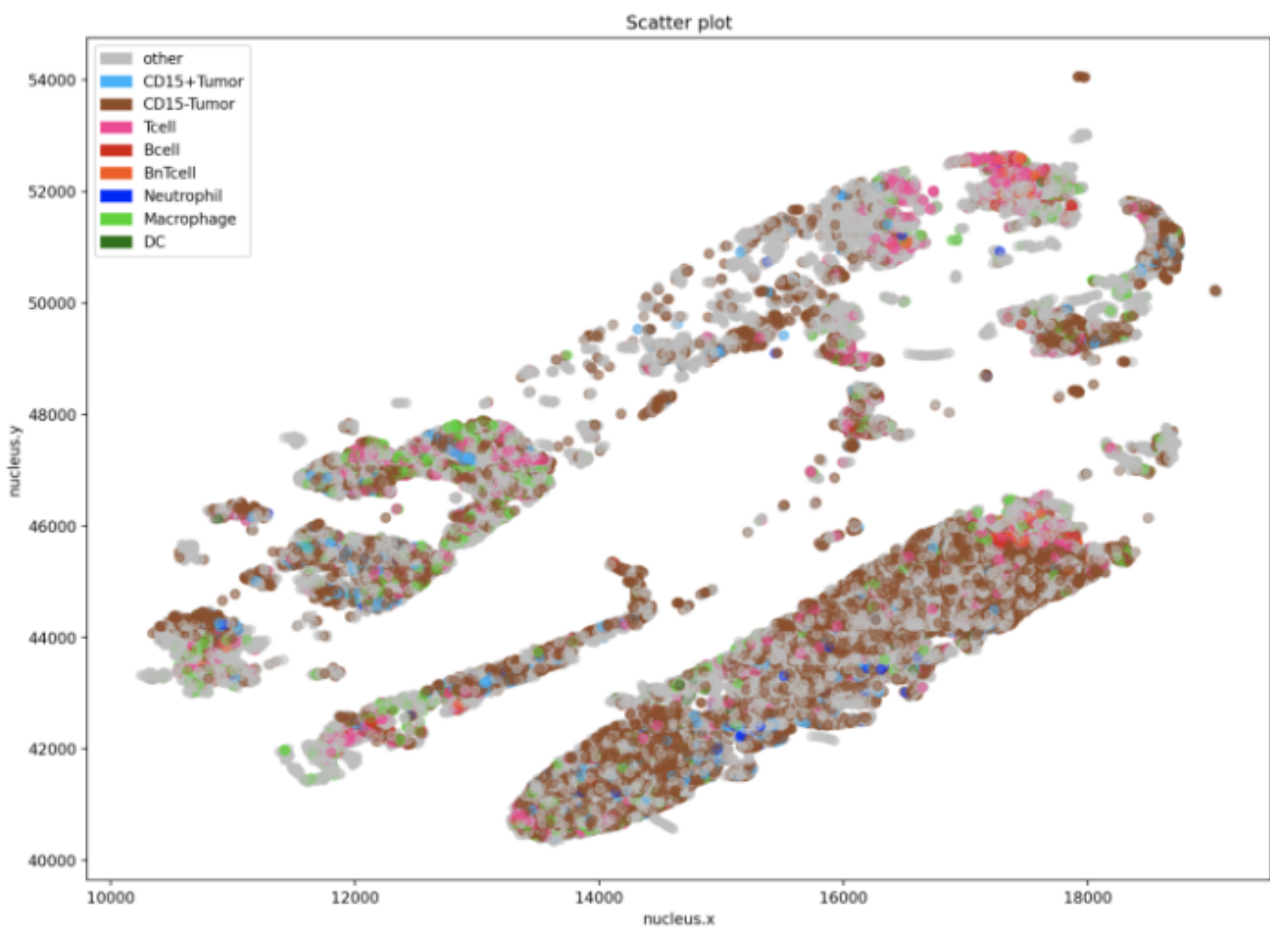


Raport: Skrypt do graficznej analizy próbek z biopsji pobranych od pacjentów rakowych.

Poniższy raport przedstawia wyniki pracy nad skryptem służącym do graficznego przedstawienia i wstępnej analizy danych przestrzennych z próbek biopsji pobranych od pacjentów rakowych. Skrypt został przygotowany w języku python z interfejsem graficznym powstałym przy użyciu biblioteki *streamlit*. Umożliwiło to na interaktywne zarządzanie działaniem programu i wprowadzanie argumentów wejściowych. Wszystkie poniżej pokazane ryciny pochodzą z pliku 1107_IF1.csv (wybrane Tcell, Bcell, BnTcell, 4 klastry, minimalna wielkość grafu 20, maksymalna odległość sąsiadów 30)

Szczególną uwagę poświęcono próbie wykrycia tzw. TLS (tertiary lymphoid structure). Są to limfoidalne tkanki formujące się obrębie lub na granicy nowotworów czy stanów zapalnych. W skład TLS wchodzi komórki układu odpornościowego, na przykład limfocyty T i B. Obecność TLS jest w badaniach łączona z lepszą prognozą pacjentów wynikającą z podejrzenia zwiększonej walki organizmu z nowotworem.

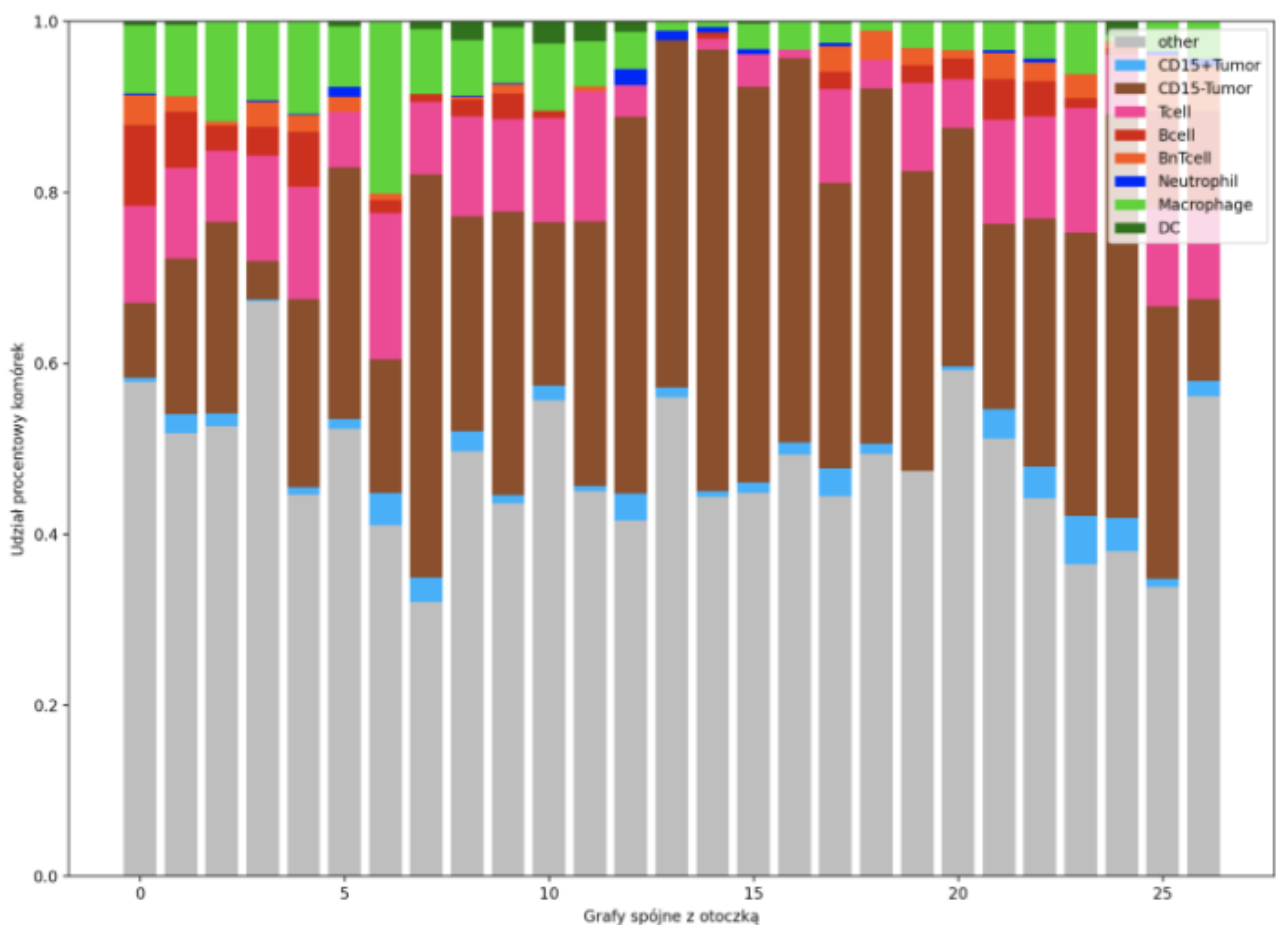
Faktyczną analizę poprzedzono graficznym przedstawieniem danych przestrzennych z zaznaczeniem typów komórek obecnych w próbce. Typy komórki zostały określone dzięki zapisanym w danych markerom molekularnym. Wynik przedstawiono poniżej (Ryc. 1)



Ryc. 1

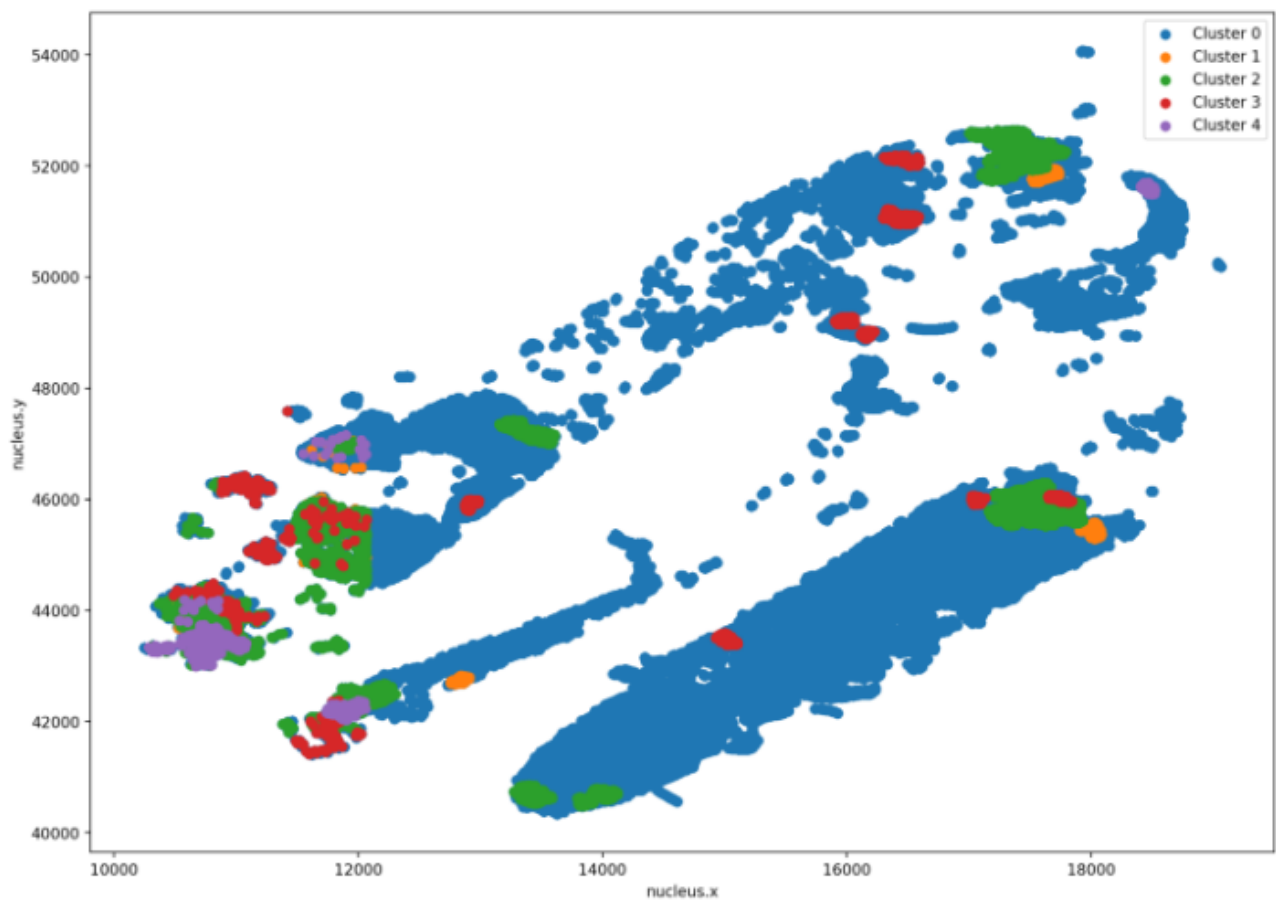
Następnie użyto koordynatów przypisanych do komórek. Każda komórka posiada współrzędną x oraz współrzędną y. Pozwala to na proste wyliczenie odległości pomiędzy komórkami. Wyliczając te odległości, używając funkcji *radius_neighbors_graph* z pakietu sklearn w python znajdowani są sąsiedzi w odległości mniejszej niż podanej w interfejsie *streamlit*. Znalezieni sąsiedzi są łączeni w spójne grafy/komponenty, biorąc pod uwagę najpierw tylko typy komórek tworzące TLS (do wybrania w interfejsie). Do komponentów dodawana jest otoczka z sąsiadów, którzy są komórkami o innych typach.

Takie rozszerzone komponenty, zostały podsumowane wektorem określającym %-owy udział komórek różnych typów. Wykres słupkowy komponentów jest jednym z wyników analizy (Ryc. 2)

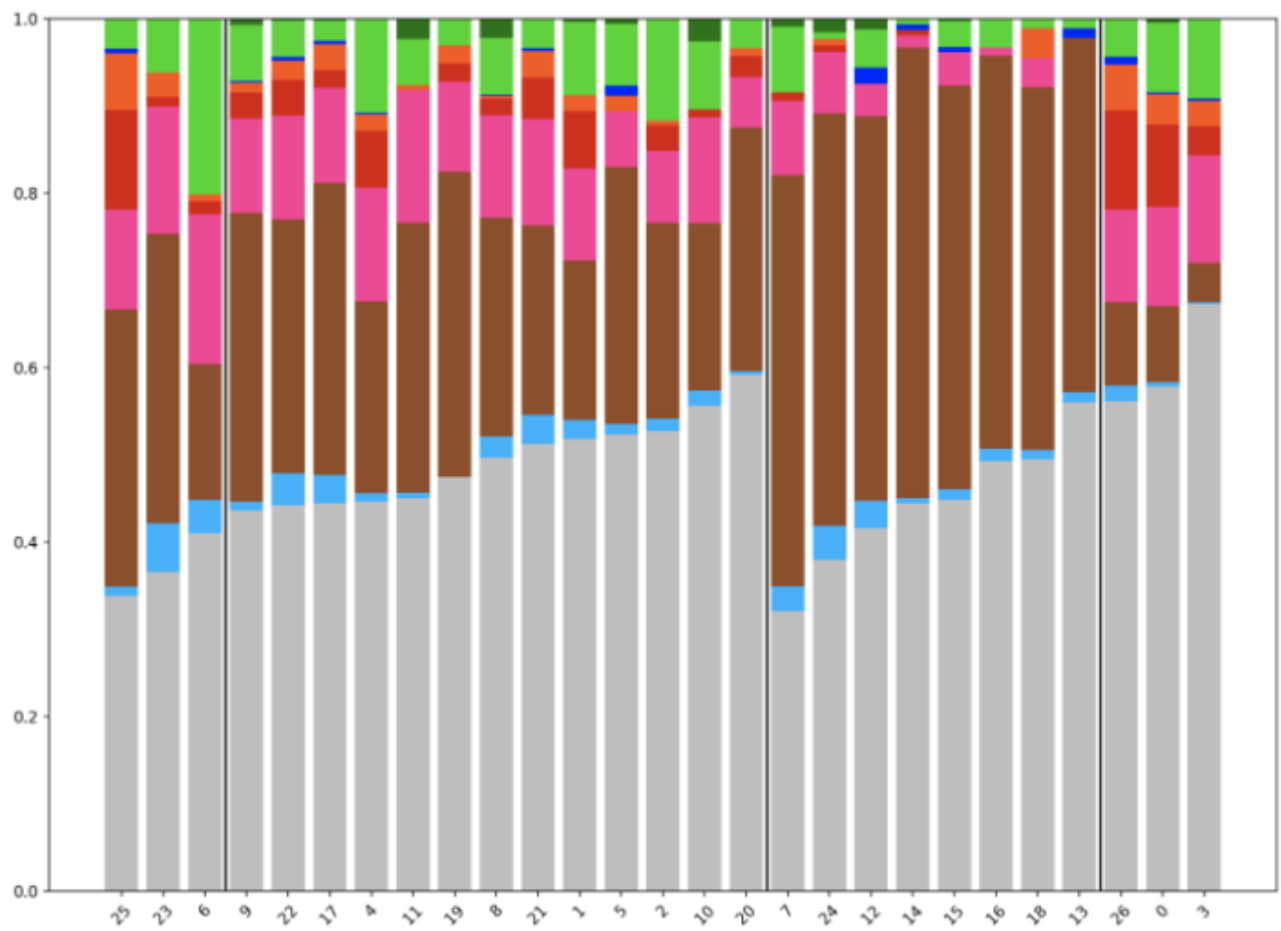


Ryc. 2

Dzięki temu, można było je łatwo przyrównać i co za tym idzie sklastrować. Liczbe klastrow można określić w interfejsie. Do klastrowania użyto metody AgglomerativeClustering. Otrzymane klastry przedstawiono zarówno na przestrzennym grafie poprzez zaznaczenie kolorem (Ryc. 3), jak i na wykresie słupkowym poprzez uporządkowanie komponentów (Ryc.4).



Ryc. 3



Ryc. 4

Przedstawiona analiza pozwala na dodanie kontekstu ułatwiając interpretację danych. %-owy udział typów komórek z jednoczesnymi informacjami o wielkości i lokalizacji skupisk komórek pozwalają na szukanie kandydatów na potencjalne TLSy.