**دانش پژوه گرامی لطفا موارد زیر را به دقت مطالعه کنید :**

تمامی سوالات را در همین دفترچه پاسخ دهید.

تمامی اعداد (فقط مقادیر پیوسته) را تا 3 رقم اعشار گرد کنید.

پاسخ های خود را خوش خط و خوانا فقط در محل طراحی شده مربوط به همان سوال بنویسید.

در صورتی که مواد و وسایل و ابزار های شما کامل نیست می توانید با بلند کردن **دست** تنها در 30 دقیقه اول آزمون آن را گزارش دهید. بعد از این مدت به درخواست شما رسیدگی نمی شود.

*****در 10 دقیقه پس از جمع آوری برگه های آزمون باید توالی های "seq22.fasta" و"seq24.fasta" و "seq25.fasta" و"seq26.fasta" را به این نشانی ایمیل کنید(نام خود را در subject ایمیل قرار دهید):**

**aliyzd1379@gmail.com**

**** ایمیل کردن لازم می باشد اما فایل ها باید روی لپ تاپ نیز ذخیره شده باشند.**

## مواد و وسایل و ابزار ها :

**کامپیوتر :**

- Online Bioinformatics Tools
- “Exam” folder on desktop

**در صورت نیاز می توانید از میان بر Ctrl + F (Find and Replace) استفاده کنید.**

**همراه خود دانش پژوه :**

- ماشین حساب مجاز
- ساعت (یا کرنومتر)
- روپوش آزمایشگاهی
- لوازم التحریر مورد نیاز



کد دانش پژوه (در این بخش چیزی ننویسید): [ ]

## بخش 1 : RNAseq (مجموعا 24 نمره)

تک سلولی یوکاریوتی تازه کشف شده MMI می تواند در غالب یکی از دو فنوتایپ A و B وجود داشته باشد.

مطالعه ی حاضر بر روی استفاده از تفاوت در الگوی بیان ژن برای تشخیص فنوتایپ این سلول تمرکز دارد.

در این راستا تعدادی سلول MMI به صورت رندم از جمعیت انتخاب شده و بعد از آماده سازی (روند معمول) مورد توالی یابی mRNA (RNAseq) قرار گرفتند.

به علت دقت پایین روش توالی یابی مورد استفاده، توالی های به دست آمده می توانند دارای خطا باشند.

توالی های حاصل از RNAseq هر نمونه (هر سلول) در پوشه Part 1 قرار گرفته اند.

**زیر بخش 1 :** هم اکنون می خواهیم با استفاده از ابزار های در دسترس، تعداد خوانش (read) های هر کدام از mRNA های متمایز را در هر کدام از نمونه ها (به عنوان ملاکی از بیان ژن مربوط به آن mRNA) مشخص کنیم. بدین منظور ابتدا باید mRNA های مجزایی که از آنها خوانش داشتیم را مشخص کنیم. سپس آنها را بر اساس طول مرتب کرده و شماره ای به هر کدام از آنها اختصاص می دهیم (بزرگترین mRNA مجزای شناخته شده mRNA 1 نام گذاری می شود).

**(مجموعا 24 نمره)**



کد دانش پژوه (در این بخش چیزی ننویسید): [ ]

**سوال 1** : تعداد خوانش های هر mRNA را در جدول زیر برای هر نمونه مشخص کنید. **(مجموعا 24 نمره)**

در صورت عدم وجود خوانش عدد صفر را گزارش کنید. روی ردیف مربوط به mRNA هایی که وجود ندارند خط بکشید.

(برای مثال در صورتی که mRNA 6 مجزا پیدا کردید روی ردیف های مربوط به 7 و 8 و 9 و 10 خط بکشید.)

در صورت نیاز می توانید به ردیف های جدول اضافه کنید. **هر خانه 0.2 نمره**

Sample

mRNAs

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | |
| 7 | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | |



کد دانش پژوه (در این بخش چیزی ننویسید): [ ]

## بخش 2 : Assembly (مجموعا 19 نمره)

در این سوال قطعات همپوشانی حاصل از برش یک توالی بلند در اختیار شما قرار گرفته است. این قطعات با روشی غیر ایده آل توالی یابی شده اند و ممکن است دارای تفاوتی هایی با توالی اصلی باشند. این قطعات ممکن است با بیش از دو قطعه دیگر همپوشانی داشته باشند. هدف شما پیدا کردن توالی یا توالی های اولیه است (طبعا دارای تعدادی جهش خواهد بود) و در این مسیر مجاز به استفاده از هر ابزار آنلاینی (به جز ارتباط با بقیه شرکت کننده ها 😊) می باشید.

- فایل های قطعات در پوشه "Part 2" با نام "whole.fasta" قرار دارد.

**سوال 1 :** حلقوی یا خطی بودن توالی ها را مشخص کنید. **4 نمره ، نمره منفی برابر**

| خطی / حلقوی |
| --- |
| |

**سوال 2 :** در کادر زیر الگو قرار گیری این 15 قطعه را نسبت به یکدیگر رسم کنید.

(وجود یا عدم وجود هم پوشانی بین هر دو قطعه واضح باشد.) **15 نمره**



کد دانش پژوه (در این بخش چیزی ننویسید): [ ]

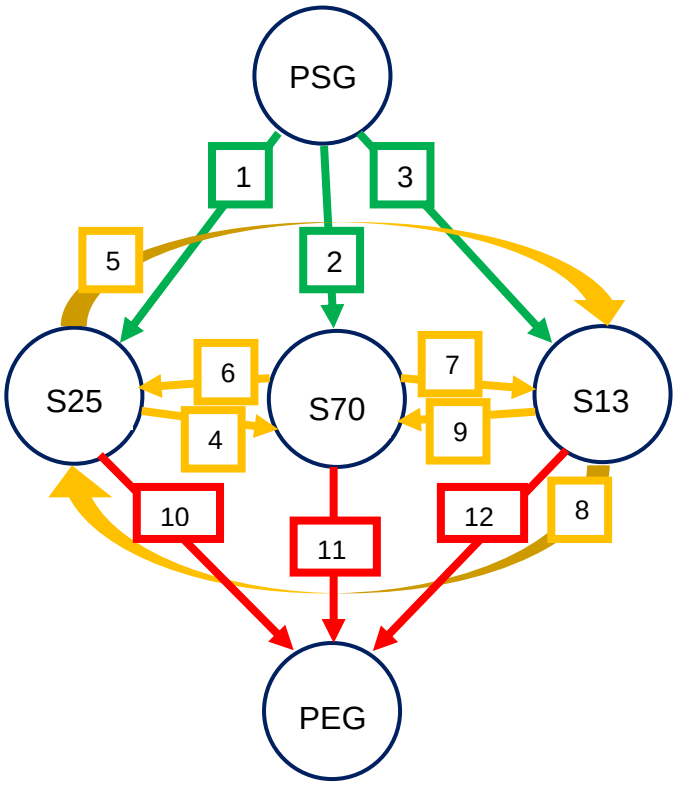
## بخش 3 : Neural pathway mapping (مجموعا 28 نمره)

مطالعه و شناسایی نقشه مسیر های عصبی (Connectomics) می تواند تاثیر بسزایی در درک ما از فرآیند های شناختی و رفتار موجودات زنده داشته باشد. شما در این مطالعه مسیر های احتمالی موجود در یک شبکه عصبی را بررسی خواهید کرد.

مطالعات انجام شده بر روی سیستم عصبی یک گونه یجانور ابتدایی نشان دهنده وجود دو گانگلیون اصلی ( primary start ganglion و primary end ganglion) و سه گانگلیون واسط (S25 و S70 و S13) در یک شبکه عصبی می باشند.

این مطالعات همچنین نشان دهنده ارتباط یک طرفه دو گانگلیون اصلی (از طریق گانگلیون های واسط) می باشند اما نحوه ارتباط بین این گانگلیون ها در این شبکه عصبی هنوز شناخته شده نیست.

تمام ارتباطات محتمل در شکل زیر نمایش داده شده اند :



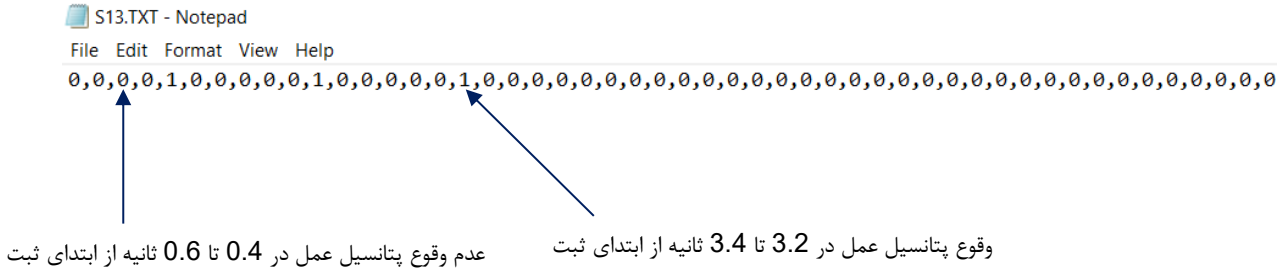

به منظور بررسی ارتباطات موجود در این شبکه عصبی در طی آزمایشی به صورت همزمان و در طی مدت ثابت از این گانگلیون ها ثبت عصبی می گیریم. سپس با تحلیل داده های به دست آمده مشخص می کنیم که آیا در هر بازه زمانی از ثبت هر یک از گانگلیون ها در حالت پتانسیل عمل (Spike) بوده اند یا خیر. تحلیل ها با فرکانس 5 هرتز انجام شده اند.

دیتای حاصل از آزمایش انجام شده در فایل های متنی فولدر Part 3 موجود می باشد.



کد دانش پژوه (در این بخش چیزی ننویسید): [ ]

به فرمت دیتای موجود توجه کنید :


S13.TXT - Notepad
File Edit Format View Help
0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,0

**سوال 2** : روش کار خود را برای پاسخ به سوال قبل به طور کامل توضیح دهید. **4 نمره**



کد دانش پژوه (در این بخش چیزی ننویسید): [ ]

## بخش 4 : Suitable input (مجموعا 17 نمره)

در صبح یک روز بهاری کار خود را به عنوان یک بیوانفورماتیسین با داده های توالی یابی DNA حاصل از دستگاه جدید آزمایشگاه تان شروع می کنید اما خروجی دستگاه شما را شگفت زده می کند !

فایل توالی خروجی این دستگاه در فایل seq21.txt در پوشه Part 4 در اختیار شما قرار گرفته است.

پس از جست و جو در اینترنت (کاری که باید قبل از خرید این دستگاه انجام می دادید :) متوجه چند اشتباه رایج در خروجی این دستگاه می شوید. این اشتباهات همیشه در خروجی این دستگاه دیده می شوند.

| توضیح | خروجی دستگاه | حالت واقعی |
|---|---|---|
| اضافه کردن کاراکتر های بی معنی | ATCLTAO G3#CWACTG | ATCTAGCACTG |
| گذاشتن N به جای باز هایی که مطمئن نیست | ATNTAGCACNG | ATCTAGCACTG |
| گذاشتن U به جای بعضی از باز های T | ATCUAGCACTG | ATCTAGCACTG |

**سوال 1 :** داده خروجی از این دستگاه را با هر ابزار آنلاینی به داده های واقعی تبدیل کنید. **8 نمره**

- حروف N باید بدون تغییر باقی بمانند.
- فایل خروجی باید به فرمت fasta باشد (عنوان توالی : ">final_seq") و آن را در پوشه فایل اولیه با نام seq22.fasta ذخیره کنید. (در فرمت fasta در هر خط نباید بیشتر از 80 کاراکتر قرار بگیرد).



کد دانش پژوه (در این بخش چیزی ننویسید): [ ]

**سوال 2 :** خروجی مرحله قبل حاصل از یک توالی متفاوت (یعنی جواب شما برای مرحله قبل الزاما با این فایل یکسان نیست) در فایل seq23.fasta در پوشه Part 4 در اختیار شما قرار داده شده است. **9 نمره**

هدف شما در این مرحله پیدا کردن یک ژن کامل پروکاریوتی در این توالیDNA است. این توالی هدف باید دارای ویژگی های زیر باشد :

1- بیش از 150 آمینو اسید طول داشته باشد.

2- با کدون ATG شروع شده باشد.

3- کدون های پایان مجاز تنها TAA و TAG است.

4- بقیه کدون ها از سیستم ترجمه standard استفاده می کنند.

5- در هر 6 فریم امکان وجود ORF وجود دارد. (در پاسخ سوال در صورت حضور ORF در فریم های reverse نسخه مکمل آورده شود(نسخه ای که با atg شروع می شود)).

توالی ژن هدف (از کدون آغاز تا آخر کدون پایان) را در پوشه فایل اولیه در فایل seq24.fasta قرار دهید.

توالی پلی پپتید حاصل از این ژن را با کد امینو اسیدی تک حرفی در پوشه فایل اولیه در فایل seq25.fasta قرار دهید.

توالی پلی پپتید حاصل از این ژن را با کد امینو اسیدی سه حرفی در پوشه فایل اولیه در فایل seq26.fasta قرار دهید.



کد دانش پژوه (در این بخش چیزی ننویسید): [ ]

## بخش 5 : Game (مجموعا 12 نمره)

پایگاه های داده پروتئین در تمامی دنیا در حال جمع آوری اطلاعات مختلف از پروتئین ها هستند. این اطلاعات شامل توالی، اطلاعات عملکردی و ساختار 3 بعدی پروتئین ها می باشد.

ساختار سه بعدی پروتئین ها به روش های مختلفی از جمله کریستالوگرافی اشعه ایکس و NMR به دست می آید که هر کدام مزایای منحصر به فرد خود را دارند.

این حجم وسیع از اطلاعات ساختار سه بعدی به همراه توالی های متناظر آنها بستر مناسبی برای به کار بردن ابزار های هوش مصنوعی برای توسعه مدل های پیش بینی ساختار سه بعدی از روی توالی پلی پپتید ها بوجود می آورد.

**سوال 1 :** در مورد روش های کسب اطلاعات از پروتئین ها، گزاره های درست و نادرست را مشخص کنید. **مجموعا 4 نمره**

**هر مورد 1 نمره نمره منفی برابر**

| | صحیح | غلط |
|---|---|---|
| پیچیدگی محاسباتی انجام روش NMR برای کشف ساختار سه بعدی پروتئین ها با افزایش تعداد آمینو اسید ها به صورت خطی افزایش می یابد. | | |
| روش کریستالوگرافی اشعه ایکس روش مناسب تری برای پپتید های کوتاه است. | | |
| فایل pdb حاصل از کریستالوگرافی اشعه ایکس دارای اطلاعات موقعیتی مولکول های آب درون کریستال می باشد. | | |
| Blastx از روی توالی ترجمه شده DNA داخل داده های توالی پروتئین انجام می شود. | | |



کد دانش پژوه (در این بخش چیزی ننویسید): [ ]

**سوال 2 :** در این بخش شما با یکی از بازی های موجود در سایت Pdbj.org بازی خواهید کرد! **8 نمره**

روش بازی :

در ابتدا وارد لینک زیر شوید (تایپ کنید):

https://numon.pdbj.org/games/pelmanism-fold-3x6s_ja.html

در این بازی به شما 18 عدد ساختار سه بعدی پروتئین داده شده است که دو به دو متعلق به یک خانواده از پروتئین ها می باشند. شما باید این 9 جفت ساختار را با حداقل تعداد تلاش پیدا کنید.

شما مجاز به تکرار این بازی به هر تعداد بار هستید اما تنها یک بار اجازه درخواست از مسئول آزمایشگاه برای عکس برداری از صفحه مرورگر خود را دارید.

**صفحه سایت:**

نحوه نمره دهی :

نمره شما = 16 – (Number of times – 9) * 0.5

نمره دهی این قسمت به صورت مقایسه ای بین شرکت کنندگان خواهد بود.



کد دانش پژوه (در این بخش چیزی ننویسید): [ ]

پایان،

این مورد رو فراموش نکنید:

*****در 10 دقیقه پس از جمع آوری برگه های آزمون باید توالی های "seq22.fasta "و"seq24.fasta "و "seq25.fasta" و"seq26.fasta " را به این نشانی ایمیل کنید(نام خود را در subject ایمیل قرار دهید):**

**aliyzd1379@gmail.com**

**** ایمیل کردن لازم می باشد اما فایل ها باید روی لپ تاپ نیز ذخیره شده باشند.**



کد دانش پژوه (در این بخش چیزی ننویسید): [ ]