

Streszczenie

Tło: Jednym z głównych celów biologii systemów jest integracja danych omicznych wysokiej przepustowości z modelowaniem obliczeniowym, aby lepiej zrozumieć interakcje molekularne i mechanizmy komórkowe. Złożona natura sieci genomowych, transkryptomicznych i proteomicznych stawia istotne wyzwania, często ograniczając zdolność tradycyjnych modeli do uchwycenia dynamicznego zachowania systemów biologicznych. Dlatego zaawansowane podejścia analityczne są niezbędne do wyjaśnienia tych złożonych interakcji i identyfikacji kluczowych elementów regulacyjnych w sieciach biologicznych.

Cel: Niniejsze badanie ma na celu opracowanie kompleksowego potoku obliczeniowego do automatycznego pozyskiwania, porządkowania i analizy danych, ułatwiającego identyfikację kluczowych zmiennych w złożonych sieciach biologicznych. Wykorzystując dane multi-omics z repozytoriów takich jak Pathway Commons, AnimalTFDB i Genomics Data Commons (GDC), badanie ma na celu wyjaśnienie interakcji czynników transkrypcyjnych (TF) z ligandami-receptorami (L-R), białko-białko (P-P) i gen-gen (G-G). Zestaw danych raka jajnika z The Cancer Genome Atlas (TCGA) służy jako studium przypadku do walidacji podejścia, początkowo koncentrując się na połączeniu między szlakami *NF-κB* i **p53**, a następnie przeprowadzając dodatkowe analizy sieciowe szlaków **Cyklad Komórkowego** i **Sygnałowego MAPK** w celu identyfikacji kluczowych węzłów regulacyjnych.

Metodologia: Badanie przeprowadzono w serii ustrukturyzowanych etapów:

- 1. Pozyskiwanie danych:** Zastosowano potok oparty na Pythonie do ekstrakcji danych biologicznych na dużą skalę z baz danych, w tym Pathway Commons, AnimalTFDB i CellTalkDB. Biblioteki Pythona, takie jak *pandas* i *requests*, zostały wykorzystane do manipulacji danymi i automatycznego dostępu do API. Dane zostały przefiltrowane na podstawie odpowiednich identyfikatorów, takich jak KEGG i PubMed ID, aby zapewnić wysoką jakość interakcji do dalszej analizy.
- 2. Projekt bazy danych:** Zgromadzone zestawy danych zostały uporządkowane w zaprojektowanej relacyjnej bazie danych MySQL, aby ułatwić efektywne zarządzanie danymi i ich pobieranie. Schemat bazy danych przechwytywał różne typy interakcji biologicznych (np. TF-do-L-R, P-P, G-G) z relacjami klucza obcego zapewniającymi integralność danych.
- 3. Filtrowanie danych i konstrukcja sieci:** Przefiltrowane dane zostały wykorzystane do skonstruowania skierowanego grafu sieciowego, reprezentującego interakcje molekularne między genami, białkami i szlakami sygnałowymi. Wagi krawędzi zostały przypisane probabilistycznie, aby odzwierciedlić znaczenie każdej interakcji.
- 4. Wizualizacja sieci i wstępna analiza w Cytoscape:** Wstępną wizualizację sieci wykonano przy użyciu Cytoscape, umożliwiając eksplorację właściwości strukturalnych, takich jak współczynniki klastrowania, średnie długości ścieżek i rozkłady stopni. Ta analiza dostarczyła informacji o cechach topologicznych sieci, identyfikując regiony o wysokiej łączności i kluczowe węzły.
- 5. Zaawansowana analiza obliczeniowa w Pythonie:** Sieć została następnie przeanalizowana przy użyciu technik opartych na Pythonie w celu odkrycia krytycznych elementów regulacyjnych:
 - **Modelowanie sieci boolowskich** symulowało dynamikę regulacyjną w celu identyfikacji kluczowych węzłów i stanów stabilnych istotnych dla procesów decyzyjnych komórki.
 - **Algorytm PageRank** oceniał centralność węzłów, aby wyróżnić wpływowe elementy regulacyjne w sieci.
 - **Modelowanie losowych spacerów** identyfikowało kluczowe węzły poprzez symulację stochastycznego przepływu przez sieć.
 - **Rekurencyjne konwolucyjne sieci neuronowe (RCNNs)** wychwytywały zależności czasowe, aby przewidywać krytyczne elementy regulacyjne z danych szeregów czasowych, zwiększając zrozumienie dynamicznej sygnalizacji.

Aby zweryfikować framework, ten sam potok został zastosowany do szlaków **Cyklad Komórkowego i Sygnałowego MAPK**. Te szlaki zostały wybrane ze względu na ich kluczową rolę w progresji nowotworu i regulacji komórkowej. Analiza obejmowała symulacje losowych spacerów, modelowanie boolowskie, analizę PageRank, miary centralności i analizę motywów w celu identyfikacji kluczowych węzłów i ich ról funkcjonalnych.

Wyniki: Wieloetapowa analiza dostarczyła kompleksowych informacji o sieci raka jajnika. Kluczowe węzły regulacyjne, takie jak *NF-κB*, *p53*, *ATM* i *TNFR1*, okazały się centralne dla stabilności sieci. Konkretnie wyniki obejmują:

- **Modelowanie sieci boolowskich** ujawniło kluczowe węzły, takie jak *NF-κB*, *p53*, *IKKα* i *ATM*, wraz z istotnymi składnikami mRNA (*A20 mRNA*, *Wip1 mRNA*, *PTEN mRNA*).
- **Analiza PageRank** zidentyfikowała prominentne węzły, w tym *NF-κB*, fosforylowane *ATM* (*ATM-p*) i fosforylowane *Chk2* (*Chk2-p*).
- **Modelowanie losowych spacerów** podkreśliło węzły takie jak *p53*, *ATM* i *NF-κB*, wskazując na ich znaczenie w procesach takich jak odpowiedź na uszkodzenie DNA i apoptoza.
- **Analiza RCNN** uwydatniła elementy regulacyjne takie jak *TNFR1*, *IKKα* i białka związane z apoptozą (*Bax*, *p21*).

Analiza szlaków **Cyklad Komórkowego i Sygnałowego MAPK** dodatkowo potwierdziła słuszność frameworka. Kluczowe węzły zidentyfikowane w szlaku **Cyklad Komórkowego** obejmowały *Cyklina D*, *ATM/ATR*, *p53* i *CDK4/6*, które są kluczowe dla regulacji cyklad komórkowego i odpowiedzi na uszkodzenie DNA. W szlaku **Sygnałowym MAPK**, węzły takie jak *EGFR*, *RAS*, *MEK* i *ERK* zostały zidentyfikowane jako centralne dla proliferacji i różnicowania komórek. Te wyniki podkreślają potencjał tych węzłów jako celów terapeutycznych w leczeniu nowotworów.

Wnioski: To badanie posuwa dziedzinę do przodu, przedstawiając kompleksowy, integracyjny potok obliczeniowy do analizy i modelowania sieci, umożliwiając precyzyjną identyfikację elementów regulacyjnych. Framework dostarcza nowych informacji o regulacji genów, szlakach sygnałowych i potencjalnych celach terapeutycznych, z szerszymi implikacjami dla medycyny personalizowanej i modelowania chorób. Pomyślne zastosowanie potoku do szlaków **Cyklad Komórkowego i Sygnałowego MAPK** podkreśla jego wszechstronność i niezawodność w identyfikacji krytycznych węzłów w różnych sieciach biologicznych. Te wyniki sugerują, że zidentyfikowane węzły, takie jak *ATM/ATR*, *p53*, *EGFR* i *ERK*, powinny być priorytetowe w dalszych badaniach nad odkrywaniem leków ukierunkowanych na szlaki związane z nowotworami.

Słowa kluczowe: inżynieria biomedyczna, biologia systemów, modelowanie sieci, sieć boolowska, PageRank, losowy spacer, RCNN, regulacja genów, cele terapeutyczne, medycyna personalizowana.