

عنوان پروژه: تحلیل و بصری سازی تغییرات تعداد کپی (CNV) در ژنوم انسان

تعداد افراد: ۳ نفر

تغییرات تعداد کپی (Copy Number Variation یا CNV) به تغییرات در تعداد کپی های یک بخش خاص از DNA در ژنوم اشاره دارد. این تغییرات می توانند شامل حذف، تکرار یا تغییرات ساختاری دیگر در قطعات بزرگ تری از DNA باشند و می توانند تأثیرات قابل توجهی بر روی بیان ژن ها و عملکرد سلولی داشته باشند.

CNV ها می توانند در سطح فردی متفاوت باشند و برخی از آن ها ممکن است با بیماری ها و اختلالات ژنتیکی مرتبط باشند. برای مثال، برخی از تحقیقات نشان داده اند که CNV ها می توانند نقش مهمی در بیماری هایی مانند اوتیسم، اسکیزوفرنی و برخی از انواع سرطان ایفا کنند.

امروزه روش های مدرن مانند توالی یابی نسل جدید (NGS) و روش های هیبریداسیون میکروآرایه به پژوهشگران این امکان را می دهند که CNV ها را به طور دقیق شناسایی و تحلیل کنند. با توجه به اینکه CNV ها می توانند تأثیرات زیادی بر روی تنوع ژنتیکی و تکامل موجودات داشته باشند، مطالعه آن ها در زمینه های مختلف زیست شناسی، پزشکی و داروسازی اهمیت ویژه ای دارد.

به طور کلی، تغییرات تعداد کپی به عنوان یک عامل کلیدی در درک پیچیدگی ژنوم انسان و تنوع زیستی مورد توجه قرار گرفته و همچنان موضوع تحقیقاتی فعال در علم ژنتیک می باشد.

یک پروژه جذاب برای شما با تمرکز بر داده های Copy Number Variation (CNV) می تواند شامل تحلیل و بصری سازی این داده ها باشد.

اهداف پروژه:

۱. تحلیل داده ها: پیاده سازی الگوریتم های تحلیل داده ها برای شناسایی الگوها و ارتباطات بین CNV ها و ویژگی های خاص (مانند بیماری ها).
۲. بصری سازی داده ها: طراحی ابزارهای بصری برای نمایش نتایج به دست آمده، مانند نقشه های حرارتی، نمودارهای توزیع و غیره.
۳. گزارش نهایی: نوشتن گزارشی جامع که شامل یافته ها، تحلیل ها و بصری سازی ها باشد.

مراحل پروژه:

۱. آشنایی با CNV: شما باید با مفهوم CNV و اهمیت آن در علم ژنتیک آشنا شوید.
 ۲. پیش پردازش داده ها: استفاده از زبان برنامه نویسی R برای آماده سازی CNV ها.
 ۳. تحلیل داده ها:
- انجام تحلیل های آماری برای شناسایی CNV های خاص مرتبط با بیماری ها.
 - استفاده از الگوریتم های آماری یا یادگیری ماشین برای پیش بینی ارتباط بین CNV ها و بیماری هدف.
۴. بصری سازی:
- استفاده از کتابخانه هایی R برای ایجاد نمودارها و گراف ها.

۵. گزارش نهایی: تدوین یک گزارش علمی شامل مقدمه، روش‌شناسی، نتایج و بحث.

ابزارها و تکنولوژی‌ها:

- زبان برنامه‌نویسی: R

توضیح پروژه:

مجموعه داده‌های موجود، حاصل از یک مطالعه Case و Control روی یک نوع خاص سرطان می‌باشد. مجموعه Case افرادی هستند که بیماری را دارند. مجموعه کنترل افرادی هستند که بیماری را ندارند. دقت کنید که وجود یک CNV در بدن انسان ممکن است:

۱. هیچ نمود بیرونی نداشته باشد،
۲. یا یک ویژگی مثلاً رنگ چشم فرد را مشخص کند،
۳. یا باعث بیماری شود.

در فایل کیس، CNVهای افراد سرطانی گزارش شده است.

در فایل کنترل، CNV های افراد سالم گزارش شده است.

در این پروژه از شما خواسته شده است

۱. مناطقی از کروموزوم را که فکر می‌کنید وجود variation در آنها باعث بروز بیماری می‌شود شناسایی کنید. برای این کار می‌توانید از تست‌های اماری یا روش‌های یادگیری ماشین استفاده کنید. در گزارش، روش انتخاب مناطق را توضیح دهید. (این مناطق باید بروز و ظهور بیشتری در کیس و بروز کمتری در مجموعه کنترل‌ها داشته باشد.)

۲. همچنین قایل مربوط به ژن‌ها نیز در اختیار شما قرار گرفته است. یکی دیگر از کارهایی که باید انجام دهید آن است که ژن‌هایی که variation در بخشی از آنها میتواند عامل بروز بیماری باشد را گزارش کنید.

مجموعه داده‌ها، شامل شماره کروموزوم، ابتدا و انتهای variation، نوع variation و شناسه بیمار است.

تحویل دادنی‌ها:

۱. گزارش شامل توضیح روش استخراج مناطق و نمودارها
۲. کد به زبان برنامه‌نویسی R
۳. فایل اکسل شامل نتایج در قالب دو برگه مختلف، شامل ژن‌ها و مناطق