



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Techniki wysokoprzepustowe [S1Bioinf1>TWP]

Przedmiot

Kierunek studiów
Bioinformatyka

Rok/Semestr
2/4

Studia w zakresie (specjalność)
–

Profil studiów
ogólnoakademicki

Poziom studiów
pierwszego stopnia

Język oferowanego przedmiotu
polski

Forma studiów
stacjonarne

Wymagalność
obligatoryjny

Liczba godzin

Wykład
30

Laboratorium
30

Inne (np. online)
0

Ćwiczenia
0

Projekty/seminaria
0

Liczba punktów ECTS

5,00

Koordynatorzy

dr hab. inż. Aleksandra Świercz
aleksandra.swiercz@put.poznan.pl

dr inż. Marcin Borowski
marcin.borowski@put.poznan.pl

dr inż. Paweł Wojciechowski
pawel.wojciechowski@put.poznan.pl

Wykładowcy

Wymagania wstępne

Student rozpoczynający ten moduł powinien posiadać podstawową wiedzę z zakresu biologii molekularnej, a także programowania i analizy bioinformatycznej sekwencji biologicznych. Powinien posiadać umiejętność rozwiązywania podstawowych problemów biologicznych i bioinformatycznych, testowania i poprawiania błędów w zaimplementowanych przez siebie programach oraz pozyskiwania informacji ze wskazanych źródeł i korzystania z baz danych. Ponadto w zakresie kompetencji społecznych student musi prezentować takie postawy, jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi.

Cel przedmiotu

1. Przekazanie studentom podstawowej wiedzy dotyczącej analizy danych pochodzących z sekwenatorów DNA, mikromacierzy oraz spektrometrów masowych, o podstawowych formatach danych uzyskiwanych z maszyn wysokoprzepustowych oraz zapoznanie z narzędziami do przetwarzania i analizy takich danych. 2. Rozwijanie u studentów umiejętności rozwiązywania problemów związanych z przetwarzaniem oraz analizą danych. 3. Kształtowanie u studentów umiejętności pracy zespołowej oraz samodzielności w rozwiązywaniu problemów.

Przedmiotowe efekty uczenia się

Wiedza:

1. Rozumie podstawowe zjawiska i procesy biologiczne, a ich interpretację opiera na podstawach empirycznych, wykorzystując metody matematyczne, w tym statystyczne
2. Zna podstawowe metody, techniki wysokoprzepustowe i narzędzia wykorzystywane w procesie rozwiązywania zadań bioinformatycznych, głównie o charakterze inżynierskim
3. Zna wybrane metody stosowane w biologii molekularnej, w tym metody wykorzystujące technologie wysokoprzepustowe
4. Zna trendy rozwojowe bioinformatyki

Umiejętności:

1. Potrafi pozyskiwać informacje z literatury, baz danych oraz innych właściwie dobranych źródeł, także w języku angielskim
2. Potrafi stosować podstawowe techniki i narzędzia informatyczne do rozwiązywania problemów biologicznych, oceniać ich przydatność
3. Pod kierunkiem opiekuna naukowego potrafi stosować metody analityczne, symulacyjne oraz eksperymentalne do formułowania i rozwiązywania zadań badawczych
4. Wie jak zastosować podstawowe metody statystyczne oraz algorytmy i techniki informatyczne do opisu procesów biologicznych i analizy danych
5. Potrafi dostrzegać systemowe i pozatechniczne aspekty podejmowanych zadań bioinformatycznych

Kompetencje społeczne:

1. Jest gotów do określania priorytetów służących realizacji zadania zdefiniowanego przez siebie lub innych
2. Rozumie, że w bioinformatyce wiedza i umiejętności bardzo szybko stają się przestarzałe - rozumie potrzebę uczenia się przez całe życie.
3. Potrafi współdziałać i pracować w grupie, przyjmując w niej różne role w szczególności podczas realizacji projektów informatycznych.

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca

a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:

- na podstawie aktywności w dyskusji na temat omawianego materiału;

b) w zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez:

- na podstawie oceny bieżącego postępu realizacji zadań;

Ocena podsumowująca

a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:

- kolokwium pisemne składające się z 5 pytań / zadań problemowych - każde zadanie punktowane 0-4 pkt (zadania mogą składać się z kilku podpunktów – za każdy podpunkt jest wówczas wyznaczona punktacja częściowa). Aby uzyskać zaliczenie należy zdobyć co najmniej 11 punktów. W przypadku nieobecności na więcej niż 1/3 odbytych wykładów zaliczenie wykładu będzie wymagało dodatkowo wykonania opracowania w formie pisemnej problemu z zakresu objętego tematyką wykładów, wskazanego przez prowadzącego i uzyskania pozytywnej oceny za to opracowanie

b) w zakresie laboratoriów / ćwiczeń weryfikowanie założonych efektów kształcenia realizowane jest przez:

- ocena końcowa stanowi średnią oceny z opracowania i przedstawienia publikacji naukowych oraz ocen za wykonanie poszczególnych ćwiczeń praktycznych. Za każde ćwiczenie / opracowanie można otrzymać maksymalnie 5 punktów (3 pkt = ocena dst; 4 pkt = ocena db, 5 pkt = ocena bdb)

Treści programowe

Program wykładu obejmuje następujące zagadnienia:

- * Podstawy analizy danych proteomicznych oraz lipidomicznych, zasad działania poszczególnych typów spektrometrów masowych, metod jonizacji cząsteczek, rodzajów detektorów i analizatorów jonów oraz danych jakie produkowane są w wyniku analizy próbek za pomocą metod spektrometrycznych.
- * Różne rodzaje mikromacierzy, przeprowadzenie analizy od pojedynczego punktu na mikromacierzy do macierzy ekspresji genów.
- * Różne rodzaje normalizacji danych, analiza statystyczna, klastrowanie genów oraz klasyfikacja próbek.
- * Sekwencjonowanie DNA, formaty plików wynikowych i podstawowe miary jakości sekwencjonowania.
- * Pojęcie genomu referencyjnego. Mapowanie odczytów do genomu referencyjnego z wykorzystaniem różnych programów. Analiza jakości mapowania i formaty plików wyjściowych.
- * Wykrywanie wariantów genetycznych (somatic and germline) - omówienie potoku przetwarzania danych, narzędzia do kontroli jakości, formaty plików wyjściowych.
- * Asemblacja de-novo - podstawowe pojęcia i algorytmy.
- * Wykrywanie wariantów strukturalnych - rodzaje wariantów, rodzaje metod i podstawowe narzędzia.
- * Analizy RNA-seq - założenia, narzędzia, formaty plików, sposoby normalizacji wyników, alternatywny splicing genów.
- * Sekwencjonowanie trzeciej generacji - omówienie technologii, podstawowych założeń, problemów i metod analizy danych. Porównanie metod sekwencjonowania krótkich i długich odczytów.

Tematyka zajęć

Program wykładu obejmuje następujące zagadnienia:

(1) Podstawy analizy danych proteomicznych oraz lipidomicznych, zasad działania poszczególnych typów spektrometrów masowych, metod jonizacji cząsteczek, rodzajów detektorów i analizatorów jonów oraz danych jakie produkowane są w wyniku analizy próbek za pomocą metod spektrometrycznych. (2) Różne rodzaje mikromacierzy. (3) Przeprowadzenie analizy od pojedynczego punktu na mikromacierzy do macierzy ekspresji genów. (4) Różne rodzaje normalizacji danych, analiza statystyczna, klastrowanie genów oraz klasyfikacja próbek. (5) Wprowadzenie do sekwencjonowania DNA. (6) Standardowe formaty plików wynikowych i podstawowe miary jakości sekwencjonowania. (7) Pojęcie genomu referencyjnego. Mapowanie odczytów do genomu referencyjnego z wykorzystaniem różnych programów. Analiza jakości mapowania i formaty plików wyjściowych. (8) Wykrywanie wariantów genetycznych (somatic and germline) - omówienie potoku przetwarzania danych, narzędzia do kontroli jakości, formaty plików wyjściowych. (9) Asemblacja de-novo - podstawowe pojęcia i algorytmy. (10) Wykrywanie wariantów strukturalnych - rodzaje wariantów, rodzaje metod i podstawowe narzędzia. (11). Analizy RNA-seq - założenia, narzędzia, formaty plików, sposoby normalizacji wyników, alternatywny splicing genów. (12) Sekwencjonowanie trzeciej generacji - omówienie technologii, podstawowych założeń, problemów i metod analizy danych. Porównanie metod sekwencjonowania krótkich i długich odczytów.

Ćwiczenia laboratoryjne prowadzone są w formie piętnastu dwugodzinnych zajęć odbywających się w laboratorium komputerowym. Pierwsze zajęcia przeznaczone są na zapoznanie studentów z zasadami użytkowania laboratorium i zaliczania ćwiczeń. Ćwiczenia realizowane są przez dwuosobowe zespoły studentów. Program zajęć laboratoryjnych obejmuje następujące zagadnienia:

(1) Analiza danych spektrometrycznych za pomocą popularnych i powszechnie używanych aplikacji komputerowych oraz baz danych. (2) Wykorzystanie biblioteki BioTools dla języka R do pisania dedykowanych skryptów analizy danych pochodzących z eksperymentów spektrometrycznych. (3) Analiza danych mikromacierzowych (4) Wykorzystanie różnych pakietów języka R, w tym Bioconductor w celu normalizacji danych, analizy ekspresji różnicowej oraz prezentacji wyników na wykresach. (5) Podstawowa analiza jakości danych sekwencjonowania - ocena jakości, filtrowanie. (6) Wykorzystanie programów do mapowania odczytów do genomu referencyjnego. Omówienie funkcjonalności programu samtools - konwersja typów, filtrowanie zmapowanych odczytów. (7) Wykorzystanie pakietu GATK do wykrywania wariantów genetycznych. Różnice między analizami somatic i germline. Filtrowanie wariantów, praca z plikami gvcf. (8) Wykrywanie wariantów strukturalnych - podstawowe algorytmy. (9) Asemblacja de-novo - przykład asemblacji i ocena jakości uzyskanego wyniku. (10) RNA-seq - przeprowadzenie przykładowego eksperymentu obliczeniowego, rodzaje normalizacji, detekcja genów dla których występuje istotna statystycznie różnica w ekspresji genów. (11) Sekwencjonowanie trzeciej generacji - formaty plików, podstawowe narzędzia oceny jakości, mapowania do genomu referencyjnego.

Metody dydaktyczne

Wykład ilustrowany prezentacją multimedialną zawierającą omawiane treści programowe, wzbogacone przykładami;

Laboratoria: ćwiczenia praktyczne z zakresu analizy danych , prezentacje, dyskusja, praca grupowa

Literatura

Podstawowa

1. N. Rodriguez-Ezpelta, M. Hackenberg, A.M. Aransay eds. „Bioinformatics for high throughput sequencing”, Springer, 2012

2. Proteomika i metabolomika, A. Drabik, A. Kraj, J. Silberring, Wydawnictwa Uniwersytetu Warszawskiego

3. Spektrometria mas, J. Charette , E. De Hoffmann , V. Stroobant, Wydawnictwa Naukowo Techniczne Uzupełniająca

<https://www.biorxiv.org/content/10.1101/020024v1>

<https://pubmed.ncbi.nlm.nih.gov/21653522/>

<https://pubmed.ncbi.nlm.nih.gov/21245279/>

<https://gatk.broadinstitute.org/hc/en-us/articles/360036194592-Getting-started-with-GATK4>

<https://gatk.broadinstitute.org/hc/en-us/articles/360035894731-Somatic-short-variant-discovery-SNVs-Indels->

<https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels->

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K, Thibault J, Chandran S, Whelan C, Lek M, Gabriel S, Daly MJ, Neale B, MacArthur DG, Banks E. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples bioRxiv, 201178. DOI: 10.1101/201178

Van der Auwera GA, Carneiro M, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella K, Altshuler D, Gabriel S, DePristo M. (2013). From FastQ Data to

High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr Protoc Bioinformatics, 43:11.10.1-11.10.33. DOI: 10.1002/0471250953.bi1110s43.

DePristo M, Banks E, Poplin R, Garimella K, Maguire J, Hartl C, Philippakis A, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell T, Kernytzky A, Sivachenko A, Cibulskis K, Gabriel S, Altshuler D, Daly M. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet, 43:491-498. DOI: 10.1038/ng.806.

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	125	5,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	60	2,50
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwium/egzaminu, wykonanie projektu)	65	2,50