

Mimo że wszystkie komórki w ludzkim ciele zawierają (prawie) tę samą sekwencję DNA, tworzą one różne tkanki i charakteryzują się rozbieżnościami strukturalnymi i funkcjonalnymi, na przykład neurony różnią się od komórek odpornościowych i innych typów komórek. Zjawisko to wynika z faktu, że w komórkach różnych typów aktywne są różne geny, w wyniku czego wytwarzane są różne mRNA i białka.

Obecnie naukowcy potrafią badać pojedyncze komórki, w celu uzyskania głębszego wglądu w różnorodność ich populacje, tworzące organizm. Sekwencjonowanie RNA pojedynczych komórek (ang. *single-cell RNA sequencing*) jest obecnie powszechnie stosowane do ukazania heterogeniczności komórkowej oraz do identyfikacji nowych ich subpopulacji. Aby dowiedzieć się więcej o roli różnych typów komórek w regulacji funkcji tkanki lub narządu, zaproponowano kilka metod obliczeniowych rozwiązujących problem przypisywania typów komórkom.

Podczas gdy dane dla pojedynczych komórek nie zawierają informacji o położeniu komórek w tkance, niedawno opracowana technologia transkryptomiki przestrzennej (ang. *spatial transcriptomics*, ST) umożliwia zachowanie tej informacji. Niemniej jednak, ST składa się z pomiarów w wybranych, odrębnych punktach zawierających nieznaną liczbę zmieszanych komórek różnych typów. Konieczna jest dekompozycja tej mieszaniny w każdym badanym punkcie.

Wspomniane powyżej problemy: 1) przypisywanie typów w danych opisujących pojedyncze komórki oraz 2) dekompozycja mieszanin w danych przestrzennych są wymagające, o dużym znaczeniu i będące przedmiotem zainteresowania wielu naukowców. W tym projekcie proponujemy nową metodę obliczeniową, która jednocześnie będzie wykonywać oba zadania 1) i 2). Użyjemy modelowania probabilistycznego i statystycznego. Model zostanie efektywnie zaimplementowany i zweryfikowany w języku Python.

Naszym zdaniem jednoczesne wykonywanie zadań 1) i 2) pozwoli uzyskać jeszcze większą dokładność i zmniejszyć niepewność statystyczną w porównaniu do podejścia polegającego na ich oddzielnym wykonywaniu. W rezultacie przyczynimy się do rozwiązania dwóch problemów 1) i 2), a tym samym do lepszego zrozumienia funkcji heterogenicznych tkanek.

Ponadto, aby wykazać użyteczność modelu, będziemy analizować dane dotyczące dwóch typów tkanek: kory mózgowej myszy i piersi zaatakowanej nowotworem (w celu zbadania interakcji złośliwych guzów ze zdrowymi komórkami w ich sąsiedztwie).

Podsumowując, dostarczymy innowacyjne narzędzie do jednoczesnego przypisywania typów komórek w danych opisujących komórki sekwencjonowane pojedynczo oraz dekompozycji mieszaniny typów komórek w transkryptomicznych danych przestrzennych. Realizacja projektu przyczyni się do zrozumienia funkcjonowania heterogenicznych tkanek, w szczególności złożonej, wzajemnej interakcji guza i jego mikrośrodowiska immunologicznego w raku piersi. Co najważniejsze, ostatecznym i kluczowym rezultatem naszego projektu będzie łatwy w użyciu pakiet o otwartym kodzie źródłowym, zawierający implementację modelu w języku Python, który będzie można wykorzystać do analizy różnorodnych tkanek.