

به نام خدا

تمرین اول

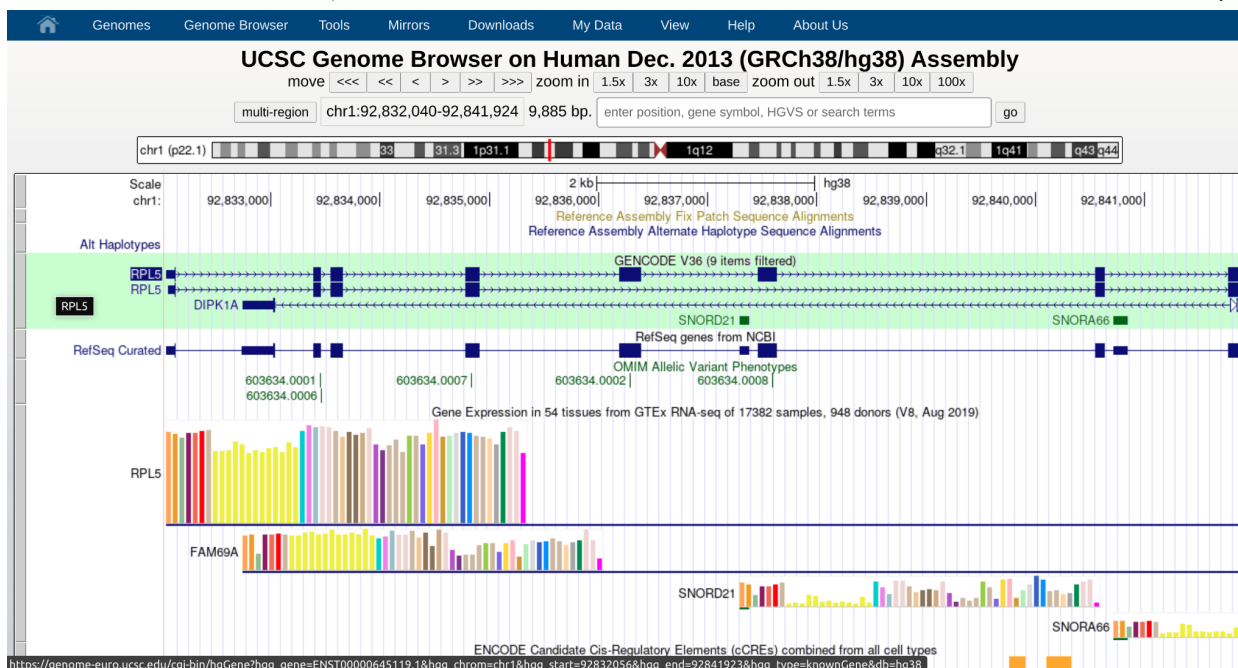
مهدی کافی ۹۹۲۱۰۷۵۳

مسئله ۱

در این مسئله یک رشته ورودی از کابر گرفته می‌شود (البته باید توجه کرد در صوتیکه کل اسکریپت به صورت یکجا اجرا شود، رشته DNA برابر رشته صورت سوال می‌شود و اگر بخواهیم که رشته جدیدی وارد کنیم باید خط شامل **readline** را به تنهایی اجرا کنیم) سپس از این رشته، رشته **reverse complement** آن را با استفاده از تابع **reverse_complement** می‌سازیم؛ سپس با داشتن **codon**ها و با استفاده از تابع **dna_to_amino_acid** هر دو رشته اصلی و مکمل را پیمایش می‌کنیم تا به کدن شروع برسیم و در صورت دیدن این کدن پنجره با قدم‌هایی به طول ۳ جلو رفته و هر بار یک آمینو اسید به رشته آمینو اسید اضافه می‌کند و سپس پس از رسیدن به کدن **stop**، رشته ساخته‌شده را در وکتور **aa** می‌ریزد و در نهایت، هر عضو یکتای این وکتور را چاپ می‌کند.

مسئله ۲

الف) با جستجوی سمبل **RPL5** در سایت **UCSC Genome Browser** به صفحه زیر میرسیم.



ب)

- کتابخانه **BSgenome**: این کتابخانه برای کار با ژنوم طراحی شده است و شامل تعداد زیادی از ژنوم‌ها است و بر پایه کتابخانه **BioStrings** ساخته شده است که برای کار با رشته‌های بسیار طولانی است.
- کتابخانه **org.Hs.eg.db**: این کتابخانه یک پایگاه داده شامل **annotation**ها بر کل طول ژنوم انسان است. از این پایگاه داده می‌توانیم برای تبدیل کردن اسامی مختلف یک ژن و یا سمبل‌های آن استفاده کنیم.
- کتابخانه **TxDb**: این کتابخانه یک پایگاه داده شامل **annotation**های تولید شده توسط **UCSC** است و همینطور اطلاعات ترنسکریپت‌ها در این پایگاه داده قرار دارد که می‌توانیم روی آن‌ها کوئری بزنیم.

ج) برای این بخش در ابتدا سمبل را از ورودی گرفته، به طور مثال سمبل RPL5 و سپس با استفاده از کتابخانه org.Hs.eg.db این سمبل را به ENTREZID یا همان GENEID در کتابخانه TxDB تبدیل می‌کنیم، سپس با استفاده از این GENEID و کتابخانه TxDB تمامی ترسکریپت‌های این ژن را به دست می‌آوریم و چاپ می‌کنیم.

برای بخش بعدی در ابتدا با استفاده از کتابخانه TxDB کروموزومی که ژن مورد نظر بر روی آن قرار دارد را پیدا می‌کنیم و با استفاده از کتابخانه BSgenome، طول این کروموزوم را به دست می‌آوریم. بار دیگر با استفاده از کتابخانه TxDB، نقاط شروع و پایان اگزون‌های این ژن را به دست می‌آوریم و سپس با محاسبه فاصله بین بزرگترین نقطه پایان و کوچکترین نقطه شروع، طول ژن را محاسبه می‌کنیم (به طور مثال با این روش طول محاسبه شده برای ژن RPL5 با مقدار گزارش شده در سایت UCSC برابر شد). سپس نسبت طول ژن به کروموزوم را محاسبه می‌کنیم.

برای رسم با کتابخانه genemodel نیاز داریم که نقاط شروع و پایان اگزون‌ها و اینترون‌ها را داشته باشیم و این فواصل باید فاقد همپوشانی باشند، به این منظور با داشتن نام ترسکریپت‌ها و دو تابع exonsByTranscript و intronsByTranscript می‌توانیم این بازه‌ها را بدون همپوشانی به دست بیاوریم. نکته‌ای که اینجا وجود دارد این است که برای هر ژن ممکن است چند مجموعه اگزون و اینترون با بازه‌های متفاوت وجود داشته باشد که این دو تابع تمام این مجموعه‌ها را به ما می‌دهند و به طور مثال برای ژن RPL5، چهار مجموعه اگزون و اینترون وجود دارد و در سایت UCSC نیز در چند خط مجزا این مجموعه‌ها رسم شده‌اند. سپس کافی است برای دادن این بازه‌ها به کتابخانه genemodel، این بازه‌ها را به فواصل بین ۱ تا آخر 3utr ببریم و برای دو بخش 5utr و 3utr بازه‌های ۱۰۰ نوکلئوتیدی در نظر گرفتیم و این نمودارها را با استفاده از genemodel رسم کردیم که با استفاده از جابه‌جا کردن نمودارها در RStudio قابل مشاهده هستند و همین‌طور فایل آن‌ها آپلود شده‌است. به عنوان مثال یکی از نمودارها در زیر آمده‌است.

