

طرح درس بیوانفورماتیک پیشرفته (۲ واحد)

رشته: دکترای تخصصی بیوتکنولوژی کشاورزی

مدرس:

قادر میرزاقدری (ایمیل: gh.mirzaghadari@uok.ac.ir)، (عملی)

پیشنیاز: آشنایی با ژنتیک و آمار تک و چند متغیره و بیوشیمی

کد درس:

مقدمه: بیوانفورماتیک علمی بین رشته ای است که به تحلیل داده های مولکولی زیستی می پردازد. در چند دهه اخیر، پیشرفت در زمینه توالی یابی ماکرومولکول ها در موجودات مختلف حجم بسیار زیادی از داده های توالی و ساختار DNA، RNA، پروتئین و اطلاعات اپی ژنتیکی و غیره را به دست داده و در پایگاه های داده ذخیره شده است. بیوانفورماتیک دانشی نوظهور و بین رشته ای مرکب از علوم ژنتیک، کامپیوتر، آمار، ریاضی، بیوشیمی و بیوفیزیک است که به پردازش و تفسیر توالی، رفتار و عملکرد ماکرومولکول های زیستی از جمله مواردی چون ایجاد و جستجوی بانک های داده های زیستی، ژنومیک مقایسه ای، تحلیل چند شکلی های ژنتیکی، تحلیل بیان ژن ها و ساختار و عملکرد و تکامل پروتئین ها و روابط تکاملی آنها و بازسازی مسیرهای متابولیسی و غیره می پردازد.

اهداف درس: توانایی کار با سیستم عامل لینوکس (Linux) برنامه R برای اهداف بیوانفورماتیک، آشنایی با توالی ملکولهای اسیدنوکلئیک یا پروتئین مورد نیاز و توانایی استخراج آنها از پایگاه های داده مختلف. توانایی مشخص کردن محل یک ژن یا توالی خاص در ژنوم با استفاده از پایگاه های مرتبط و توانایی تعیین ویژگی های توالی ها. مشخص کردن دومین های پروتئینی در پروتئین های همولوگ، توانایی استخراج فایل pdb حاوی اطلاعات ساختار سه بعدی یک پروتئین و نشان دادن ساختارهای مربوطه، شناسایی پروموتورها، اینترون ها، اگزون ها و غیره. توانایی استفاده از Primer3 برای طراحی پرایمر مناسب برای مقاصد مختلف مانند بررسی بیان ژن. آشنایی با تحلیل داده های میکرواری و توانایی استخراج داده های میکرواری برای یک مطالعه بیولوژیکی مشخص، از پایگاه های داده مربوطه و توانایی تحلیل و استنتاج آن. توانایی بررسی مسیرهای انتقال پیام در سلول، مفهوم Gene Ontology و دسته بندی های آن، استفاده از داده پایگاه های مرتبط، روش های غنی سازی دسته های ژنی (Gene set Enrichment)، شناسایی و ترسیم شبکه میانگنش پروتئین-پروتئین مرتبط با یک دسته ژنی، آشنایی با اصول اولیه، انواع پلتفرم ها و نحوه انجام (Next Generation Sequencing (NGS)، توانایی تحلیل داده های NGS و توانایی بررسی میزان بیان ژن های مختلف و ایزوform های آنها، توانایی مشاهده الگوی بیان ژن های مربوط به داده های NGS در ابزار IGV. توانایی بررسی ارتباط بین چند شکلی نوکلئوتید (SNP) و بیان ژن. آشنایی با Perl، Python و Biopython و انجام آنالیزهای مقدماتی بر روی توالی با استفاده از Biopython.

منابع:

۱. لینک های معرفی شده در صفحه درس بیوانفورماتیک پیشرفته از جمله جزوات مربوط به R
۲. David Edwards (ed.), Plant Bioinformatics: Methods and Protocols, Methods in Molecular Biology, vol. 1374, Springer Science+Business Media New York 2016
۳. کتاب بیوانفورماتیک: راهنمای درسی و عملی. مؤلف: جانانان پوسنر، مترجمان: سعیدرضا وصال، محمد عارفیان، ناشر: دانشگاه فردوسی مشهد

ارزیابی بخش عملی (۱۰ نمره): حضور منظم (۱ نمره)، پروژه (۵ نمره)، امتحان پایان ترم (۴ نمره)

مقررات: ۱. دانشجوی باید حضور منظم در کلاس داشته باشد. ۲. غیبت بیش از دو جلسه باعث حذف درس می گردد. ۳. پروژه ها و تکالیف فقط در تاریخ تعیین شده تحویل گرفته می شود و بعد از آن نمره ای تعلق نمی گیرد. ۴. عدم ارائه هر پروژه به منزله نمره صفر برای آن پروژه می باشد.

برنامه کلاس عملی در طول ترم (برخی عناوین دو یا چند جلسه زمان خواهند برد):

۱	نصب لینوکس (Linux) و مرور دستورات پایه
۲	کار با لینوکس
۳	همتایابی و همردیفی چندگانه (Local and multiple alignment). فایل نیوویک
۴	نصب R و دستورات پایه، بردارها، جداول و لیست ها، رسم گراف های پایه در R
۵	کار با بسته های مرتبط با بیوانفورماتیک مثل pheatmap, ggtree, ggplot2 و ...
۶	جستجوی ژن، توالی اسید نوکلئیک و پروتئین در GeneBank و EnsemblPlant
۷	درخت فیلوژنی و بوت استریپینگ
۸	کار با دیتابیس ژنومی EnsemblePlan، آشنایی با سایت FTP و فایل های موجود مثل gtf و توالی های ژنومی مختلف
۹	استخراج توالی ژنهای خاص، فایل gtf و پروتئین آنها (UniGene, Gene, EST, EnsemblPlants) به صورت آفلاین
۱۰	بررسی توالی و ساختار پروتئین و شناسایی موتیف های کانسرو (NCBI, UniProt, InterPro, MeMe, Pfam, PDB)
۱۱	رسم ساختار ژنها و دومین های پروتئینی با استفاده TBtools و افزودن درخت فیلوژنی به آنها
۱۲	طراحی پرایمر با Primer 3
۱۳	آشنایی با داده های NGS و کار با آنها در لینوکس
۱۴	تحلیل داده های RNASeq: استخراج، کنترل کیفیت و ترمیم کردن، نمونه گیری ...
۱۵	تحلیل داده های RNASeq و بررسی بیان ژنها (الاینمنت و اسمبلینگ)، مشاهده الگوی بیان ژنها در ابزار IGV
۱۶	تحلیل داده های RNASeq و بررسی بیان ایزوفرم ها و رسم نمودارهای مربوطه در R
۱۷	آشنایی و کار با فایل های VCF، واریانت کالینگ و شناسایی SNPها از خوانش های ایلومینا
۱۸	استخراج داده های میکروآری مرتبط با یک آزمایش و تحلیل داده ها با R
۱۹	تحلیل داده های میکروآری با R و آنتولوژی ژنها
۲۰	آشنایی و کار با فایل های VCF، واریانت کالینگ و شناسایی SNPها از خوانش های ایلومینا
۲۱	مطالعه ارتباطی سراسرژنومی (Genome wide association study, GWAS)
۲۲	نصب Perl, Python و Biopython و انجام آنالیزهای مقدماتی بر روی توالی با استفاده از Biopython