Uygulama Ödevinde Yapılanlar

Genomik hesaplama dersinin ikinci ödevi için paylaşılan dosyalar incelendi. Ardından ödev için önerilen dosyanın indirilmesi için okul mail adresi ile COSMIC web sayfasına üye olunmuştur. Data > Download sekmesinde bulunan "All Mutations in Census Genes" dosyası işlenmek üzere indirildi.

Ardından indirilen dosya işlenmek üzere okundu ve tüm kayıtlar gösterilmek üzere bir txt uzantılı dosyaya yazıldı. TSV dosyasını okurken python dilinde bulundan pandas kütüphanesinden yaralanılmıştır. Pandas kütüphanesinde bulunan csv dosya okuma fonksiyonuna ilk parametre olarak dosya verilmiştir ; ikinci parametrede dosyamızda yer alan verilerin birbirinden ayrılış şekli tab belirtilmiştir.

Bu işlemler sonrası ödevin birinci sorusu cevaplandırıldı. TSV dosyasında hastaya ait numunenin sınıflandırılmasını gösteren kolon "primary site" isimli kolondur. Bu kolona sorgu uygulamadan önce sorun teşkil edebilecek kolon isimlerindeki boşluklar, '_' ile değiştirildi. Bunu yaparken okunan dosyamızın kolon isimlerini bir for döngüsünde işlendi. Ardından pandas kütüphanesinin sağlamış olduğu query fonksiyonunu kullanarak liver carcinoma içeren kolonları sorgulandı. Aynı hastalar tekrarlı gösterilmemesin diye duplicate verilerin süzülmesi için hazır olan drop_duplicates fonksiyonu ile süzme işlemi yapılarak tekrarlar kaldırıldı. Ardından görsel olarak karmaşıklığı önlemek için sadece hasta id kolonu olan ID_sample verileri dosyaya kaydedildi . Sorunun ikinci aşaması olarak istenen bu hastaların mutasyonlu genlerinin listesi çıkarılması ve bir dosyaya yazılması işlemi gen ve hasta id gurplandırması ile yapıldı ve hasta id değerine göre sıralı olarak gerçekleştirildi.

İkinci soruda istenen şekilde liver carcinomada en çok görülen 30 gen ve yüzdeleri bir dosyaya yazıldı. Dosyaya istenenlere ek olarak en çok görülen geni belirlerken kullanılan genin görülme sayıları kolon olarak eklendi.

Üçüncü soruda istenildiği gibi cevap 1 ve cevap 2 fonksiyonları skin carcinoma içinde çalıştırıp sonuçlar dosyaya kaydedildi.

Dördüncü soru için skin ve liver carcinoma hastalıklarına ait ortak genler dosyaya yazıldı. Yazma işleminde tekrarlı veri olmaması sağlandı.

Beşinci soru için 3 hastalığın hepsinde bulunan ortak mutasyon genleri bulundu ve ardından dosyaya yazıldı.

Ek olarak ; hastalıklar ve hasta sayı dağılımları analiz edilip dosyaya yazıldı, hastalıklar ve yaş ortalamaları analiz edilip dosyaya yazıldı , liver carcinoma ve skin carcinoma hastalığı görünenler arasında en küçük yaşa sahip , en büyük yaşa sahip ve yaş ortalaması analiz edilip dosyalara yazıldı , mutasyonlu genler ve görülme sayıları gruplandırılarak dosyaya yazıldı , tüm hastalar ve bu hastalardaki toplam mutasyonu gen sayısı bulunup dosyaya yazıldı.

Uygulama Ödevine Ait Program Kodunun Tamamı

```
import pandas as pd

class TSV():

    def __init__(self):
        self.dosya_adi = "CosmicMutantExportCensus.tsv"
        self.dosya = pd.read_csv(self.dosya_adi , sep="\t")
        pd.set_option('display.max_rows' , self.dosya.shape[0]+1)
```

```
    def tsv_dosya_yaz(self,isim):
        yazilacak_dosya = open(r'C:\\Users\ASUS\Desktop\Uygulama_Odevi_2\\'+ isim +'.txt', "w")
        yazilacak_dosya.write(str(self.dosya))

    def analiz_dosya_yaz(self,sonuc,isim):
        yazilacak_dosya = open(r'C:\\Users\ASUS\Desktop\Uygulama_Odevi_2\\'+ isim +'.txt', "w")
        yazilacak_dosya.write(str(sonuc))

    def hasta_id_getir(self,primary_site , primary_histology,hastalik_adi):
        self.dosya = pd.read_csv(self.dosya_adi, sep="\t")
        self.dosya.columns = [column.replace(" ", "_") for column in self.dosya.columns]
        self.dosya.query("Primary_site == '" + primary_site + "'", inplace=True)
        self.dosya.query("Primary_histology == '" + primary_histology + "'", inplace=True)
        hastalik = self.dosya.drop_duplicates(subset=['ID_sample'])
        pd.set_option('display.max_rows', hastalik.shape[0] + 1)
        self.analiz_dosya_yaz(hastalik[['ID_sample']],hastalik_adi+"_hasta_id")
        return float(hastalik[['ID_sample']].count())

    def hasta_mutasyonlu_gen_listesi(self,primary_site,primary_histology,hastalik_adi):
        self.dosya = pd.read_csv("CosmicMutantExportCensus.tsv", sep="\t")
        self.dosya.columns = [column.replace(" ", "_") for column in self.dosya.columns]
        self.dosya.query("Primary_site == '" + primary_site + "' & Primary_histology == '" +
primary_histology + "'", inplace=True)
        mutasyonlu_gen_listesi = self.dosya[['ID_sample', 'GENOMIC_MUTATION_ID']]
        mutasyonlu_gen_listesi=mutasyonlu_gen_listesi.sort_values(by='ID_sample')
        self.analiz_dosya_yaz(mutasyonlu_gen_listesi,hastalik_adi+"_hastalarinin_mutasyonlu_genleri")

    def gen_bul(self,primary_site,primary_histology):
        self.dosya = pd.read_csv("CosmicMutantExportCensus.tsv", sep="\t")
        self.dosya.columns = [column.replace(" ", "_") for column in self.dosya.columns]
        self.dosya.query("Primary_site == '" + primary_site + "'", inplace=True)
        self.dosya.query("Primary_histology == '" + primary_histology + "'", inplace=True)
        self.dosya = self.dosya.drop_duplicates(subset=['Gene_name'])
        sonuc = self.dosya[['Gene_name']]
        return sonuc

    def ortak_gen_bul(self,primary_site_1,primary_histology_1 ,primary_site_2,primary_histology_2):
        hastalik_1 = self.gen_bul(primary_site_1,primary_histology_1)
        hastalik_2 = self.gen_bul(primary_site_2, primary_histology_2)
        birlesim_liste = pd.concat([hastalik_1, hastalik_2])
        birlesim_liste["duplicated"] = birlesim_liste.duplicated()
        birlesim_liste.query("duplicated == True", inplace=True)
        self.analiz_dosya_yaz(birlesim_liste[['Gene_name']],"ortak_gen_" + primary_site_1 +
primary_histology_2 + "&" + primary_site_2+primary_histology_2)

    def en_cok_gorunen_gen_yuzde(self,toplam_hasta_sayisi,primary_site,primary_histology ):
        dosya_1 = pd.read_csv("CosmicMutantExportCensus.tsv", sep="\t")
        dosya_1.columns = [column.replace(" ", "_") for column in dosya_1.columns]
        dosya_1.query("Primary_site == '" + primary_site + "'", inplace=True)
        dosya_1.query("Primary_histology == '" + primary_histology + "'", inplace=True)
        dosya_1.drop_duplicates(subset=['ID_sample'])
        dosya_1 =
dosya_1.groupby(['Gene_name'])['ID_sample'].count().reset_index(name="Hasta_Sayisi")

        dosya = pd.read_csv("CosmicMutantExportCensus.tsv", sep="\t")
        dosya.columns = [column.replace(" ", "_") for column in dosya.columns]
        dosya.query("Primary_site == '" + primary_site + "'", inplace=True)
        dosya.query("Primary_histology == '" + primary_histology + "'", inplace=True)
```

```
        dosya =
dosya.groupby(['Gene_name'])['ID_sample'].count().sort_values(ascending=False).head(30).reset_index(
name="Gorulme_Sayisi")

        dosya['Yüzde'] = dosya_1['Hasta_Sayisi'].mul(100)
        dosya['Yüzde'] = dosya['Yüzde'].divide(other=toplam_hasta_sayisi)
        self.analiz_dosya_yaz(dosya , primary_site + primary_histology +
"_en_cok_gorunen_30_gen_ve_yuzdeleri")

    def hastalik_hasta_sayi_dagilimi(self):
        self.dosya = pd.read_csv("CosmicMutantExportCensus.tsv", sep="\t")
        self.dosya.columns = [column.replace(" ", "_") for column in self.dosya.columns]
        self.dosya = self.dosya.drop_duplicates(subset=['ID_sample'])

sonuc=self.dosya.groupby(['Primary_site'])['Primary_site'].count().sort_values(ascending=False).reset_in
dex(name="Hasta_Sayisi")
        self.analiz_dosya_yaz(sonuc,"hastalik_hasta_sayisi")

    def hasta_mutasyon_sayisi(self):
        self.dosya = pd.read_csv("CosmicMutantExportCensus.tsv", sep="\t")
        self.dosya.columns = [column.replace(" ", "_") for column in self.dosya.columns]
        sonuc =
self.dosya.groupby(['ID_sample'])['ID_sample'].count().sort_values(ascending=False).reset_index(name=
"Mutasyon_Sayisi")
        self.analiz_dosya_yaz(sonuc,"hasta_mutasyon_sayilari")

    def hastalik_yas_ortalama(self):
        dosya = pd.read_csv("CosmicMutantExportCensus.tsv", sep="\t")
        dosya=dosya.drop_duplicates(subset=['ID_sample'])
        dosya=dosya.groupby(['Primary site'])['Age'].mean().reset_index(name="Yas_Ortalamasi")
        self.analiz_dosya_yaz(dosya,"hastalik_yas_ortalamalari")

    def mutasyonlu_gen_isimleri_gorulme_sayilari(self):
        dosya = pd.read_csv("CosmicMutantExportCensus.tsv", sep="\t")
        dosya =
dosya.groupby(['GENOMIC_MUTATION_ID'])['ID_sample'].count().sort_values(ascending=False).reset
_index(name="Görülme_Sayisi")
        self.analiz_dosya_yaz(dosya, "mutasyonlu_genler_ve_sayilari")

    def hastalik_yas_incelemesi(self,primary_site,primary_histology):
        self.dosya = pd.read_csv("CosmicMutantExportCensus.tsv", sep="\t")
        self.dosya.columns = [column.replace(" ", "_") for column in self.dosya.columns]
        self.dosya.query("Primary_site == '" + primary_site + "' & Primary_histology == '" +
primary_histology + "'" ,inplace=True)
        sonuc ="En Yüksek Yaş : " + str(self.dosya['Age'].max())+"\nEn Düşük Yaş : " +
str(self.dosya['Age'].min())+"\nOrtalama Yaş : " + str(self.dosya['Age'].mean())
        self.analiz_dosya_yaz(sonuc , primary_site+primary_histology+"_yas_incelemesi")

    def ortak_mutasyonlu_genler_3_hastalik(self):
        liver_dosya = pd.read_csv("CosmicMutantExportCensus.tsv", sep="\t")
        liver_dosya.columns = [column.replace(" ", "_") for column in liver_dosya.columns]
        liver_dosya.query("Primary_site == 'liver'", inplace=True)
        liver_dosya.query("Primary_histology == 'carcinoma'", inplace=True)
        liver = liver_dosya.drop_duplicates(subset=['GENOMIC_MUTATION_ID'])
        liver_listesi = liver[['GENOMIC_MUTATION_ID']]

        skin_dosya = pd.read_csv("CosmicMutantExportCensus.tsv", sep="\t")
        skin_dosya.columns = [column.replace(" ", "_") for column in skin_dosya.columns]
```

```
        skin_dosya.query("Primary_site == 'skin'", inplace=True)
        skin_dosya.query("Primary_histology == 'carcinoma'", inplace=True)
        skin = skin_dosya.drop_duplicates(subset=['GENOMIC_MUTATION_ID'])
        skin_listesi = skin[['GENOMIC_MUTATION_ID']]

        malignant_dosya = pd.read_csv("CosmicMutantExportCensus.tsv", sep="\t")
        malignant_dosya.columns = [column.replace(" ", "_") for column in malignant_dosya.columns]
        malignant_dosya.query("Primary_site == 'skin'", inplace=True)
        malignant_dosya.query("Primary_histology == 'malignant_melanoma'", inplace=True)
        malignant = malignant_dosya.drop_duplicates(subset=['GENOMIC_MUTATION_ID'])
        malignant_listesi = malignant[['GENOMIC_MUTATION_ID']]

        birlesim_liste = pd.concat([liver_listesi, skin_listesi])
        birlesim_liste["duplicated"] = birlesim_liste.duplicated()
        birlesim_liste.query("duplicated == True", inplace=True)
        birlesim_liste = birlesim_liste[['GENOMIC_MUTATION_ID']]

        birlesim_liste_2 = pd.concat([malignant_listesi, birlesim_liste])
        birlesim_liste_2["duplicated"] = birlesim_liste_2.duplicated()
        birlesim_liste_2.query("duplicated == True", inplace=True)
        pd.set_option('display.max_rows', birlesim_liste_2.shape[0] + 1)
        sonuc = birlesim_liste_2[['GENOMIC_MUTATION_ID']]
        self.analiz_dosya_yaz(sonuc , "malignant_skin_liver_ortak_mutasyonlu_genleri")

def main():
    analiz = TSV()
    analiz.tsv_dosya_yaz("tum_dosya")
    skin_sayisi = analiz.hasta_id_getir("skin", "carcinoma" , "skin_carcinoma")
    liver_sayisi = analiz.hasta_id_getir("liver", "carcinoma" , "liver_carcinom")
    analiz.hasta_mutasyonlu_gen_listesi("liver", "carcinoma" , "liver_carcinom")
    analiz.hasta_mutasyonlu_gen_listesi("skin", "carcinoma" , "skin_carcinoma")
    analiz.ortak_gen_bul("liver", "carcinoma" ,"skin", "carcinoma" )
    analiz.en_cok_gorunen_gen_yuzde(skin_sayisi,"skin", "carcinoma")
    analiz.en_cok_gorunen_gen_yuzde(liver_sayisi,"liver", "carcinoma")
    analiz.hastalik_hasta_sayi_dagilimi()
    analiz.hasta_mutasyon_sayisi()
    analiz.hastalik_yas_ortalama()
    analiz.mutasyonlu_gen_isimleri_gorulme_sayilari()
    analiz.hastalik_yas_incelemesi("liver", "carcinoma")
    analiz.hastalik_yas_incelemesi("skin", "carcinoma")
    analiz.ortak_mutasyonlu_genler_3_hastalik()

if __name__ == "__main__":
    main()
```

Program Çıktıları

Programa eklenen pandas kütüphanesi sayesinde dosya okuma işlemi yapıldı. Burada dosya türünden dolayı ayırma işlemi sep="\t" parametresi ile sağlandı. Tüm satırların gözükmesi adına set_option fonksiyonu kullanıldı. Ardından tüm kayıtlar txt dosyasına yazıldı.

tum_dosya.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

```
1349905      SEPT6   ENST00000343984.5  ...   74.00     2
1349906      SEPT6   ENST00000343984.5  ...     NaN     2
1349907      SEPT6   ENST00000343984.5  ...   55.00     2
1349908      SEPT6   ENST00000343984.5  ...   70.00     2
1349909      SEPT6   ENST00000343984.5  ...   70.00     2
1349910      SEPT6   ENST00000343984.5  ...   74.00     2
1349911      SEPT6   ENST00000343984.5  ...   63.00     2
1349912      SEPT6   ENST00000343984.5  ...   63.00     2
1349913      SEPT6   ENST00000343984.5  ...     NaN     2
1349914      SEPT6   ENST00000343984.5  ...     NaN     2
1349915      SEPT6   ENST00000343984.5  ...     NaN     2
1349916      SEPT6   ENST00000343984.5  ...     NaN     2
1349917      SEPT6   ENST00000343984.5  ...   56.00     2
1349918      SEPT6   ENST00000343984.5  ...   41.00     2
1349919      SEPT6   ENST00000343984.5  ...     NaN     2
1349920      SEPT6   ENST00000343984.5  ...     NaN     2
1349921      SEPT6   ENST00000343984.5  ...   77.83     2
1349922      SEPT6   ENST00000343984.5  ...   44.00     2
1349923      SEPT6   ENST00000343984.5  ...   74.00     2
1349924      SEPT6   ENST00000343984.5  ...     NaN     2
1349925      SEPT6   ENST00000343984.5  ...   66.00     2
1349926      SEPT6   ENST00000343984.5  ...     NaN     2
1349927      SEPT6   ENST00000343984.5  ...     NaN     2
1349928      SEPT6   ENST00000343984.5  ...   59.00     2
1349929      SEPT6   ENST00000343984.5  ...   67.00     2
1349930      SEPT6   ENST00000343984.5  ...     NaN     2
1349931      SEPT6   ENST00000343984.5  ...   59.00     2
1349932      SEPT6   ENST00000343984.5  ...   58.00     2
1349933      SEPT6   ENST00000343984.5  ...     NaN     2
1349934      SEPT6   ENST00000343984.5  ...   64.00     2
1349935      SEPT6   ENST00000343984.5  ...   63.00     2
1349936      SEPT6   ENST00000343984.5  ...     NaN     2
1349937      SEPT6   ENST00000343984.5  ...     NaN     2
1349938      SEPT6   ENST00000343984.5  ...   60.00     2
1349939      SEPT6   ENST00000343984.5  ...   55.00     2
1349940      SEPT6   ENST00000343984.5  ...   30.00     2

[1349941 rows x 38 columns]
```

Şekil 1 TSV dosyasının tüm kayıtlarının okunup yazıldığı txt dosyası

Sonrasında hasta_id_getir fonksiyonu üzerinden skin ve liver carcinoma hasta idleri ayrı dosyalara yazıldı ve fonksiyon sonu hastalıkların hasta sayıları döndürüldü. Burada liver carcinoma hasta sayısının skin carcinoma hasta sayısından daha fazla olduğu görülmüş oldu.

liver_carcinom_hasta_id.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

```
        ID_sample
11        1918838
15        2194690
19        2194706
40        1918814
52        1652979
69        1919262
71        2194723
73        1918809
76        2120944
79        1919209
80        2340950
88        1919255
98        1918823
102       2194434
130       1919254
142       2194710
143       1919259
172       1918817
200       2120960
201       2340466
226       2120951
240       2120980
241       1918803
253       2120915
266       2194719
276       1566773
286       1918872
```

Şekil 2 Liver carcinoma hastalarının idleri yazılan txt dosyası

skin_carcinoma_hasta_id.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

	ID_sample
10	2521295
78	2688645
228	2385650
271	2292475
289	2292492
329	2292470
377	2292472
439	1604872
465	2688647
468	2292462
519	2292471
592	2724580
707	2292469
753	2716511
785	2726338
808	2292478
817	2716810
859	2584584
1019	2521292
1082	2688389
1124	2292493
1221	2384186
1299	2292466
1366	2584589
1389	2292486
1499	2385646
1525	2292460
1583	2637734

Şekil 3 Skin carcinoma hastalığına sahip hastaların idlerinin yazdığı txt dosyası

Liver ve skin carcinoma hastalarında görülen mutasyonlu genler sorgulanırken primary site ve primary histology kolonlarından yararlanıldı. Ardından hasta idleri ve mutasyonlu genleri dosyaya yazıldı ve hasta idleri üzerinden sıralama yapılarak aynı hastalara ait mutasyon genleri alt alta getirildi. Bu çıktılar sayesinde bu hastalıklara bazı gen mutasyonlar tek başına sebep olurken bazı hastalarda birden fazla gen mutasyonu gerektiği gördü.

liver_carcinom_hastalarinin_mutasyonlu_genleri.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

	ID_sample	GENOMIC_MUTATION_ID
631014	684195	COSV56056643
1048672	690746	COSV54736624
1048860	696439	COSV54736310
479906	705028	COSV55497369
485330	705030	COSV55497419
475158	705036	COSV55497469
505029	705037	COSV55497378
487147	705038	COSV55497369
511857	705039	COSV55497419
467472	705040	COSV55497388
488703	705041	COSV55497469
506382	711677	COSV55508630
1055140	721023	COSV54736310
479443	721024	COSV55497419
1053678	724820	COSV54736624
1050240	724821	COSV54736624
493448	725765	COSV55497369
1053242	730643	COSV54736624
469061	730676	COSV55499223
1048866	730720	COSV54736624
1051644	730742	COSV54736624
484905	733807	COSV55497419
476493	735437	COSV55497469
1182527	755707	COSV59272241
1181514	755708	COSV59268145
1183113	755709	COSV59272813
348064	798715	COSV53842469

Şekil 4 Liver carcinoma hastalarının mutasyonlu genlerinin txt dosyası

skin_carcinoma_hastalarinin_mutasyonlu_genleri.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

	ID_sample	GENOMIC_MUTATION_ID
337401	694377	COSV54239883
338564	704597	COSV54236743
500448	704598	COSV55497369
470858	704618	COSV55497469
469935	704619	COSV55497469
1052141	704877	COSV54736383
1049879	704879	COSV54736383
1052875	704880	COSV54736383
1053310	704881	COSV54736383
1049609	704882	COSV54736416
478278	704882	COSV55498802
1051040	704883	COSV54736480
1055760	704885	COSV54736991
1054954	704886	COSV54736991
488266	704888	COSV55497419
338478	704891	COSV54236734
338186	704892	COSV54236734
338169	704893	COSV54239883
1053637	704910	COSV54736383
337154	704910	COSV54236734
1052747	704913	COSV54736991
494822	704914	COSV55497419
490744	704915	COSV55497419
495698	704916	COSV55497419
473063	704917	COSV55497419
337321	704918	COSV54236734
337113	704919	COSV54236734
337916	704920	COSV54236734
337850	710390	COSV54236734
337640	710391	COSV54236734
336940	710392	COSV54236734
337508	710393	COSV54236734
337021	710394	COSV54236734
338995	710395	COSV54236734
337540	710396	COSV54236734

Şekil 5 Skin carcinoma hastalarının mutasyonlu genlerinin txt dosyası

İstenilen liver ve skin carcinoma ortak genleri ortak_gen_bul fonksyionu ile çıkarıldı. Burada her iki hastalığa ait genler tekrarsız olarak çıkarıldı . Çıkarılan listeler concat fonksiyonu ile bağlandı ve tekrar eden genler ortak olduğundan duplicate fonksiyonunun “dublicated == True” ifadesi kullanılarak tekrar edenlerden ikincisi süzülüp ortak genler elde edildi. Burada liver 1407 civarı gene sahipken ortak gen 706 olduğu gözlendi ve bu da aslında carcinomada yaklaşık %50 civarı benzer genler olduğu kanısına varıldı.

ortak_gen_livercarcinoma&skincarcinoma.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

	Gene_name
10	LEPROTL1
289	CCR4
575	KLF4
1019	RSPO3
1583	PREX2
6542	CD28
6804	PTK6
6986	C15orf65
7076	RAD17
7422	BCL9L
8339	BCR
10063	FEN1
10130	KIAA1549
11769	SMAD4
14375	NTRK3
19358	FNBP1
20358	FAT3
28159	ARHGAP5
29168	FOXO4
29407	STAT6
29765	GLI1
30497	ATP1A1
31199	DNMT3A
35557	RBM10
36499	PTCH1
38637	CREB3L1
39274	KIT
49286	RMI2

Şekil 6 Liver ve skin carcinoma hastalıklarına ait ortak genleri gösteren txt dosyası

Liver ve skin carcinoma hastalıkları için istenen ayrı ayrı en çok görünen 30 gen ve yüzdeleri için en_cok_gorunen_gen_yuzde isimli fonksiyon yazıldı ve bu fonksiyonda istenen ek olarak en çok görülen 30 genin görülme sayısının analizi de yapıldı. Bu fonksiyon içerisinde oluşturulan birinci sorgu kısmında incelenen hastalığa sahip hasta sayısı tekrarsız olarak çıkarılıp ikinci sorgu kısmında genin görünme sayısı çıkarıldı . Son aşamada ise yeni bir kolon oluşturulup yüzde hesabu yapıldı ve dosyaya yazıldı.

Bu fonksiyonun çalıştırılması sonucu gen görülme sayısı ve hastalardaki yüzdelerine bakılınca bir hastaya ait bir gende birden fazla mutasyon olabileceği görülmüş oldu . Bu sonuçta aslında görülme sayısı ile hastalarda görülme yüzdesi arasındaki beklenen doğru orantının olmama sebebini net bir şekilde ortaya koymaktadır. Ortak genlerin olduğu fakat yüzde farklarının gözlendiği görüldü.

livercarcinoma_en_cok_gorunen_30_gen_ve_yuzdeleri.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

	Gene_name	Gorulme_Sayisi	Yüzde
0	CNTNAP2	5214	3.286653
1	CSMD3	3262	2.743641
2	LRP1B	3079	2.972278
3	ROBO2	2971	3.000857
4	PTPRD	2482	0.400114
5	DCC	2404	2.715061
6	CTNNA2	2341	1.800514
7	ERBB4	2318	4.144041
8	GPC5	1763	3.400972
9	PTPRT	1641	4.201200
10	CTNND2	1537	6.287511
11	TP53	1435	12.317805
12	CTNNB1	1325	3.972564
13	NBEA	1312	6.087454
14	FHIT	1293	0.600171
15	ZFHX3	1165	1.543298
16	CNBD1	1097	8.430980
17	CAMTA1	972	1.800514
18	EPHA3	961	23.835382
19	RGS7	859	1.228923
20	FAT3	840	6.859103
21	ALK	834	4.201200
22	PREX2	759	0.142898
23	NTRK3	714	3.915404
24	KMT2C	692	0.600171
25	ZNF521	663	8.916833
26	GPHN	616	2.857959
27	LPP	608	4.744213
28	PRKCB	578	3.258074
29	FOXP1	564	4.829951

Şekil 7 Liver carcinoma hastalarında en çok görülen 30 gen ve yüzdelerinin dosyası

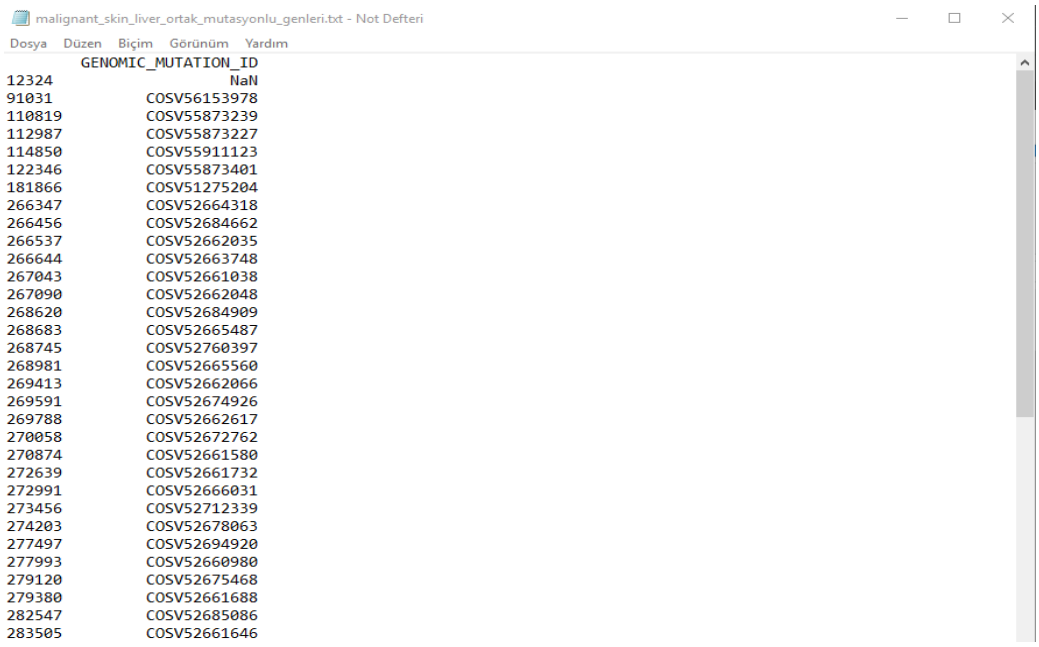
skincarcinoma_en_cok_gorunen_30_gen_ve_yuzdeleri.txt -

Dosya Düzen Biçim Görünüm Yardım

	Gene_name	Gorulme_Sayisi	Yüzde
0	TP53	976	2.924791
1	MUC16	701	0.557103
2	LRP1B	511	3.760446
3	NOTCH1	459	2.506964
4	PTCH1	440	1.044568
5	CSMD3	437	1.183844
6	CNTNAP2	411	2.715877
7	NOTCH2	389	1.183844
8	PTPRD	353	0.417827
9	KMT2D	257	2.576602
10	PTPRT	251	2.924791
11	ROBO2	242	8.844011
12	KMT2C	240	1.671309
13	GPC5	224	2.855153
14	DCC	221	1.253482
15	ERBB4	221	1.323120
16	ZFHX3	214	3.342618
17	FAT4	211	0.696379
18	MUC4	205	7.729805
19	NBEA	203	1.880223
20	CTNNA2	192	5.431755
21	ROS1	190	6.197772
22	GRIN2A	168	0.905292
23	CTNND2	165	2.994429
24	FAM135B	163	0.766017
25	CDKN2A	152	2.715877
26	RGS7	151	1.253482
27	CARD11	150	5.153203
28	FAT1	149	3.412256
29	FAT3	147	2.437326

Şekil 8 Skin carcinoma hastalarında en çok görülen 30 gen ve yüzdelerinin dosyası

Malignant melanoma , skin ve liver carcinoma hastalıklarının ortak genlerinin olan 71 gen aslında bu genler ile başka genlerin mutasyonunun birleşmesi sonucu bu hastalıkların spesifik olarak ortaya çıkmasını sağladığı gözlenmiştir. Bu da demek oluyor ki bir genin etkileştiği diğer gen veya aynı gendeki mutasyonun farklı noktada olması hastalığın farkını ortaya koymaktadır diye düşünülebilinir.



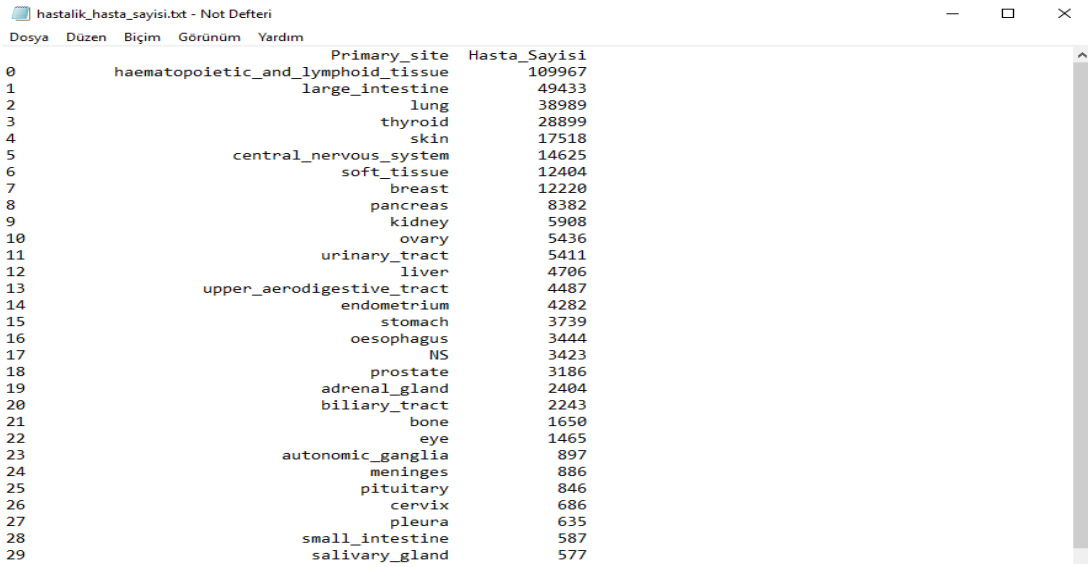
malignant_skin_liver_ortak_mutasyonlu_genleri.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

	GENOMIC_MUTATION_ID
12324	NaN
91031	COSV56153978
110819	COSV55873239
112987	COSV55873227
114850	COSV55911123
122346	COSV55873401
181866	COSV51275204
266347	COSV52664318
266456	COSV52684662
266537	COSV52662035
266644	COSV52663748
267043	COSV52661038
267090	COSV52662048
268620	COSV52684909
268683	COSV52665487
268745	COSV52760397
268981	COSV52665560
269413	COSV52662066
269591	COSV52674926
269788	COSV52662617
270058	COSV52672762
270874	COSV52661580
272639	COSV52661732
272991	COSV52666031
273456	COSV52712339
274203	COSV52678063
277497	COSV52694920
277993	COSV52660980
279120	COSV52675468
279380	COSV52661688
282547	COSV52685086
283505	COSV52661646

Şekil 9 Malignant melanoma , skin ve liver carcinoma hastalıklarının ortak mutasyonlu genlerinin olduğu dosya

Ödevde verilen sorulara ek olarak yapılan bir diğer inceleme hastalıklar ve hasta dağılımlarının analizidir. Burada görülmüştür ki hastalıklar arasında en yaygın olan kan hastalıklarıdır. Bunu kalın bağırsak ve akciğer hastalıkları takip etmektedir. Bu analiz sonucunda nadir olarak hastalık görülen organ ve vücut bölgeleri de görülmüştür. Dosyadaki çeşitliliğin fazlalığı da görülmüştür. Ayrıca karaciğer hasta sayısına bakınca bu dosyada bir organa ait birden fazla hastalığında dosyada olduğu görülmektedir.



hastalik_hasta_sayisi.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

	Primary_site	Hasta_Sayisi
0	haematopoietic_and_lymphoid_tissue	109967
1	large_intestine	49433
2	lung	38989
3	thyroid	28899
4	skin	17518
5	central_nervous_system	14625
6	soft_tissue	12404
7	breast	12220
8	pancreas	8382
9	kidney	5908
10	ovary	5436
11	urinary_tract	5411
12	liver	4706
13	upper_aerodigestive_tract	4487
14	endometrium	4282
15	stomach	3739
16	oesophagus	3444
17	NS	3423
18	prostate	3186
19	adrenal_gland	2404
20	biliary_tract	2243
21	bone	1650
22	eye	1465
23	autonomic_ganglia	897
24	meninges	886
25	pituitary	846
26	cervix	686
27	pleura	635
28	small_intestine	587
29	salivary_gland	577

Şekil 10 Hastalıklar ve hasta dağılımlarının dosyası

Ek olan bir diğer analiz ise hastaların mutasyon sayılarıdır. Bu analiz ile görülmüştür ki bir hasta bir mutasyonla o hastalığa yakalanabileceği gibi birden fazla mutasyonla da yakalanmış olabilir. Bu da gösteriyor ki her gen bölgesindeki mutasyon aynı etkiye sahip değildir. Hatta aynı gen bölgesinde olup nükleotit farkı bile kişinin hastalığında etkili faktördür. Analizde en çok mutasyonu olan hasta 2460 mutasyona sahipken bir başka hasta 1 mutasyon ile başka bir hastalığa yakalanmıştır.

hasta_mutasyon_sayilari.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

	ID_sample	Mutasyon_Sayisi
0	2194686	2460
1	2197226	1969
2	1842414	1796
3	2185962	1303
4	1652971	1293
5	2443644	1274
6	2371738	1193
7	2443648	1069
8	2121224	1057
9	1652981	1045
10	2263330	1033
11	2196284	1014
12	1819100	988
13	2193467	976
14	2194688	943
15	1919258	935
16	2195248	898
17	2194733	856
18	1652979	825
19	2465276	813
20	1783352	793
21	2193470	792
22	1919262	781
23	2371742	774
24	2195318	770
25	1783335	762

Şekil 11 Hastalar ve mutasyon sayılarının dosyası

Bir başka ek analiz ise hastalıklar ve bu hastalıklara yakalanan hastaların yaş ortalamaları oldu. Burada görülen en net durum genellikle organlarda gözlenen hastalıkların 45 yaş ve üstünde çıkmasıdır. Ayrıca nadir görülen organ hastalıklarında yaş değerine yer verilmemiş olması dikkat çekicidir. Buradaki sonuçlara bakarak aslında hastalıkların veya mutasyon etkilerinin ilerleyen yaşlarda daha çok ortaya çıktığı görülmüştür.

hastalik_yas_ortalamalari.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

	Primary site	Yas_Ortalamasi
0	NS	52.228261
1	adrenal_gland	44.757853
2	autonomic_ganglia	11.300393
3	biliary_tract	62.910668
4	bone	32.591820
5	breast	54.069123
6	central_nervous_system	35.256162
7	cervix	49.569554
8	endometrium	59.864809
9	eye	56.892889
10	fallopian_tube	73.000000
11	female_genital_tract_site_indeterminate	56.000000
12	gastrointestinal_tract_site_indeterminate	50.583333
13	genital_tract	18.981651
14	haematopoietic_and_lymphoid_tissue	47.883924
15	kidney	44.926088
16	large_intestine	63.505714
17	liver	49.684462
18	lung	61.975558
19	mediastinum	NaN
20	meninges	55.296891
21	oesophagus	61.226545
22	ovary	53.649935
23	pancreas	61.948017
24	paratesticular_tissues	57.000000
25	parathyroid	50.758929
26	penis	67.117647
27	pericardium	NaN

Şekil 12 Hastalıklara ait yaş ortalamalarının dosyası

Mtasyonlu genler ve görülme sayılarının incelemesi işlemini de gönüllülükle gerçekleştirdim. Burada bazı genlerin 1 kez görülmesi bazılarının 41489 kez görülmesi ve diğer sayı dağılımları aslında bana bazı genlerin tek başına hastalık yapabileceği düşüncesine ek bir delil niteliğindeydi. Yüksek görülen mutasyonlu genleri yorumlamak gerekirse birden fazla hastalıkta farklı mutasyonlu gen etkileşimleri ile rol oynuyor olabilir ya da yaygın görülen hastalıklarda etkindir denilebilir.

mutasyonlu_genler_ve_sayilari.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

	GENOMIC_MUTATION_ID	Görülme_Sayisi
0	COSV67569051	41489
1	COSV56056643	28296
2	COSV55497369	15137
3	COSV55497419	10351
4	COSV55497388	5639
5	COSV55497469	5136
6	COSV61615239	4877
7	COSV55873195	3106
8	COSV51542664	2679
9	COSV51765161	2500
10	COSV55873239	2478
11	COSV55497479	2402
12	COSV55497461	2004
13	COSV57169334	1829
14	COSV54736340	1695
15	COSV57116546	1680
16	COSV52661038	1566
17	COSV55873227	1538
18	COSV53390026	1507
19	COSV55497582	1469
20	COSV54736310	1279
21	COSV62687862	1185
22	COSV51765119	1122
23	COSV52661580	1100
24	COSV57116551	1067
25	COSV52660980	1016
26	COSV52662066	968
27	COSV55386424	926
28	COSV61615256	877

Şekil 13 Mutasyonlu genler ve görülme sayıları

Liver ve skin carcinoma hastalarının en yüksek , en küçük ve ortalama değerlerinin analizi yapıldı. Her iki hastalıkta 60 yaş üstünde daha yoğun görülse de küçük yaşlarda da nadir olarak görülmektedir. Liver carcinomanın skin carcinomadan daha küçük yaşlarda görülmesi yaygınlığın sebebi midir düşüncesi aslında ortalama değerdeki yakınlık ile çürütülmektedir. Fakat küçük yaşlarda görülmesi liver carcinoma sebebi olan mutasyonların etkisinin ortaya çıkışının daha çabuk olabileceğinin çevresel faktörlerden ziyade genetik aktarımla ortaya çıktıkları düşüncesini desteklemektedir.

skincarcinomayas_incelemesi.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

```
En Yüksek Yaş : 100.0
En Düşük Yaş : 9.0
Ortalama Yaş : 64.20215940534968
```

Şekil 14 Skin carcinoma hastalığının yaş incelemesinin sonuç dosyası

livercarcinomayas_incelemesi.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

```
En Yüksek Yaş : 90.0
En Düşük Yaş : 1.5
Ortalama Yaş : 63.932721382289415
```

Şekil 15 Liver carcinoma hastalığının yaş incelenmesinin sonuç