiędzy komórkami. Dla tych genów wykonano dodatkowo skalowanie (standaryzację) wraz z wyrugowaniem zmienności ekspresji mającej biologicznie nieinteresujące źródło (tj. związanej z efektem liczby UMI w komórce lub z efektem procentowej zawartości transkryptów genów mitochondrialnych). Na tak przygotowanym zbiorze danych przeprowadzono analizę składowych głównych (w efekcie ograniczając zbiór do pierwszych 30 składowych głównych) i redukcję wymiarów algorytmem UMAP (na domyślnych wartościach parametrów metody RunUMAP). Następnie korzystając z algorytmu grupowania w zaimplementowanego w metodzie FindClusters (rozdzielczość 0.1) w sposób nienadzorowany podzielono komórki w połączonym zbiorze danych na klastry i zaprezentowano wyniki w przestrzeni dwuwymiarowej (UMAP-1; UMAP-2). Dodatkowo w tej samej przestrzeni zwizualizowano informację, o próbce z której pochodzi każda z komórek (Ryc. 6.17). Otrzymane wyniki wskazywały, że większość klastrów zawiera komórki pochodzące w zdecydowanej większości z jednej próbki. Sugerowało to, że wpływ efektu próbki na profile transkryptomiczne komórek jest silniejszy niż różnice w ekspresji genów wynikające z różnic pomiędzy typami/stanami komórek. Jest to najczęściej sytuacja niepożądana, utrudniająca identyfikację populacji komórek i ich opis, a także selekcję genów o istotnie zróżnicowanej ekspresji pomiędzy typami komórek. W związku z tym wykorzystano procedurę integracji (standardowej) próbek, dostępną w pakiecie Seurat.



Ryc. 6.17. Wyniki grupowania komórek w połączonym zbiorze danych (wszystkie 5 próbek) i informacja o próbce, z której pochodzą poszczególne komórki.

Przygotowując dane do integracji wybrano 2000 genów o największej zmienności ekspresji, a także 2000 komórek będących wzajemnymi najbliższymi sąsiadami (ang. mutual nearest neighbors) dla analizowanych pięciu próbek. Następnie zintegrowano zbiór metodą IntegrateData, powtórzono analizę składowych głównych, redukcję wymiarów algorytmem UMAP i grupowanie komórek (rozdzielczość 0.1). Wyniki przedstawiono na przestrzeni dwuwymiarowej (UMAP-1, UMAP-2) na Ryc. 6.18.

W przestrzeni (UMAP-1, UMAP-2) komórki zlokalizowały się w trzech obszarach: przypisane do klastra 3, przypisane do klastra 7 i przypisane do pozostałych klastrów. W ramach ogólnej charakterystyki zbioru zmniejszono rozdzielczość w algorytmie grupowania do 0.01 i w efekcie uzyskano trzy klastry (Ryc 6.19). Po analizie genów o statystycznie istotnie zwiększonej ekspresji w każdym z klastrów stwierdzono, że w klastrze 1 znalazły się komórki prekursorowe dla oligodendrocytów, a w klastrze 2 znalazły się komórki NK/T. Obie te populacje nie były w zakresie zaplanowanej analizy, w związku z czym usunięto je ze zbioru (pozostało 22353 komórek). Wystandaryzowaną ekspresję genów o statystycznie istotnej i najbardziej zwiększonej ekspresji w tych trzech klastrach przedstawiono na mapie ciepła Ryc. 6.20. Następnie powtórzono procedurę integracji próbek. W celu wyboru liczby składowych głównych, wykorzystywanych na dalszych etapach analizy, użyto metody JackStraw zaimplementowaną w pakiecie Seurat i na podstawie uzyskanych wyników zdecydowano, żeby pozostawić w zbiorze 40 składowych głównych. Redukcję wymiarów przeprowadzono ponownie algorytmem UMAP do przestrzeni (UMAP-1; UMAP-2). Wynik, wraz z informacją o próbce, z której pochodzi każda z komórek, przedstawiono na Ryc. 6.21.



Ryc. 6.18 Wyniki wstępnego grupowania komórek w zintegrowanym zbiorze danych (wszystkie 5 próbek) i informacja o próbce z której pochodzą poszczególne komórki.



Ryc. 6.19 Wynik grupowania komórek przy rozdzielczości 0.01.



Ryc. 6.20 Mapa ciepła wystandaryzowanej ekspresji genów o najbardziej zwiększonej ekspresji w poszczególnych klastrach.



Ryc. 6.21 Komórki zintegrowanego zbioru danych (po filtrowaniu) przedstawione na dwuwymiarowej przestrzeni (UMAP-1; UMAP-2) wraz z informacją o próbce, z której pochodzi każda komórka.

# 6.2.3 Ocena wpływu rozdzielczości grupowania na liczbę grup komórek i przypisanie komórek do poszczególnych subpopulacji

Analogicznie jak w przypadku analizy komórek CD11b+ pochodzących z mikrośrodowiska mysiego glejaka GL261, dokonano oceny jak zmiana wartości parametru rozdzielczości w algorytmie grupowania wpływa na uzyskiwaną liczbę klastrów i stabilność podziałów. Efekt zmian rozdzielczości przedstawiono (korzystając z pakietu clustree (Zappia & Oshlack, 2018)) na Ryc.6.22. W przypadku grupowania komórek w zbiorach scRNA-seq priorytetem jest właściwa interpretacja biologiczna uzyskanych populacji. Grupowanie, jako metoda nienadzorowana, podczas przypisywania komórek do klastrów nie uwzględnia zewnętrznych informacji (np. o znanych markerach komórek określonego typu) analizując jedynie profile transkryptomiczne pojedynczych komórek. Mając to na uwadze, zastosowano iteracyjne podejście do oceny wyników grupowania i weryfikacji uzyskanych klastrów, opierającej się na wiedzy biologicznej. Wyżej opisane analizy tego zbioru miały przede wszystkim na celu ocene jego jakości, wybór komórek będących w zakresie analizy i przygotowanie danych do dalszych prac. Na Ryc. 6.22 można zaobserwować stabilność przypisania komórek do klastrów – od rozdzielczości 0.1 zwiększanie liczby klastrów w większości przypadków związane jest z podziałem istniejącego klastra, ale nie powoduje przemieszczania się komórek pomiędzy klastrami z wyższego poziomu. Na potrzeby charakterystyki zbioru porównywano wyniki grupowania otrzymane dla rozdzielczości (0.2, 0.3, 0.4, 0.5) i wybrano rozdzielczość 0.4. Wynik grupowania przedstawiono na Ryc. 6.23. Identyfikacja typów/stanów komórek została przeprowadzona przez dr. Kacpra Walentynowicza, we współpracy z autorem niniejszej rozprawy.

## 6.2.4 Identyfikacja wybranych subpopulacji komórek CD11b+ w mikrośrodowisku glejaka wielopostaciowego.

W pierwszym kroku, na podstawie ekspresji wybranych genów markerowych, stwierdzono, że w zbiorze znajdują się monocyty, mikroglej/makrofagi i grupa komórek proliferujących (Ryc. 6.24). Korzystając z wartości sygnatury mikrogleju i sygnatury makrofagów (zestawu genów o wysokiej i specyficznej ekspresji w określonym typie komórek) nie zidentyfikowano klastrów zdominowanych przez komórki jednego z tych dwóch typów (na Ryc. 6.25 przedstawiono na wykresach skrzypcowych rozkład wartości obu sygnatur). W celu scharakteryzowania subpopulacji przeanalizowano geny o statystycznie istotnej i najbardziej podwyższonej ekspresji w poszczególnych klastrach. Na Ryc. 6.26 przedstawiono wystandaryzowaną ekspresję tych genów w formie mapy ciepła, a na Ryc. 6.27 zidentyfikowane typy/stany komórek. Klaster o największej liczbie komórek (0) charakteryzował się podwyższoną ekspresją genów kompleksu MHCII (HLA-DPB1, HLA-DQA1, HLA-DRA, HLA-DPA1), w związku z czym opisano go jako komórki prezentujące antygen (APC). Na Ryc. 6.28 przedstawiono dodatkowo wartości sygnatury prezentowania antygenów w całym zbiorze.



Ryc. 6.22 Graf ilustrujący wpływ zmiany wartości parametru rozdzielczości w metodzie grupowania na liczbę klastrów, na które został podzielony zbiór komórek. Krawędzie grafu przedstawiają sposób przemieszczania komórek między kolejnymi wariantami grupowania. Ramką oznaczono wybraną rozdzielczość 0.4.

Drugi największy klaster, nr 1, został opisany jako zróżnicowane komórki o profilu prozapalnym. Wyróżniał się podwyższoną ekspresją przede wszystkim chemokin CCL3, CCL3L1,

CCL4, CCL4L2 powiązanych z odpowiedzią zapalną (Lin et al., 2018), a także genów kodujących czynniki transkrypcyjne EGR1 i EGR2 (McCowan et al., 2021). Wartości sygnatury przedstawiono na Ryc. 6.29.Klaster 2 charakteryzował się podwyższoną ekspresją genów kodujących białka PLIN2 (ang. Perilipin 2), TIMP1 (ang. TIMP metallopeptidase inhibitor 1), VIM (ang. Vimentin) i BNIP3 (ang. BCL2 Interacting Protein 3) co sugerowało, że komórki znalazły się w obszarze występowania hipoksji (Guo et al., 2001). Wartości sygnatury przedstawiono na Ryc. 6.32. Kolejny pod względem wielkości, klaster (3), został opisany jako komórki dendrytyczne o stłumionej odpowiedzi (fenotyp M2), wykazujące m.in. podwyższoną ekspresję genu kodującego białko TMEM176B (ang. Transmembrane Protein 176B) (Segovia et al., 2019). Wartości sygnatury immunosupresyjnej przedstawiono na Ryc. 6.30. W klastrze (4) geny o statystycznie istotnej i najbardziej podwyższonej ekspresji, to geny kodujące rybosomalne RNA. Następne dwa klastry (5) i (6) zostały opisane jako klastry komórek o zdolnościach fagocytarnych, z tą różnicą, że komórki w klastrze (5) charakteryzowały się dodatkowo ekspresją genów powiązanych ze zdolnością przetwarzania lipidów, o podwyższonej ekspresji genów kodujących białka GPNMB, (ang. Glycoprotein-NMB), LIPA (ang. Lipase A), CD9 i TREM2 (ang. Triggering receptor expressed on myeloid cells 2) (Remmerie, Martens and Scott, 2020).



Ryc. 6.23. Wynik grupowania komórek (wartość parametru rozdzielczości=0.4).



Ryc. 6.24. Wynik identyfikacji głównych grup komórek.



Ryc. 6.25. Wykresy skrzypcowe ilustrujące rozkłady wartości sygnatury mikrogleju i makrofagów w poszczególnych klastrach.



Ryc. 6.26. Mapa ciepła ilustrująca względną ekspresję (z-score) genów o najbardziej podwyższonej ekspresji w poszczególnych klastrach. Przedstawiono nazwy genów wybrane jako istotne do identyfikacji typu/stanu komórki.



Ryc. 6.27. Wynik identyfikacji typów/stanów komórek przypisanych do poszczególnych klastrów

Do klastra (7) zostały przypisane komórki o podwyższonej ekspresji genów kodujących białka mitochondrialne. Pamiętając o zastosowanym na etapie filtrowania mniej restrykcyjnym podejściu (odfiltrowano znaczniki, dla których ponad 20% odczytów zostało zmapowanych do genów mitochondrialnych) stwierdzono, że są to komórki umierające. Komórki przypisane do kolejnego pod względem wielkości klastra (8) charakteryzowały się podwyższoną ekspresją genów białek szoku cieplnego (*ang. heat shock proteins*), co może sugerować że były to komórki narażone na . Do klastra (9) zostały natomiast przypisane komórki zidentyfikowane jako monocyty (o podwyższonej ekspresji genów kodujących białka S100A8, S100A9, FCN1 (Averill *et al.*, 2011) i LYZ, VCAN, S100A12). Komórki proliferujące zostały przypisane do klastrów (10) i (11).



Ryc. 6.28. Wartości sygnatury prezentowania antygenów w pojedynczych komórkach



Ryc. 6.29. Wartości sygnatury prozapalnej w pojedynczych komórkach

Klaster 10 charakteryzuje się podwyższoną ekspresją genów powiązanych z fazą S cyklu komórkowego (PCNA, RPA3, CDK1, DUT), a z kolei w klastrze 11 komórki wykazywały podwyższoną ekspresję genów powiązanych z fazą G2M (KPNA2, HMGN2, CDC20, CDKN3). Do następnego pod względem wielkości klastra 12 przypisane zostały komórki o podwyższonej ekspresji genów kompleksu MHCII, a także genów kodujących NFKBIA (ang. NFKB Inhibitor Alpha) i CXCR4 (ang. C-X-C chemokine receptor type 4). Na podstawie wyników analizy wpływu rozdzielczości grupowania na liczbę klastrów zauważono, że klaster ten został wyodrębniony jako subpopulacja komórek opisanych jako APC.



Ryc. 6.30. Wartości sygnatury immunosupresyjnej w pojedynczych komórkach

Komórki przypisane do klastra 13 charakteryzowały się podwyższoną ekspresją genów indukowanych interferonem (ISG15, IFIT1, IFIT3, IFITM3). Wartości sygnatury przedstawiono na Ryc. 3.31. Najmniejszy ze zidentyfikowanych klastrów, klaster 14 został opisany jako homeostatyczne makrofagi (z podwyższoną ekspresją genów kodujących białka LYVE1, STAB1, SELENOP, CD163, MRC1 (Weinberger *et al.*, 2020)).

Podsumowując, skutecznie przetworzono, przygotowano i wykonano analizę profili transkryptomicznych komórek CD11b+ z mikrośrodowiska glejaka wielopostaciowego pochodzących od pięciu pacjentów. Umożliwiło to podział komórek mikrośrodowiska na subpopulacje i ich charakterystykę.



Ryc. 6.31. Wartości sygnatury indukowanej interferonem w pojedynczych komórkach



Ryc. 6.32. Wartości sygnatury hipoksji w pojedynczych komórkach



Ryc. 6.33. Wartości sygnatury proliferacji w pojedynczych komórkach

### 7 DYSKUSJA

#### 7.1 Identyfikacja populacji komórek CD11b+ w mikrośrodowisku mysiego glejaka GL261

W wyniku przeprowadzonych analiz bioinformatycznych danych scRNA-seq z komórek CD11b+ wysortowanych z prawidłowego mózgu oraz w mózgu z glejakiem wykazano, że komórki CD11b+ zarówno w prawidłowym mózgu jak i w mikrośrodowisku mysiego glejaka GL261 nie są jednorodną populacją. Stwierdzono znaczną heterogenność ich profili transkryptomicznych wskazującą na infiltrację komórek z krwi oraz zmiany fenotypu komórek mikrogleju zachodzące w środowisku guza. Przeanalizowano także dane scRNA-seq komórek CD11b+ w celu charakterystyki i porównania populacji w próbkach pobranych od zwierząt naiwnych i od zwierząt po operacji pozorowanej, aby stwierdzić, czy sama procedura implantacji nie wywołuje długotrwałych zmian w składzie i funkcjonalności mikrogleju.

W ramach opisanych w niniejszej pracy analiz wykonano porównanie profili transkryptomicznych pojedynczych komórek CD11b+ z próbek pobranych z mózgów myszy naiwnych i z próbek pobranych z mózgów myszy, u których wykonano operację pozorowaną, zgodną z procedurą implantacji komórek GL261. Dane uzyskane ze wszystkich próbek zostały przeanalizowane łącznie, bez przeprowadzania procedury integracji danych, aby zachować w zbiorze do dalszej analizy jak najwięcej ewentualnych różnic występujących pomiędzy próbkami. Na podstawie wyników grupowania stwierdzono, że proporcje komórek pochodzących z obu warunków były podobne. Potwierdzono również, że wszystkie zidentyfikowane grupy komórek były obecne w próbkach pobranych z mózgów myszy naiwnych i z próbek pobranych z mózgów myszy, które poddano operacji pozorowanej. Dodatkowo porównano ekspresję genów wskazujących na aktywację mikrogleju (geny kodujące białka kompleksu MHCII) i genów prozapalnych (*Ccl12, Stat1, lfitm3*) w komórkach pochodzących z próbek z obu grup. W przypadku porównania ekspresji tych genów w grupach komórek z próbek pobranych od zwierząt naiwnych i zwierząt poddanych operacji pozorowanej nie wykazano istotnych różnic. Wyniki te wskazują, że sama operacja nie zmienia proporcji i funkcjonalności komórek w 14 dni po zabiegu, gdy prowadzone są badania mikrośrodowiska. Na podstawie wykonanych analiz stwierdzono, że wykorzystanie w doświadczeniu zwierząt naiwnych było uzasadnione.
W doświadczeniu na zwierzętach z implantacją komórek glejaka GL261 zbadano 4 grupy: samce i samice naiwne oraz samce i samice, którym zaimplantowano komórki GL261 (w każdej grupie były dwa powtórzenia biologiczne). Dane uzyskane ze wszystkich próbek przeanalizowano w dwóch etapach: 1) integrując dane pochodzące z powtórzeń biologicznych w obrębie każdej z czterech grup 2) przyporządkowując komórki w sposób nienadzorowany do grup. Celem było uzyskanie takiego podziału, w którym będzie można wiarygodnie opisać dominujący fenotyp komórek, korzystając z informacji o genach o statystycznie istotnie podwyższonej ekspresji w danej grupie i aktualnej wiedzy biologicznej. Dodatkowo istotne było to, aby podział na grupy był możliwie stabilny, tzn. aby wraz ze zwiększaniem wartości parametru rozdzielczości komórki dzieliły się na podgrupy, a nie były przypisywane do zupełnie innych grup. Na tym etapie stwierdzono, jak istotne było wybranie odpowiedniej wartości parametru rozdzielczości grupowania, który wpływa na liczbę grup, do których przypisane są komórki. Z danych z każdej z czterech zbadanych grup doświadczalnych wybrano do dalszej analizy komórki mikrogleju i BAM, a także monocyty/makrofagi w przypadku danych z dwóch grup, w których zaimplantowano komórki GL261. Warto zauważyć, że procentowy udział komórek z obu powtórzeń biologicznych był bardzo zbliżony w grupach obejmujących te typy komórek, co świadczy o powtarzalności zaobserwowanych efektów i zwiększa wiarygodność wyników uzyskanych w przeprowadzonych analizach.

Kolejne analizy obejmowały wyłącznie komórki mikrogleju, monocytów/makrofagów i BAM pochodzące ze wszystkich próbek, ze wszystkich zbadanych grup (analizy zostały przeprowadzone na jednym, połączonym zbiorze danych). Już przy bardzo niskiej wartości parametru rozdzielczości (równym 0.04) komórki zostały przypisane do grup zidentyfikowanych jako mikroglej, monocyty/makrofagi i BAM. Sugeruje to występowanie wyraźnych różnic pomiędzy profilami transkryptomicznymi komórek przypisanych do tych grup. Co okazało się szczególnie ciekawe, to przy niewiele wyższej wartości parametru rozdzielczości (równym 0.07) grupa komórek mikrogleju podzieliła się na dwie główne podgrupy (zidentyfikowane na dalszych etapach analiz jako homeostatyczny mikroglej i aktywowany mikroglej). W kolejnych iteracjach grupowania, wraz ze zwiększaniem wartości parametru rozdzielczości, tworzyły się dalsze podgrupy w obrębie homeostatycznego mikrogleju i aktywowanego mikrogleju. Sugeruje to, że na poziomie transkryptomicznym komórki przypisane do tych dwóch głównych grup mikrogleju wyraźnie się różnią, a obserwowane podziały tworzą subpopulacje homeostatycznego mikrogleju i aktywowanego mikrogleju. Po sprawdzeniu z jakich próbek pochodzą komórki przypisane do poszczególnych grup stwierdzono, że w grupach zidentyfikowanych jako homeostatyczny mikroglej i BAM znajdowały się komórki ze wszystkich próbek, a w grupach zidentyfikowanych jako aktywowany mikroglej i makrofagi niemal wyłącznie komórki z próbek pobranych od zwierząt, którym zaimplantowano komórki GL261. W związku z tym przyjęto, że komórki przypisane do tych dwóch ostatnich grup, to komórki o fenotypie charakterystycznym dla mikrośrodowiska mysiego glejaka GL261.

Na podstawie wyników kolejnych iteracji grupowania wybrano do dalszych analiz grupy komórek uzyskane przy rozdzielczości o wartości 0.6. Przeanalizowano z jakich próbek pochodziły komórki przypisane do poszczególnych grup i wykazano, że w niektórych grupach występuje dysproporcja w liczbie komórek pochodzących od samic/samców. W grupach opisanych jako homeostatyczny mikroglej i komórki BAM nie wykryto wyraźnych różnic pomiędzy płciami. Natomiast, co szczególnie interesujące, takie różnice zaobserwowano w populacji aktywowanego mikrogleju i jednej z subpopulacji makrofagów (opisanej jako populacja przejściowa monocytarno/makrofagowa). Sugeruje to występowanie różnic zależnych od płci w mechanizmie odpowiedzi na obecność nowotworu. W przypadku obu tych grup wykryto, że poziom ekspresji genów kodujących białka MHCII (m.in. H2-Ab1, H2-Eb1, H2-Aa), a także genu kodującego Cd74 był istotnie wyższy w komórkach pochodzących z próbek pobranych od samców, u których zostały zaimplantowane komórki GL261. Wyniki te zostały dodatkowo potwierdzone w badaniach biochemicznych, w których mikroglej mysi pochodzący od samców został wystawiony na działanie czynników wydzielanych przez komórki glejaka. Różnice zależne od płci w mechanizmie odpowiedzi na glejaka są tematem wartym dalszych badań m.in. ze względu na różnice w częstotliwości występowania glejaka wielopostaciowego u ludzi (1.6:1 u mężczyzn w porównaniu do kobiet (Khan et al., 2021)), a także z uwagi na to, że w przypadku meta-analizy obejmującej różne typy nowotworów stwierdzono lepszą odpowiedź na immunoterapię u mężczyzn niż u kobiet (Conforti et al., 2018). Dalsze badania przeprowadzone były w Pracowni Neurobiologii Molekularnej z użyciem scRNA-seq i przeciwciał (CITE-seq), co pozwoliło lepiej definiować subpopulacje, wykorzystując również informacje o obecności wybranych białek powierzchniowych. Badania te pokazały, że u samców zwiększona jest proporcja immunosupresyjnych makrofagów w guzach, a u samic jest wyższa proporcja monocytów z wysoką ekspresją genów indukowanych interferonem. Badania immunocytochemiczne pokazały znacząco wyższy procent aktywnych komórek mieloidalnych (HLA immunoreaktywnych) u mężczyzn z GBM niż u kobiet (Ochocka *et al.,* 2023). Wyniki te wskazują na odmienny, zależny od płci, przebieg akumulacji i aktywacji komórek mieloidalnych u osób z GBM, co może tłumaczyć zarówno inną częstotliwość występowania GBM, jak też różnice w odpowiedzi na terapie.

# 7.2 Zróżnicowanie populacji komórek CD11b+ w mikrośrodowisku ludzkiego glejaka wielopostaciowego

W celu charakterystyki subpopulacji CD11b+ w próbkach ludzkich GBM zebranych śródoperacyjnie od pacjentów poddawanych resekcji guza wykorzystano dane scRNA-seq. Próbki pobrano od 5 pacjentów ze zdiagnozowanym glejakiem wielopostaciowym G4 (IDH prawidłowy). Stwierdzono, że komórki CD11b+ w mikrośrodowisku ludzkiego glejaka wielopostaciowego również nie tworzą jednorodnej populacji, zauważalna była heterogenność profili transkryptomicznych i występowanie subpopulacji. Nie udało się wyodrębnić osobnych grup komórek mikrogleju i makrofagów, natomiast zidentyfikowano różne fenotypy funkcjonalne komórek mieloidalnych (GAMs, *ang. glioma-associated microglia/macrophages*).

Dane uzyskane ze wszystkich próbek zostały przeanalizowane łącznie (jako jeden zbiór danych), jednak z uwagi na występowanie w połączonym zbiorze silnego efektu próbki (zidentyfikowane w sposób nienadzorowany grupy komórek zawierały w zdecydowanej większości komórki pochodzące z jednej próbki) zdecydowano się na wykorzystanie zaimplementowanej w pakiecie Seurat procedury integracji danych. Pozwoliło to na uzyskanie zbioru danych, w którym wpływ efektu próbki na profile transkryptomiczne pojedynczych komórek został zredukowany. W rezultacie zidentyfikowane grupy komórek zawierały komórki pochodzące z różnych próbek i wpływ fenotypu na profile transkryptomiczne dominował nad efektem próbki. Choć zidentyfikowano pewną liczbę komórek innych typów (NK, NKT, OPC), w dalszych analizach skupiono się komórkach mikrogleju/makrofagów, gdyż były to komórki dominujące w uzyskanych danych z mikrośrodowiska glejaka. Trudności w rozdzieleniu mikrogleju i makrofagów interpretujemy jako efekt upodobniania się funkcjonalnego komórek mieloidalnych w glejaku w trakcie wieloletniego wzrostu guza. Subpopulacje komórek zidentyfikowane jako GAMs nie miały wyraźnie podwyższonej ekspresji typowych markerów mikrogleju albo makrofagów, co sugerowało że mogły upodobnić się do siebie profilami transkryptomicznymi.

Analizując populację mikrogleju/makrofagów grupowanie komórek wykonano wielokrotnie (iteracyjnie), wykorzystując różne wartości parametru rozdzielczości, identyfikując za każdym razem geny, których ekspresja była podwyższona w poszczególnych uzyskanych grupach. Wartość parametru rozdzielczości przekładała się na liczbę grup, a w efekcie na przypisanie poszczególnych komórek. Przeanalizowano diagram przepływu komórek pomiędzy grupami wraz ze zmianami rozdzielczości i wybrano rozdzielczości, dla których przypisanie komórek do grup wykazywało dużą stabilność. Wyniki różnych iteracji grupowania porównano i wybrano wartość parametru rozdzielczości, który pozwolił uzyskać populacje charakteryzujące się profilami ekspresji wskazującymi na występowanie podtypów funkcjonalnych mikrogleju/makrofagów w mikrośrodowisku ludzkiego glejaka wielopostaciowego.

Zidentyfikowano populację komórek prezentujących antygen (APC), które miały podwyższoną ekspresję genów kodujących białka MHCII (HLA-DPB1, HLA-DQA1, HLA-DRA, HLA-DPA1) i genów kodujących apolipoproteiny APOC1 i APOE. Komórki o takim fenotypie mogą promować bardziej inwazyjny i agresywny rozwój nowotworu (Sarantopoulos, Ene and Aquilanti, 2024). Z tej grupy komórek na pewnym etapie zwiększania rozdzielczości grupowania wyodrębniła się mniejsza podgrupa, charakteryzująca się dodatkowo podwyższoną ekspresją genów kodujących NFKBIA (*ang. NFKB Inhibitor Alpha*) i CXCR4 (*ang. C-X-C chemokine receptor type 4*) znany też jako białko Fusyna lub CD184 (*ang. cluster of differentiation 184*). NFKBIA jest inhibitorem czynnika NFκB oddziałującym z podjednostką REL NFκB i hamującą kompleksy tego białka uczestniczące w stanach zapalnych (Guo *et al.*, 2024). CXCR4 jest znany ze swojej roli w zasiedlaniu komórek progenitorowych w szpiku kostnym. W przypadku urazu, stresu lub blokady CXCR4 zwiększa się mobilizacja komórek progenitorowych w krwi. CXCR4 reguluje chemotaksji i/lub zatrzymaniu CXCR4+ komórek progenitorowych sterowanych przez ligand CXCL12, który jest wydzielany przez komórki zrębu szpiku kostnego albo na obrzeżach w miejscach uszkodzenia lub stres komórkowy. Podobnie CXCR4 wpływa na mobilizację innych typów komórek, takich jak limfocyty i neutrofile, ale także CXCR4+ komórki nowotworowe (Caspar *et al.*, 2022) Komórki te mogą reprezentować migrujące do guza monocyty, które przekształciły w immunosupresyjne makrofagi.

Druga największa populacja została opisana jako zróżnicowane komórki o profilu prozapalnym, które miały przede wszystkim podwyższoną ekspresję genów kodujących chemokiny CCL3, CCL3L1, CCL4, CCL4L2. Prozapalny fenotyp sugeruje przeciwnowotworową funkcję tych komórek. Do komórek biorących najpewniej udział w odpowiedzi przeciwnowotworowej można też zaliczyć grupę komórek, w której zaobserwowano podwyższoną ekspresję genów indukowanych interferonem (m.in. ISG15, IFIT1, IFIT3, IFITM3, IFIT2). Komórki te mogą reprezentować monocyty świeżo migrujące do guza, które mają fenotyp zapalny (przeciwnowotworowy) i nie przekształciły jeszcze w immunosupresyjne makrofagi. Ponieważ w pobranym fragmencie guza jest obecna krew, nie ma możliwości stwierdzenia, czy pochodziły one z guza czy też z otaczającej krwi śródoperacyjnej.

Kolejna grupa komórek charakteryzowała się podwyższoną ekspresją genów kodujących białka, których obecność powiązana jest z występowaniem hipoksji, m.in. PLIN2 (*ang. Perilipin 2*), TIMP1 (*ang. TIMP metallopeptidase inhibitor 1*), VIM (*ang. Vimentin*) i BNIP3 (*ang. BCL2 Interacting Protein 3*). Polaryzacja makrofagów do takiego fenotypu może następować w niszach w pobliżu tkanek nekrotycznych. Mogą one następnie destabilizować połączenia adherencyjne śródbłonka poprzez aktywację sygnalizacji parakrynnej adrenomedulliny i w efekcie stymulację powstawania wysoce przepuszczalnych nowych naczyń krwionośnych, które utrudniają dostarczenie leku do guza (co zaobserwowano w modelu przeszczepu heterogenicznego glejaka do myszy (*ang. xenograft*) (Wang *et al.*, 2024). Wykazano, że zablokowanie adrenomedulliny, wydzielanej przez makrofagi o fenotypie wskazującym na występowanie hipoksji, prowadzi do zmniejszenia przepuszczalności naczyń krwionośnych, a w efekcie do zwiększenia stężenia podawanego leku wewnątrz guza, co pozwala osiągnąć korzyści terapeutyczne. Proporcja makrofagów o takim fenotypie w mikrośrodowisku, a także poziom ekspresji adrenomedulliny umożliwia prognozowanie występowania naczyń krwionośnych o wysokiej przepuszczalności, jak

również jest negatywnym czynnikiem prognostycznym. Z tego względu, ta grupa makrofagów może być potencjalnym celem terapeutycznym (Wang et al., 2024). Występowanie hipoksji może mieć też wpływ na przestrzenną strukturę glejaka wielopostaciowego. W najnowszych badaniach stwierdzono, że guz można podzielić na pięć warstw, w których dominują komórki nowotworowe w określonym stanie (Greenwald *et al.*, 2024). Bliskość obszaru hipoksji/nekrozy powiązana została z wyraźniejszą, bardziej uporządkowaną strukturą warstwy. Z drugiej strony, w nowotworach, w których hipoksja/nekroza nie występowały (np. glejaki ze zmutowanym IDH) komórki nowotworowe o odmiennych stanach były wymieszane w przestrzeni, nie tworząc tak wyraźnej struktury.

Zidentyfikowano również dwie grupy komórek o własnościach fagocytarnych, jedną z podwyższoną ekspresją genów kodujących MSR1 i CDC42SE1, a drugą z podwyższoną ekspresją genów kodujących GPNMB, LIPA, CD9 i TREM2. Glikoproteina-NMB (GPNMB, ang. Glycoprotein-NMB) jest glikoproteiną transbłonową o podwyższonej ekspresji w nowotworach i komórkach mieloidalnych. GPNMB moduluje odporność wrodzoną i nabytą poprzez działanie polegające na tłumieniu odpowiedzi prozapalnych, jednocześnie promując rekrutację i polaryzację komórek immunosupresyjnych (Lazaratos, Annis and Siegel, 2022). Sygnatura drugiej z nich sugerowała zdolność tych komórek do przetwarzania lipidów (Remmerie, Martens and Scott, 2020). Badania wskazują, że makrofagi o wysokiej ekspresji GPNMB biorą udział w przetwarzaniu bogatej w cholesterol mieliny, co powoduje polaryzację w stronę immunosupresyjnego fenotypu. Dodatkowo makrofagi należące do tej populacji przekazują lipidy będące pochodnymi mieliny do komórek glejaka i w ten sposób wspierają rozwój guza. Wykazano, że obecność tego typu makrofagów w mikrośrodowisku koreluje z występowaniem mezenchymalnego, agresywnego, podtypu komórek glejaka i gorszym rokowaniem dla pacjenta. Z tego względu wykrywanie makrofagów należących do tej grupy może mieć w przyszłości wartość diagnostyczną i ułatwiać wybór spersonalizowanej terapii dla pacjenta. W szczególności może być alternatywą dla obecnie stosowanych kryteriów (badania częstości występowania mutacji somatycznych w tkance nowotworowej (TMB, ang. tumor mutational burden) w predykcji odpowiedzi pacjenta na immunoterapię (Kloosterman et al., 2024).

W analizowanym zbiorze mikrogleju/makrofagów wyszczególniono populację komórek dendrytycznych o stłumionej odpowiedzi (fenotyp M2, podwyższona ekspresja genu TMEM176B), komórek o podwyższonej ekspresji genów kodujących białka rybosomalne, a także komórek będących prawdopodobnie w trakcie podziału (wykazujących podwyższoną ekspresję genów powiązanych z fazą S albo fazą G2M cyklu komórkowego).

Przeprowadzone analizy pozwoliły wyodrębnić podtypy funkcjonalne mikrogleju/makrofagów w mikrośrodowisku ludzkiego glejaka wielopostaciowego. Jest to heterogenna populacja, której dokładne poznanie (na poziomie genomicznym, transkryptomicznym, proteomicznym) jest istotne dla pełnego zrozumienia mechanizmów patologii glejaków.

## 7.3 Ograniczenia analizy profili transkryptomicznych pojedynczych komórek scRNA-seq

Wysokoprzepustowe metody analizy ekspresji genów w pojedynczych komórkach mają szereg ograniczeń wynikających z ich specyfiki technicznej, a także z biologicznych właściwości analizowanych próbek. Planując badania i interpretując ich wyniki wskazane jest, aby te ograniczenia uwzględnić.

#### 7.3.1 Ograniczenia techniczne

#### 7.3.1.1 Niska ilość mRNA z pojedynczych komórek

Proces wychwytywania transkryptów z pojedynczych komórek ma ograniczoną efektywność i zazwyczaj odczytywanych jest tylko około kilku tysięcy unikalnych transkryptów z jednej komórki. Z tego powodu w danych może występować szum techniczny, powiązany z niewystarczającym pokryciem, fałszywie ujemnymi informacjami o braku ekspresji niektórych genów (gdy żaden transkrypt dla tych genów nie został wychwycony, pomimo występowania w komórce) lub zniekształconą proporcją wychwyconych transkryptów, powiązaną z preferencyjnym wychwytywaniem transkryptów pewnej grupy genów względem pozostałych. Prowadzi to do analizowania zaburzonych danych profili ekspresji pojedynczych komórek i w efekcie do trudności z wiarygodną interpretacją otrzymanych wyników.

### 7.3.1.2 Błędy i nierównomierność amplifikacji

Proces powielania fragmentów cDNA nie jest równomierny i w praktyce niektóre sekwencje są powielane bardziej efektywnie niż inne. Ze względu na to, że dane scRNA-seq są sekwencjonowane relatywnie płytko (per komórka) powoduje to problem we właściwej interpretacji zerowej liczby odczytów w danych dla pojedynczego genu. Trudno o jednoznaczną odpowiedź na pytanie, czy gen ten w rzeczywistości nie miał ekspresji w tej konkretnej komórce, czy może amplifikacja nie zadziałała wystarczająco efektywnie i ekspresja nie została wykryta. Dodatkowym problemem występującym na etapie amplifikacji są błędy, które ten proces czasem wprowadza. Do pewnego stopnia można zminimalizować wpływ tego efektu poprzez wykorzystanie UMI zaprojektowanych w sposób umożliwiający (na etapie analiz bioinformatycznych) korekcję błędów w danych po sekwencjonowaniu.

#### 7.3.1.3 Efekt serii/próbki (ang. batch effect)

Wpływ na próbkę (a docelowo na odczytane profile transkryptomiczne pojedynczych komórek) wszelkiego rodzaju czynników nie powiązanych z mechanizmami biologicznymi jest zjawiskiem niepożądanym. Źródłem takiego wpływu może być m.in. niejednolity sposób przygotowania lub przechowywania próbek, różnice w parametrach sekwencjonowania lub różne serie wykorzystanych odczynników. Wskazane jest, aby cały proces od pobrania próbki do sekwencjonowania był przeprowadzony w maksymalnie jednolity sposób i możliwie dokładnie udokumentowany. Umożliwia to na etapie interpretacji wyników na stawianie bardziej wiarygodnych hipotez dotyczących obserwowanych różnic w populacjach komórek lub powiązanie różnic z ewentualnymi problemami na etapie przygotowania danych.

#### 7.3.2 Ograniczenia biologiczne

#### 7.3.2.1 Zróżnicowane profile transkryptomiczne komórek będących częścią jednej populacji

Komórki będące biologicznie częścią jednej populacji mogą mieć zróżnicowane profile transkryptomiczne i może to wynikać m.in. ze zmian powiązanych z cyklem komórkowym lub zjawiskiem "wybuchów transkrypcji" (transcriptional bursts) (Tunnacliffe and Chubb, 2020).

Zjawisko to opisuje sytuację, w której polimeraza w krótkich okresach intensywnie syntetyzuje mRNA ("wybuchy"), a poza nimi przez długie okresy transkrypcja nie zachodzi. Profil transkryptomiczny pojedynczej komórki opisuje stan tej komórki w jednym tylko momencie, w którym jej mRNA zostało oznakowane w kropli (lub w momencie, w którym komórka została utrwalona, w przypadku wykorzystania odpowiedniego protokołu). Innym powodem zróżnicowania profili transkryptomicznych komórek będących częścią jednej populacji może być wpływ nawet niewielkich różnic w reakcji komórek na proces przygotowywania próbki. W części komórek mogą zostać uruchomione programy transkrypcyjne powiązane z odpowiedzią na stres i gdy odpowiedź ta będzie wystarczająco silna, to różnice będą widoczne w analizowanych danych. Pozostawia to pole do błędnej interpretacji wyników i np. wysunięcia hipotezy, że komórki odpowiedziały stresem na obecność nowotworu, gdy w rzeczywistości była to jedynie odpowiedź części pojedynczej populacji komórek na procedurę laboratoryjną.

#### 7.3.2.2 Odczytywanie krótkich fragmentów transkryptów

W najczęściej wykorzystywanych systemach 10x Genomics transkrypty odczytywane są na długości blisko 100 par zasad od jednego z końców (3' lub 5' w zależności od protokołu). W przypadku wielu genów uniemożliwia to wiarygodną analizę alternatywnego splicingu lub izoform genów. Podobna sytuacja ma miejsce w przypadku wykrywania mutacji, które obecne są w większej odległości od wybranego w protokole końca transkryptu niż długość odczytu informacje o takich mutacjach nie znajdą się w danych i nie będą mogły zostać uwzględnione podczas interpretacji wyników. W badaniach, w których informacja o występujących w transkryptach mutacjach jest szczególnie istotna, podejmowano próby zastosowania mieszanego podejścia, wykorzystując urządzenia umożliwiające sekwencjonowanie krótkich odczytów i urządzenia umożliwiające sekwencjonowanie długich odczytów (biblioteki były sekwencjonowane dwukrotnie, po jednym razie na urządzeniu każdego typu).

#### 7.3.2.3 Zależności przestrzenne

W wielu badaniach zostało wykazane, że komórki tego samego typu znajdujące się, w różnych obszarach w przestrzeni próbki miały zróżnicowane profile transkrypcyjne.

Wysokoprzepustowe metody mikroprzepływowe nie pozwalają na zachowanie informacji o lokalizacji poszczególnych komórek w obrębie próbki (informacja ta jest tracona w trakcie dysocjacji tkanki podczas przygotowywania próbki). Z tego względu w wynikach analiz można czasem zaobserwować, że komórki jednego typu występują w różnych stanach (np. objawiających się w aktywowaniu charakterystycznych ścieżek sygnałowych), ale nie ma możliwości, aby te różnice zinterpretować w kontekście przestrzennym. W szczególności nie ma możliwości, aby uwzględnić bezpośrednie sąsiedztwo konkretnej komórki, a następnie ocenić czy występują między sąsiadującymi komórkami znane mechanizmy komunikacji. Powstały metody do statystycznej estymacji prawdopodobieństwa i intensywności komunikacji pomiędzy zidentyfikowanymi populacjami komórek (Jin *et al.,* 2021), działające w oparciu o dane o ekspresji genów kodujących białka znanych ligandów i receptorów i identyfikację występowania znanych par ligand-receptor pomiędzy populacjami komórek. Wskazane jest jednak, aby interpretując te wyniki pamiętać o tym, że ekspresji genów kodujących białka nie da się w jednoznaczny sposób przełożyć na obecność białek.

## 8 WNIOSKI

- 1. Zoptymalizowano metodykę analizy i wizualizacji danych scRNA-seq z komórek mieloidalnych w mikrośrodowisku doświadczalnych glejaków.
- Wybrano wartości parametrów dla uzyskania optymalnych wyników analiz i wizualizacji danych scRNA-seq z komórek mieloidalnych w mikrośrodowisku doświadczalnych glejaków.
- 3. Wykazano, że komórki mieloidalne w mózgach myszy naiwnych i po operacji pozorowanej mają podobne profile ekspresji.
- Wybrano wartości parametrów dla uzyskania optymalnych wyników analiz i wizualizacji danych scRNA-seq komórek mieloidalnych w mikrośrodowisku ludzkich glejaków wielopostaciowych.
- 5. Zidentyfikowano i scharakteryzowano subpopulacje mieloidalne w mikrośrodowisku ludzkich guzów glejaków wielopostaciowych.

# 9 LISTA PUBLIKACJI Z UDZIAŁEM AUTORA ROZPRAWY

Ochocka, N.\*, **Segit, P.\***, Walentynowicz, K. A., Wojnicki, K., Cyranowski, S., Swatler, J., Mieczkowski, J., & Kaminska, B. (2021). Single-cell RNA sequencing reveals functional heterogeneity of glioma-associated brain macrophages. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-21407-w

Kaminska, B., Ochocka, N., & **Segit, P.** (2021). Single-cell omics in dissecting immune microenvironment of malignant gliomas—Challenges and perspectives. In *Cells* (Vol. 10, Issue 9). https://doi.org/10.3390/cells10092264

Ochocka, N., **Segit, P.**, Wojnicki, K., Cyranowski, S., Swatler, J., Jacek, K., Grajkowska, W., & Kaminska, B. (2023). Specialized functions and sexual dimorphism explain the functional diversity of the myeloid populations during glioma progression. *Cell Reports*, *42*(1). https://doi.org/10.1016/j.celrep.2022.111971

# **10 BIBLIOGRAFIA**

Averill, M.M. *et al.* (2011) 'S100A9 differentially modifies phenotypic states of neutrophils, macrophages, and dendritic cells: Implications for atherosclerosis and adipose tissue inflammation', *Circulation*, 123(11). doi:10.1161/CIRCULATIONAHA.110.985523.

Baharlou, H. *et al.* (2019) 'Mass Cytometry Imaging for the Study of Human Diseases— Applications and Data Analysis Strategies', *Frontiers in Immunology*. doi:10.3389/fimmu.2019.02657.

Bandura, D.R. *et al.* (2009) 'Mass cytometry: Technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry', *Analytical Chemistry*, 81(16). doi:10.1021/ac901049w.

Barteneva, N.S., Fasler-Kan, E. and Vorobjev, I.A. (2012) 'Imaging Flow Cytometry: Coping with Heterogeneity in Biological Systems', *Journal of Histochemistry and Cytochemistry*. doi:10.1369/0022155412453052.

Bellail, A.C. *et al.* (2004) 'Microregional extracellular matrix heterogeneity in brain modulates glioma cell invasion', *International Journal of Biochemistry and Cell Biology*. doi:10.1016/j.biocel.2004.01.013.

Borst, K., Dumas, A.A. and Prinz, M. (2021) 'Microglia: Immune and non-immune functions', *Immunity*. doi:10.1016/j.immuni.2021.09.014.

Brodie, T.M. and Tosevski, V. (2017) 'High-dimensional single-cell analysis with mass cytometry', *Current Protocols in Immunology*, 2017. doi:10.1002/cpim.31.

Caspar, B. *et al.* (2022) 'CXCR4 as a novel target in immunology: moving away from typical antagonists', *Future Drug Discovery*, 4(2). doi:10.4155/fdd-2022-0007.

Conforti, F. *et al.* (2018) 'Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis', *The Lancet Oncology*, 19(6). doi:10.1016/S1470-2045(18)30261-4.

Cossarizza, A. *et al.* (2019) 'Guidelines for the use of flow cytometry and cell sorting in immunological studies (second edition)', *European Journal of Immunology*, 49(10). doi:10.1002/eji.201970107.

Darmanis, S. *et al.* (2017) 'Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma', *Cell Reports*, 21(5). doi:10.1016/j.celrep.2017.10.030.

Friebel, E. *et al.* (2020) 'Single-Cell Mapping of Human Brain Cancer Reveals Tumor-Specific Instruction of Tissue-Invading Leukocytes', *Cell*, 181(7). doi:10.1016/j.cell.2020.04.055.

Gabrusiewicz, K. *et al.* (2011) 'Characteristics of the alternative phenotype of microglia/macrophages and its modulation in experimental gliomas', *PLoS ONE*, 6(8). doi:10.1371/journal.pone.0023902.

Gieryng, A. *et al.* (2017) 'Immune microenvironment of gliomas', *Laboratory Investigation*, 97(5), pp. 498–518. doi:10.1038/labinvest.2017.19.

Greenwald, A.C. *et al.* (2024) 'Integrative spatial analysis reveals a multi-layered organization of glioblastoma', *Cell*, 187(10). doi:10.1016/j.cell.2024.03.029.

Groborz, K. and Drąg, M. (2021) 'WIZUALIZACJA ENZYMÓW PROTEOLITYCZNYCH ZA POMOCĄ SOND CHEMICZNYCH ORAZ CYTOMETRII MASOWEJ', *Wiadomości Chemiczne* [Preprint]. doi:10.53584/wiadchem.2021.12.2.

Guo, K. *et al.* (2001) 'Hypoxia induces the expression of the pro-apoptotic gene BNIP3', *Cell Death and Differentiation*, 8(4). doi:10.1038/sj.cdd.4400810.

Guo, Q. *et al.* (2024) 'NF-κB in biology and targeted therapy: new insights and translational implications', *Signal Transduction and Targeted Therapy*. doi:10.1038/s41392-024-01757-9.

Hara, T. *et al.* (2021) 'Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma', *Cancer Cell*, 39(6). doi:10.1016/j.ccell.2021.05.002.

Hashimshony, T. *et al.* (2016) 'CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq', *Genome Biology*, 17(1). doi:10.1186/s13059-016-0938-8.

Hinton, G. and Roweis, S. (2002) 'Stochastic Neighbor Embedding', in *NIPS 2002: Proceedings of the 15th International Conference on Neural Information Processing Systems*. doi:10.1007/978-3-031-10602-6\_16.

Jin, S. *et al.* (2021) 'Inference and analysis of cell-cell communication using CellChat', *Nature Communications*, 12(1). doi:10.1038/s41467-021-21246-9.

Jollife, I.T. and Cadima, J. (2016) 'Principal component analysis: A review and recent developments', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. doi:10.1098/rsta.2015.0202.

Kaminska, B., Ochocka, N. and Segit, P. (2021) 'Single-cell omics in dissecting immune microenvironment of malignant gliomas—Challenges and perspectives', *Cells*. doi:10.3390/cells10092264.

Khan, M.T. *et al.* (2021) 'Identification of Gender-Specific Molecular Differences in Glioblastoma (GBM) and Low-Grade Glioma (LGG) by the Analysis of Large Transcriptomic and Epigenomic Datasets', *Frontiers in Oncology*, 11. doi:10.3389/fonc.2021.699594.

Kierdorf, K. *et al.* (2019) 'Macrophages at CNS interfaces: ontogeny and function in health and disease', *Nature Reviews Neuroscience*. doi:10.1038/s41583-019-0201-x.

Klemm, F. *et al.* (2020) 'Interrogation of the Microenvironmental Landscape in Brain Tumors Reveals Disease-Specific Alterations of Immune Cells', *Cell*, 181(7). doi:10.1016/j.cell.2020.05.007. Kloosterman, D.J. *et al.* (2024) 'Macrophage-mediated myelin recycling fuels brain cancer malignancy', *Cell* [Preprint]. doi:10.1016/j.cell.2024.07.030.

Korin, B. *et al.* (2017) 'High-dimensional, single-cell characterization of the brain's immune compartment', *Nature Neuroscience*, 20(9). doi:10.1038/nn.4610.

Korin, B., Dubovik, T. and Rolls, A. (2018) 'Mass cytometry analysis of immune cells in the brain', *Nature Protocols*, 13(2). doi:10.1038/nprot.2017.155.

Kowalczyk, M.S. *et al.* (2015) 'Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells', *Genome Research*, 25(12). doi:10.1101/gr.192237.115.

Lazaratos, A.M., Annis, M.G. and Siegel, P.M. (2022) 'GPNMB: a potent inducer of immunosuppression in cancer', *Oncogene*. doi:10.1038/s41388-022-02443-2.

Lin, G.L. *et al.* (2018) 'Non-inflammatory tumor microenvironment of diffuse intrinsic pontine glioma', *Acta neuropathologica communications*, 6(1). doi:10.1186/s40478-018-0553-x.

Louis, D.N. *et al.* (2016) 'The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary', *Acta Neuropathologica*. doi:10.1007/s00401-016-1545-1.

Van Der Maaten, L. and Hinton, G. (2008) 'Visualizing data using t-SNE', *Journal of Machine Learning Research*, 9.

Macosko, E.Z. *et al.* (2015) 'Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets', *Cell*, 161(5). doi:10.1016/j.cell.2015.05.002.

Mathewson, N.D. *et al.* (2021) 'Inhibitory CD161 receptor identified in glioma-infiltrating T cells by single-cell analysis', *Cell*, 184(5). doi:10.1016/j.cell.2021.01.022.

McCowan, J. *et al.* (2021) 'The transcription factor EGR2 is indispensable for tissue-specific imprinting of alveolar macrophages in health and tissue repair', *Science Immunology*, 6(65). doi:10.1126/sciimmunol.abj2132.

McInnes, L. *et al.* (2018) 'UMAP: Uniform Manifold Approximation and Projection', *Journal of Open Source Software*, 3(29). doi:10.21105/joss.00861.

McKinnon, K.M. (2018) 'Flow cytometry: An overview', *Current Protocols in Immunology*, 2018. doi:10.1002/cpim.40.

Mortazavi, A. *et al.* (2008) 'Mapping and quantifying mammalian transcriptomes by RNA-Seq', *Nature Methods*, 5(7). doi:10.1038/nmeth.1226.

Müller, S. *et al.* (2017) 'Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment', *Genome Biology*, 18(1). doi:10.1186/s13059-017-1362-4.

Neftel, C. *et al.* (2019) 'An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma', *Cell*, 178(4). doi:10.1016/j.cell.2019.06.024.

Ochocka, N. *et al.* (2021) 'Single-cell RNA sequencing reveals functional heterogeneity of glioma-associated brain macrophages', *Nature Communications*, 12(1). doi:10.1038/s41467-021-21407-w.

Ochocka, N. *et al.* (2023) 'Specialized functions and sexual dimorphism explain the functional diversity of the myeloid populations during glioma progression', *Cell Reports*, 42(1). doi:10.1016/j.celrep.2022.111971.

Ochocka, N. and Kaminska, B. (2021) 'Microglia diversity in healthy and diseased brain: Insights from single-cell omics', *International Journal of Molecular Sciences*. doi:10.3390/ijms22063027.

Osorio, D. and Cai, J.J. (2021) 'Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA-sequencing data quality control', *Bioinformatics*, 37(7). doi:10.1093/bioinformatics/btaa751.

Ozawa, T. *et al.* (2014) 'Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma', *Cancer Cell*, 26(2). doi:10.1016/j.ccr.2014.06.005.

Park, Y.W. *et al.* (2023) 'The 2021 WHO Classification for Gliomas and Implications on Imaging Diagnosis: Part 1—Key Points of the Fifth Edition and Summary of Imaging Findings on Adult-Type Diffuse Gliomas', *Journal of Magnetic Resonance Imaging*. doi:10.1002/jmri.28743.

Patel, A.P. *et al.* (2014) 'Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma', *Science*, 344(6190). doi:10.1126/science.1254257.

Picelli, S. *et al.* (2014) 'Full-length RNA-seq from single cells using Smart-seq2', *Nature Protocols*, 9(1). doi:10.1038/nprot.2014.006.

Pombo Antunes, A.R. *et al.* (2021) 'Single-cell profiling of myeloid cells in glioblastoma across species and disease stage reveals macrophage competition and specialization', *Nature Neuroscience*, 24(4). doi:10.1038/s41593-020-00789-y.

Prinz, M. *et al.* (2021) 'Microglia and Central Nervous System-Associated Macrophages mdash from Origin to Disease Modulation', *Annual Review of Immunology*. doi:10.1146/annurev-immunol-093019-110159.

Qin, D. *et al.* (1998) 'Microfabrication, Microstructures and Microsystems', in. doi:10.1007/3-540-69544-3\_1.

Quail, D.F. and Joyce, J.A. (2017) 'The Microenvironmental Landscape of Brain Tumors', *Cancer Cell*. doi:10.1016/j.ccell.2017.02.009.

Remmerie, A., Martens, L. and Scott, C.L. (2020) 'Macrophage Subsets in Obesity, Aligning the Liver and Adipose Tissue', *Frontiers in Endocrinology*. doi:10.3389/fendo.2020.00259.

Sade-Feldman, M. *et al.* (2018) 'Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma', *Cell*, 175(4). doi:10.1016/j.cell.2018.10.038.

Sankowski, R. *et al.* (2019) 'Mapping microglia states in the human brain through the integration of high-dimensional techniques', *Nature Neuroscience*, 22(12). doi:10.1038/s41593-019-0532-y.

Sarantopoulos, A., Ene, C. and Aquilanti, E. (2024) 'Therapeutic approaches to modulate the immune microenvironment in gliomas.', *NPJ precision oncology*, 8(1), p. 241. doi:10.1038/s41698-024-00717-4.

Schena, M. *et al.* (1995) 'Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray', *Science*, 270(5235). doi:10.1126/science.270.5235.467.

Segovia, M. *et al.* (2019) 'Targeting TMEM176B Enhances Antitumor Immunity and Augments the Efficacy of Immune Checkpoint Blockers by Unleashing Inflammasome Activation', *Cancer Cell*, 35(5). doi:10.1016/j.ccell.2019.04.003.

Stuart, T. *et al.* (2019) 'Comprehensive Integration of Single-Cell Data', *Cell*, 177(7). doi:10.1016/j.cell.2019.05.031.

Tang, F. *et al.* (2009) 'mRNA-Seq whole-transcriptome analysis of a single cell', *Nature Methods*, 6(5). doi:10.1038/nmeth.1315.

Tunnacliffe, E. and Chubb, J.R. (2020) 'What Is a Transcriptional Burst?', *Trends in Genetics*. doi:10.1016/j.tig.2020.01.003.

Venteicher, A.S. *et al.* (2017) 'Decoupling genetics, lineages, and microenvironment in IDHmutant gliomas by single-cell RNA-seq', *Science*, 355(6332). doi:10.1126/science.aai8478.

Verhaak, R.G.W. *et al.* (2010) 'Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1', *Cancer Cell*, 17(1). doi:10.1016/j.ccr.2009.12.020.

Villani, A.C. *et al.* (2017) 'Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors', *Science*, 356(6335). doi:10.1126/science.aah4573.

Wang, Q. *et al.* (2017) 'Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment', *Cancer Cell*, 32(1). doi:10.1016/j.ccell.2017.06.003.

Wang, W. *et al.* (2024) 'Identification of hypoxic macrophages in glioblastoma with therapeutic potential for vasculature normalization', *Cancer Cell*, 42(5), pp. 815-832.e12. doi:10.1016/j.ccell.2024.03.013.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: A revolutionary tool for transcriptomics', *Nature Reviews Genetics*. doi:10.1038/nrg2484.

Weinberger, T. *et al.* (2020) 'Ontogeny of arterial macrophages defines their functions in homeostasis and inflammation', *Nature Communications*, 11(1). doi:10.1038/s41467-020-18287-x.

Zhang, Y. *et al.* (2021) '1p/19q co-deletion status is associated with distinct tumor-associated macrophage infiltration in IDH mutated lower-grade gliomas', *Cellular Oncology*, 44(1). doi:10.1007/s13402-020-00561-1.