

Analiza rakowego środowiska immunologicznego

Michał Stanowski

May 7, 2024

1 Założenia projektu i stopień ich osiągnięcia

Celem projektu było przeanalizowanie próbek 2D biopsji pacjentów w celu znalezienia potencjalnych ognisk walki z nowotworem (TLS, ang. *tertiary lymphoid structures*). Zgodnie z definicją są to miejsca szczególnie bogate w limfocyty B, T i komórki dendrytyczne. W literaturze struktury TLS-ów są bardzo różne, natomiast zazwyczaj w jądrze takiej struktury występuje dużo limfocytów B i T; często limfocyty B otaczają komórki dendrytyczne. Następnie, po znalezieniu kandydatów na TLS-y, sklastrowano je przy użyciu algorytmu KMeans w poszukiwaniu TLS-ów o charakterystycznym składzie komórkowym.

1.1 Budowa słownika

Na początku korzystając z pliku *IF1_phen_to_cell_mapping.csv* wykonano słownik mapujący fenotypy związane z ekspresją danego typu receptora błonowego na konkretne typy komórek. Dodano kolumnę *celltype* do danych z próbki pacjenta.

1.2 Budowa grafów

Następnie korzystając z faktu, że znane były pozycje jąder komórkowych przy użyciu funkcji *radius_neighbors_graph* z pakietu *sklearn.neighbors* zbudowano graf o promieniu $30\text{ }\mu\text{m}$ tylko na limfocytach B i T (oraz niejednoznacznie przypisanych do *celltype* BnT). Taki sam graf rozpięto na wszystkich komórkach, aby później wykorzystać sąsiedztwo limfocytów B i T do poszukiwania pełnego składu TLS-ów.

1.3 Poszukiwanie spójnych składowych

W grafie rozpiętym na limfocytach B i T przy użyciu funkcji *connected_components* z pakietu *scipy.sparse.csgraph* szukano spójnych składowych, żeby wyodrębnić izolowane TLS-y, ogniska walki z nowotworem. Zarówno tutaj, jak i wcześniej w przypadku rozpinań grafu sąsiedztwa zastosowano arbitralne progi odcięcia. Zdecydowano się na, zupełnie niepoparty żadnymi dowodami naukowymi, próg wynoszący 20 limfocytów B i T budujących spójną składową.

1.4 Dodawanie sąsiadów

Następnie, po wyodrębnieniu odpowiednio dużych spójnych składowych, za pomocą wcześniej zbudowanego grafu sąsiedztwa dodano do spójnych składowych inne komórki, które są ich sąsiadami. Zadbano o to, żeby w spójnych składowych nie było duplikatów.

1.5 Budowa wektorów i klastrowanie

Następnie każdy z TLS przekształcono w wektor udziału procentowego konkretnych komórek wraz informacją o pliku, z którego pochodził. Te wektory następnie sklastrowano algorytmem KMeans. Aby wybrać odpowiednią liczbę klastrów w kodzie stosowana jest metoda łokcia (ang. *elbow method*). Te klastry następnie są wyświetlane przy pomocy PCA na wykresie dwuwymiarowym (Fig. 1). Możliwe jest także wyświetlenie

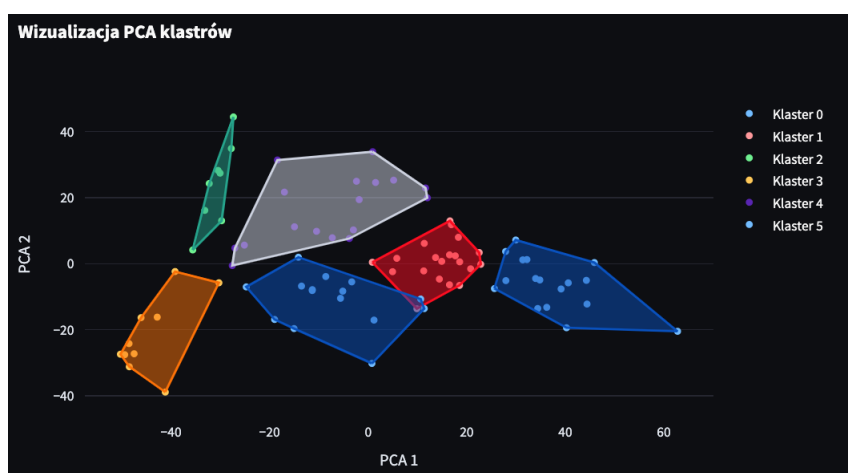


Figure 1: Wykres PCA przedstawiający kolejne klastry obliczone algorytmem KMeans

wykresu słupkowego TLS-ów należących do poszczególnych klastrów, aby zaobserwować podobieństwo w składzie komórkowym TLS-ów należących do danego klastra (Fig. 2).

1.6 Możliwości wizualizacyjne

Oprócz wyżej wspomnianych wykresów PCA klastrów oraz słupkowego TLS-ów po sklastrowaniu, program jest w stanie wykonać:

- interaktywny wykres TLS-ów przed sklastrowaniem, który po najechaniu na dany TLS pokazuje informacje o próbce pacjenta, procentowym udziale w TLS-ie danej komórki oraz numerze TLS-a (Fig. 3). Jest on użyteczny do wizualizacji procesu klastrowania, gdyż identyczny wykres tworzony jest po klastrowaniu.
- interaktywny wykres biopsji pacjenta (Fig. 4). Po najechaniu na konkretną komórkę wyświetla się informacja o typie komórki.
- wykres biopsji z naniesionymi na niego kandydatami na TLS-y (Fig. 5).

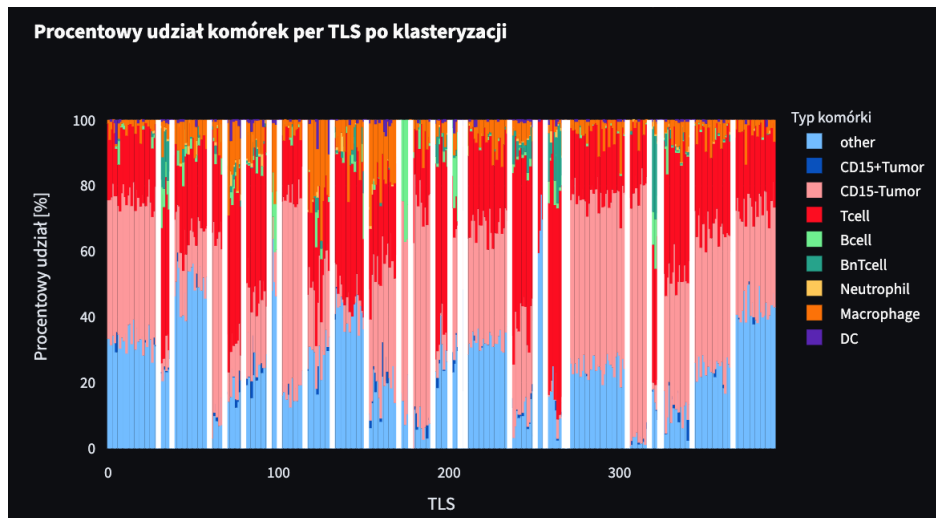


Figure 2: Wykres PCA przedstawiający kolejne klastry obliczone algorytmem KMeans

- wykres nałożonych na siebie TLS-ów należących do jednego klastra wraz z tabelą umożliwiającą lokalizację TLS-ów na oryginalnych wykresach biopsji (Fig. 6)

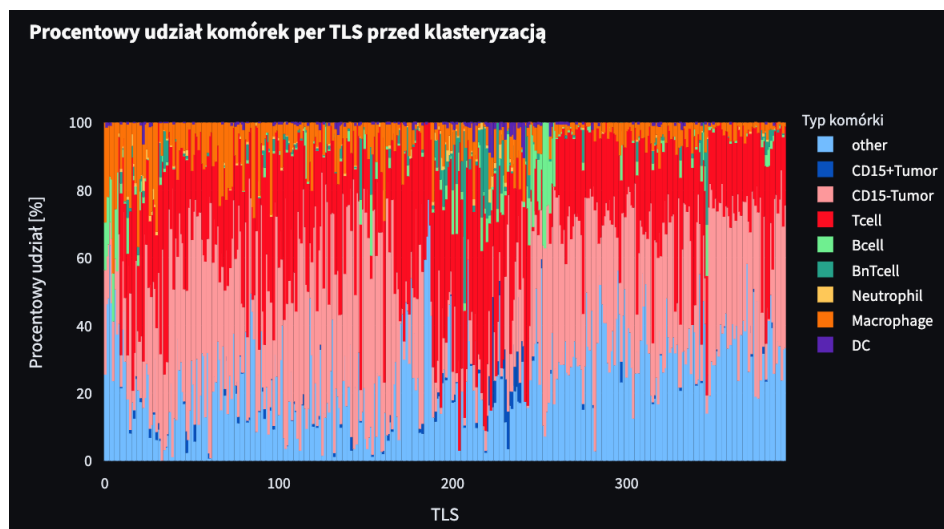


Figure 3: Wykres PCA przedstawiający kolejne klastry obliczone algorytmem KMeans

2 Przykładowy wynik

Program został wykorzystany do analizy wszystkich biopsji IF1. W wyniku klastrowania TLS-ów otrzymano 50 klastrow z 1406 kandydatami na TLS-y. W jednym TLS-ie limfocyty B stanowiły 19,87% komórek, co jest rekordowym wynikiem. Obliczone klastry

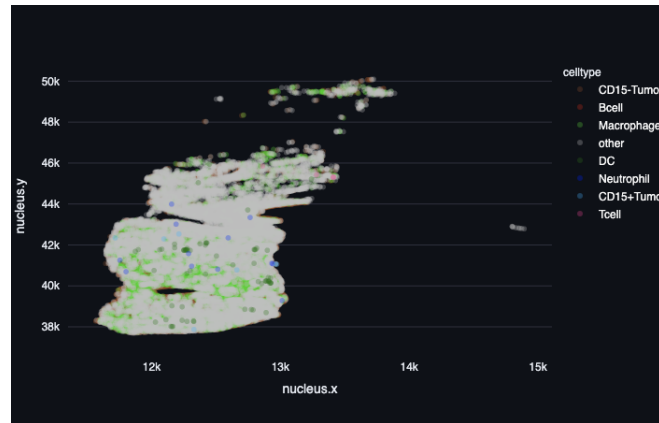


Figure 4: Wykres dwuwymiarowy biopsji pacjenta

zostały przedstawione na Fig. 7. Szczególnie interesujące okazały się klastry 24 i 40. W przypadku klastra 24 (Fig. 8) TLS-y pochodziły z 10 różnych próbek. Jądro klastra złożone jest z limfocytów B i T, znajdują się w nim także pojedyncze komórki dendrytyczne. W przypadku klastra 40 (Fig. 9) TLS-y także pochodziły z 10 różnych próbek. W centrum też występują limfocyty B i T, aczkolwiek w tym przypadku limfocyty B otaczają komórki dendrytyczne w paru miejscach, więc byłoby to zgodne z niektórymi publikacjami.

3 Minusy projektu i powody do zastanowienia

3.1 Tworzenie grafów i szukanie spójnych składowych

Projekt nie ma żadnego znaczenia poza programistycznym, ponieważ progi odcięcia dla sąsiadów komórek ($30 \mu\text{m}$) oraz wielkości spójnych składowych są zupełnie niepoparte żadnymi badaniami naukowymi. Wynika to z minimalnej liczby badań nt. TLS-ów i braku jakichkolwiek danych na temat progów liczebności komórek składających się na TLS-y oraz odległości między komórkami w tym TLS-ie.

3.2 Ilość klastrów

Ilość klastrów i związana z tym ilość kandydatów na TLS-y jest ogromna (w porównaniu z innymi osobami z grupy). Nie jest jednak znana przyczyna tak czułego na TLS-y algorytmu.

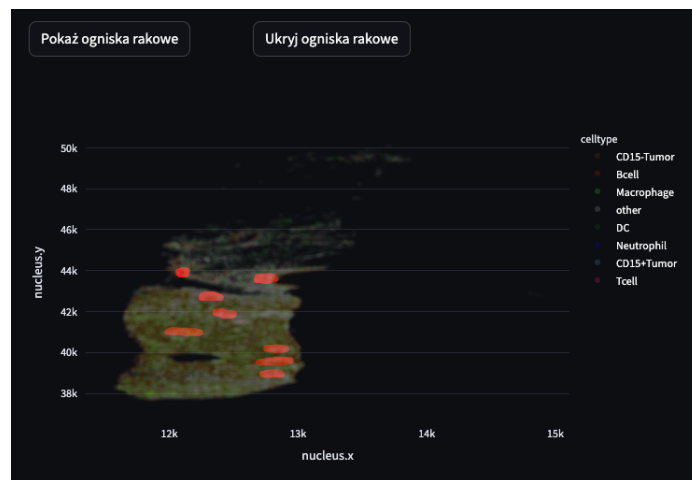


Figure 5: Wykres dwuwymiarowy biopsji pacjenta z naniesionymi na niego kandydatami na ogniska rakowe (czerwone)



Figure 6: Wykres przestrzennie nałożonych na siebie TLS-ów należących do jednego klastra wraz z pozycjami w oryginalnych biopsjach

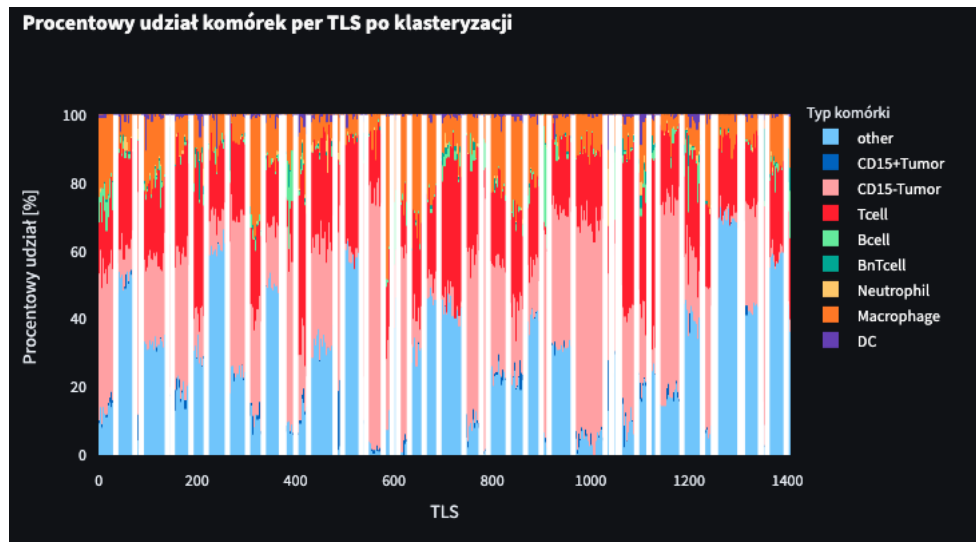


Figure 7: Wykres słupkowy po sklastrowaniu wszystkich TLS-ów pochodzących z biopsji IF1

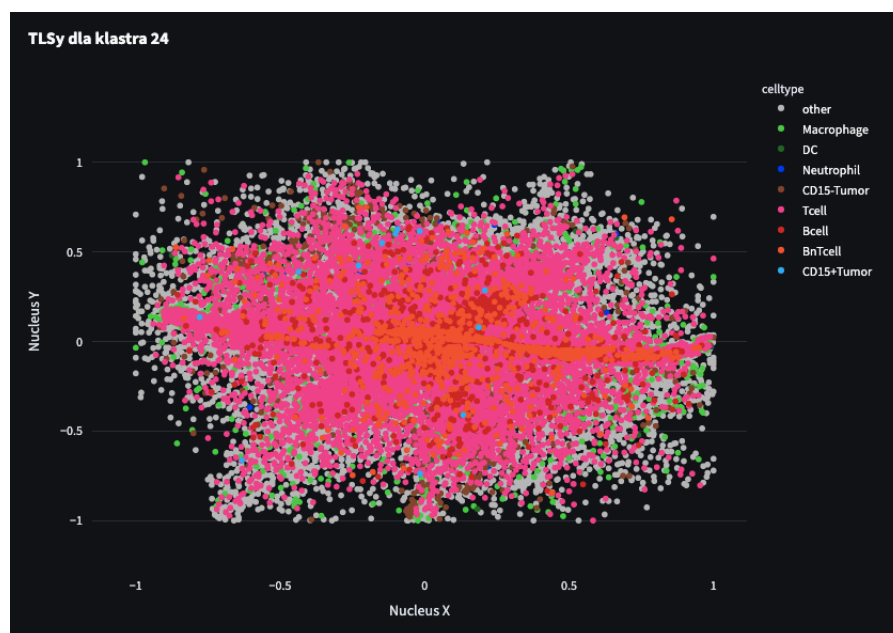


Figure 8: Klaster 24

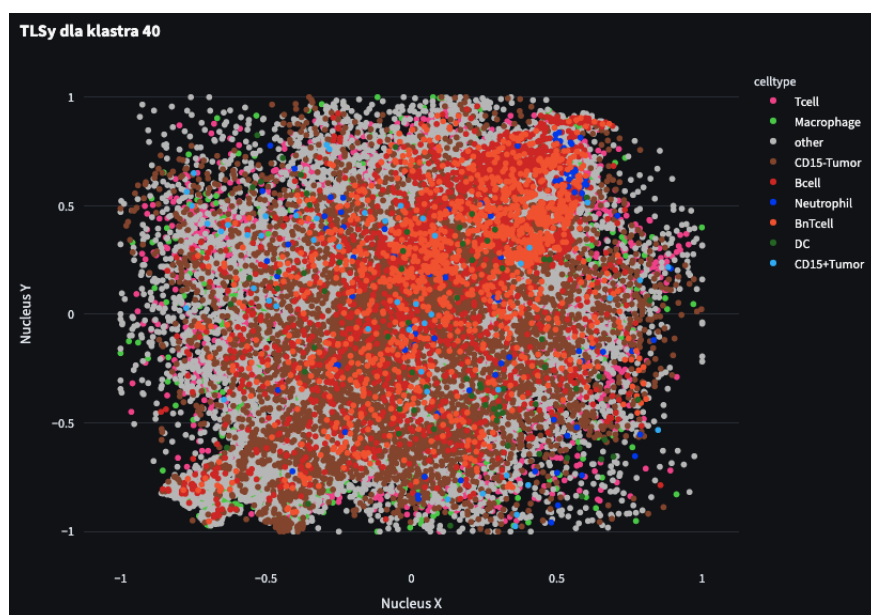


Figure 9: Klaster 40