Branch: master ▾   **Bioinformatics** / Bioinformatics.ipynb   Find file   Copy path

🐙 **Amritpal** Disulfide bond score added but all files not updated.   4db12bc   43 seconds ago

0 contributors

2949 lines (2949 sloc)   129 KB   ⟨⟩ ▤   Raw   Blame   History   ✏️ 🗑️

```python
In [11]:  #!/usr/bin/env python
          # coding: utf-8

          # In[38]:


          import numpy as np
          import pandas as pd
          import re
          import matplotlib.pyplot as plt
          import sys
          import time



          #Input file must be given to run the code.
          #For Example Run(in terminal) 'python3 Bioinformatics input.xlsx'

          #Filename =sys.argv[1] # Read input file
          t0=time.time()
          df = pd.read_excel('input.xlsx')# Read EXCEL Files and storing in database 'df'
          t1 = time.time()

          total = t1-t0
          data= pd.DataFrame()# Creating New Data Frame


          # Storing Usefull information From the database 'df'
          data['Entry']=df['Entry'] #Storing Entry
          data['Entry name']=df['Entry name']  #Storing Entry Name
          data['Protein names']=df['Protein names']  #Storing Protein Names
          data['Gene names']=df['Gene names'] # Storing Gene Names
          data['Length']=df['Length'] # Storing Length
          data['Organism']=df['Organism'] # Storing Type of Organism
          data['Mass']=df['Mass'] #Storing Mass
          data['Status']=df['Status']

          #Now removing the extra data we dont need and containg only organism 'Human'
          human_data=data[data['Organism'].str.contains("Homo sapiens")] #Selecting only organism which cont
          ains 'Homo sapiens'
          data= human_data

          data['Disulfide bond']=df['Disulfide bond'] # Storing Disulfide bond
          data['Glycosylation']= df['Glycosylation']# Storing Glycosylation Position
          data=data.dropna().reset_index(drop=True) # Drop all rows which contains Not a number and reset th
          e index

          # Working on getting the relative positions of the Disulfide bond
          disulfide_column1= data['Disulfide bond'] # Storing data temporary as 'disulfide_column'
          disulfide_column2= data['Disulfide bond']
          def get_sulfide_value(newdata):
              return re.findall('\d+ \d+',newdata)

          bond= disulfide_column1.apply(get_sulfide_value) # Function call which gives all the positons
          data['Disulfide bond']=disulfide_column2.apply(get_sulfide_value)# Storing extracted disulfide val
```

*[handwritten annotation]: fixed (*cannot change depends upon input) reading file 94.838C*

```
                                                                              ue in data Frame.
```

In [12]: `total`

Out[12]: `24.38349413871765`

In [13]:
```python
#Working on getting the positions of N-linked Glycosylation
Glyco_data= data['Glycosylation'] # Storing data temporary as 'Glyco_data'
def getGlycoNLinked(data):   #Making a Function to get the positions
    return re.findall('\d+\s* N-linked',data) # The return data will contain 'position and N-linke
d pattern'
data['Glycosylation']= Glyco_data.apply(getGlycoNLinked) # the return data will be like '49 N-link
ed'
#Now Removing the extra word 'N-linked' and getting all positions of Glycosylation.
temp_data= data['Glycosylation'].astype(str)
def get_gly_value(data):
    return re.findall('\d+',data) # return all positions
data['Glycosylation']= temp_data.apply(get_gly_value)

t0 = time.time()
# Creating a temporary dataframe for total length in columns and length in rows(like number of pai
rs in each index).
a= pd.DataFrame()
b= pd.DataFrame()
a =pd.Series(data['Glycosylation'])
b =pd.Series(data['Disulfide bond'])
lenB=b.str.len()
lenA=a.str.len()

rel_pos = pd.Series(a.size,dtype=np.str)
for i in range(a.size):
    rel_pos[i]=''

InsidePairs=pd.Series(b.size,dtype=np.str)
InsidePairsLength=np.zeros(b.size,dtype=np.int)
for i in range(a.size):
    InsidePairs[i]=''

# Now using loop and properties of Python library 're' to get values and finding positions between
disulphide bond and glycosation.
i=0
while i<b.size:
    for j in range(lenA[i]):
        Position=int(a[i][j])
        rel_pos[i]=rel_pos[i]+str(a[i][j])+'{|'
        for k in range(lenB[i]):
            split = re.findall('\d+',b[i][k])
            firstPair= int(split[0])
            secondPair=int(split[1])
            if Position< firstPair:
                rel_pos[i] = rel_pos[i]+'o'+str(Position-firstPair)
                rel_pos[i] = rel_pos[i]+'o'+str(Position-secondPair)
            elif Position>= firstPair and Position<=secondPair:
                rel_pos[i]=  rel_pos[i]+'i' + str(Position-firstPair)
                rel_pos[i] = rel_pos[i]+'i'+ str(Position-secondPair)
                InsidePairs[i]= InsidePairs[i] + str('('+ str(firstPair)+','+str(secondPair)+')')
                InsidePairsLength[i]=InsidePairsLength[i]+ (secondPair-firstPair)
            else:
                rel_pos[i] = rel_pos[i]+'o'+str(Position-firstPair)
                rel_pos[i] = rel_pos[i]+'o'+str(Position-secondPair)
        rel_pos[i]=rel_pos[i]+'|'
    rel_pos[i]=rel_pos[i]+'}'
    i=i+1
data['rel_pos']= rel_pos
data['In_Gly_Dis_Pair']=InsidePairs
t1 = time.time()

total = t1-t0
```

(handwritten annotations: "Nested loops", "Running Time O(n³)", "55.35µ")

In [14]: `total`

Out[14]: `55.35723352432251`

```
In [20]:    #Caluclating Score of each pairs of disulfide bonds
            # O score if other pair outside.
            # 0.5 score if other pair half inside and half outside.
            # 1.0 score if other pair full inside.
            st=time.time()
            array = pd.Series(a.size,dtype=np.str)
            for x in range(a.size):
                array[x]=''

            for i in range(0,len(data)):
                temp_array = np.zeros(lenB[i])
                for j in range(0,lenB[i]):
                    for k in range(j+1,lenB[i]):
                        if k!=lenB[i]:
                            split1 = re.findall('\d+',data['Disulfide bond'][i][j])
                            p1e1=int(split1[0])
                            p1e2=int(split1[1])
                            split2 = re.findall('\d+',data['Disulfide bond'][i][k])
                            p2e1=int(split2[0])
                            p2e2=int(split2[1])

                            if(p2e1>p1e2):
                                temp_array[j]=temp_array[j]+0
                                temp_array[k]=temp_array[k]+0
                            elif (p2e1<p1e2) and (p2e2>p1e2):
                                temp_array[j]=temp_array[j]+0.5
                                temp_array[k]=temp_array[k]+0.5
                            elif (p2e1<p1e2) and (p2e2<p1e2):
                                temp_array[j]=temp_array[j]+1.0
                                temp_array[k]=temp_array[k]+0.0
                    array[i]=array[i]+str(temp_array[j])+','

            en=time.time()
            total=en-st
```

```
In [21]:    print(total)
            data['Disulfide_score']=array
```

```
55.47823166847229
```

```
In [33]:    ## Calculating Interbond Distance
            start=time.time()
            interbond_distance= pd.Series(a.size,dtype= np.str)

            for i in range(a.size):
                interbond_distance[i]=''

            for i in range(a.size):

                if lenB[i]==0:
                    interbond_distance[i]='No pair'

                if lenB[i]==1:
                    interbond_distance[i]='Only One pair'

                for r in range(lenB[i]-1):
                    split1= re.findall('\d+',data['Disulfide bond'][i][r])
                    split2= re.findall('\d+',data['Disulfide bond'][i][r+1])
                    interbond_distance[i]= interbond_distance[i]+ str(int(split2[0])-int(split1[1]))+ '|'

            data['interbond_distance']=interbond_distance

            #Intrabond distance formation
            intrabond=pd.Series(b.size,dtype=np.str)
            for i in range(a.size):
                intrabond[i]=''
            j=0;
            while j<b.size:
                for k in range(lenB[j]):
                    split = re.findall('\d+',b[j][k])
                    firstPair= int(split[0])
```

Nested loops

Running time

$$O(n^3)$$

55.4788c

Continued

```
            firstPair= int(split[0])
            secondPair=int(split[1])
            intrabond[j]=intrabond[j]+str(secondPair-firstPair)+'|'
        j=j+1
data['intrabond']=intrabond


#number of pairs of disuphide bonds
data['No. Disulphide bonds']=lenB


#1st disulphide pair to N-terminus distance
first_pos = pd.Series(a.size,dtype=np.int)
for r in range(a.size):
    split = re.findall('\d+',data['Disulfide bond'][r][0])
    firstPair= int(split[0])
    first_pos[r]= firstPair
data['first_position_occurance']=first_pos


#last disulphide pair to C-terminus distance
last_pos= pd.Series(a.size,dtype=np.int)
for r in range(a.size):
    split= re.findall('\d+',data['Disulfide bond'][r][(lenB[r]-1)])
    lastPair= int(split[1])
    last_pos[r] = lastPair
data['last_position_occurance']=last_pos

#Calculate average distance of intrabond distance
def intrabond_Average(data):
    return re.findall('\d+',data)
average= data['intrabond'].apply(intrabond_Average)
lenIntra=average.str.len()

averageintrabond= np.zeros(a.size)
for i in range(a.size):
    totalsum=0
    for j in range(lenIntra[i]):
        totalsum= totalsum+int(average[i][j])
    averageintrabond[i]=totalsum/lenIntra[i]
data['average_Intrabond']=averageintrabond
end=time.time()
total=end-start
print(total)
```

```
34.521262645721436
```

In [30]:
```
#calculating average distance of those pairs of disulphide bonds that have Glycosylation Inside.
s=time.time()
temp_length=pd.Series(b.size,dtype=np.str)
#for i in range(a.size):
#    temp_length[i]=''
inside_length = np.zeros(a.size,dtype=np.int)
for r in range(a.size):
    inside_length[r]=len(re.findall('\d+',data['In_Gly_Dis_Pair'][r]))/2
    temp_length[r]=re.findall('\d+',data['In_Gly_Dis_Pair'][r])
    if inside_length[r]!=0:
        InsidePairsLength[r]=InsidePairsLength[r]/inside_length[r]
data['avg_Inside_Pairs_Length']=InsidePairsLength
e=time.time()
t=e-s
print(t)
```

```
7.344398021697998
```

In [25]:
```
#If the glyco is inside then putting the sulphide position 'nil' at that point.
s=time.time()
i=0
while i<b.size:
    for j in range(lenA[i]):
        Position=int(a[i][j])
        for k in range(lenB[i]):
            if bond[i][k]!='nil':
                split = re.findall('\d+',b[i][k])
```

```
                    split = re.findall('\d+',b[i][k])
                    firstPair= int(split[0])
                    secondPair=int(split[1])
                    if Position>= firstPair and Position<=secondPair:
                        bond[i][k]='nil'
        i=i+1
e=time.time()
print(e-s)
```

18.13939380645752

18·13 sec

In [26]:
```
s=time.time()
#calculating length between sulfide bonds.
temp=pd.Series(b.size,dtype=np.str)
for i in range(a.size):
    temp[i]=''

for i in range(len(bond)):
    for j in range(lenB[i]):
        if bond[i][j]!='nil':
            split= re.findall('\d+',bond[i][j])
            temp[i]=str(temp[i] + str(int(split[1])-int(split[0]))+'|')
data['glyco_outside_bond']=bond
e=time.time()
print(e-s)
```

9.262168884277344

9.2 sec

In [27]:
```
s=time.time()
#get the integer values.
def get_gly_outside_values(temp):
    return re.findall('\d+',temp)
intrabond_outside= temp.apply(get_gly_outside_values)
data['intrabond_glyco_outside_bond']=intrabond_outside
e=time.time()
print(e-s)
```

0.0371861457824707

0·03 sec

In [28]:
```
#calculating avegrage lenth of sulphide bonds which do not have glyco inside.
s=time.time()
new =pd.Series(intrabond_outside)
newlen=new.str.len()
average= np.zeros(a.size)
for i in range(a.size):
    totalsum=0
    for j in range(newlen[i]):
        totalsum= totalsum+int(intrabond_outside[i][j])
    if newlen[i]!=0:
        average[i]=totalsum/newlen[i]
data['average_Intrabond_outside_glyco_bond']=average
e=time.time()
print(e-s)
```

1.001692771911621

1 sec

In [ ]:

In [76]:
```
data.to_excel('output.xlsx')
```

In [77]:

Out[77]:

	Protein names	Gene names	Length	Organism	Mass	Status	Disulfide bond	Glycosylation	...	
IAN	HLA class I histocompatibility antigen, A alph...	HLA-A HLAA	365	Homo sapiens (Human)	40,841	reviewed	[125 188, 227 283]	[110]	...	
IAN	Neutrophil elastase (EC 3.4.21.37) (Bone marro...	ELANE ELA2	267	Homo sapiens	28,518	reviewed	[55 71, 151 208, 181 187	[88, 124, 173]	...	

	3.4.21.37) (Bone marro...			(Human)			[181 187, 198 223]		
MAN	Granzyme B (EC 3.4.21.79) (C11) (CTLA-1) (Cath...	GZMB CGL1 CSPB CTLA1 GRB	247	Homo sapiens (Human)	27,716	reviewed	[49 65, 142 209, 173 188]	[71, 104]	... 7
MAN	HLA class I histocompatibility antigen, C alph...	HLA-C HLAC	366	Homo sapiens (Human)	40,649	reviewed	[125 188, 227 283]	[110]	... 3
MAN	HLA class I histocompatibility antigen, B alph...	HLA-B HLAB	362	Homo sapiens (Human)	40,460	reviewed	[125 188, 227 283]	[110]	... 3
MAN	NPC intracellular cholesterol transporter 2 (E...	NPC2 HE1	151	Homo sapiens (Human)	16,570	reviewed	[27 140, 42 47, 93 99]	[58, 135]	... -
MAN	Secreted frizzled-related protein 4 (sFRP-4) (...	SFRP4 FRPHE	346	Homo sapiens (Human)	39,827	reviewed	[24 85, 32 78, 69 108, 97 136, 101 125]	[38, 68, 116, 194, 240]	... -
MAN	Macrophage mannose receptor 1 (MMR) (C-type le...	MRC1 CLEC13D CLEC13DL MRC1L1	1456	Homo sapiens (Human)	166,012	reviewed	[35 49, 74 91, 168 194, 182 209, 247 340, 316 ...	[104, 344, 529, 926, 930, 1160, 1205]	... 2
MAN	Semaphorin-7A (CDw108) (JMH blood group antige...	SEMA7A CD108 SEMAL	666	Homo sapiens (Human)	74,824	reviewed	[120 126, 143 152, 266 366, 291 335, 493 511, ...	[105, 157, 258, 330, 602]	... -
MAN	Butyrophilin subfamily 3 member A2	BTN3A2 BT3.2 BTF3 BTF4	334	Homo sapiens (Human)	36,428	reviewed	[52 126, 166 220]	[115]	... 4
MAN	CMRF35-like molecule 1 (CLM-1) (CD300 antigen-...	CD300LF CD300F CLM1 IGSF13 IREM1 NKIR UNQ3105/...	290	Homo sapiens (Human)	32,335	reviewed	[40 108, 54 62]	[88]	... -
MAN	T-lymphocyte activation antigen CD80 (Activati...	CD80 CD28LG CD28LG1 LAB7	288	Homo sapiens (Human)	33,048	reviewed	[50 116, 162 216]	[53, 89, 98, 186, 207, 211, 226, 232]	... 4
MAN	Carboxypeptidase A6 (EC 3.4.17.-)	CPA6 CPAH	437	Homo sapiens (Human)	51,008	reviewed	[265 288]	[89, 153, 427]	... (
MAN	N(4)-(beta-N-acetylglucosaminyl)-L-asparaginas...	AGA	346	Homo sapiens (Human)	37,208	reviewed	[64 69, 163 179, 286 306, 317 345]	[38, 308]	... 9
MAN	Scavenger receptor cysteine-rich type 1 protei...	CD163 M130	1156	Homo sapiens (Human)	125,451	reviewed	[76 141, 89 151, 120 130, 184 248, 197 258, 22...	[105, 140, 767, 1027]	... -
				Homo			[1480 1569, 1513		

MAN	Collagen alpha-4(IV) chain	COL4A4	1690	Homo sapiens (Human)	164,038	reviewed	1566, 1525 1531, 1588 1686, 1...	[142, 669]	...
IAN	Cryptic protein (Cryptic family protein 1)	CFC1	223	Homo sapiens (Human)	24,612	reviewed	[90 97, 91 103, 105 114]	[52]	...
MAN	A disintegrin and metalloproteinase with throm...	ADAMTS19	1207	Homo sapiens (Human)	134,048	reviewed	[401 466, 441 448, 460 540, 499 524, 569 593, ...	[260, 797, 907, 949, 1009]	...
MAN	5-hydroxytryptamine receptor 7 (5-HT-7) (5-HT7...	HTR7	479	Homo sapiens (Human)	53,555	reviewed	[155 231]	[5, 66]	...
MAN	CD226 antigen (DNAX accessory molecule 1) (DNA...	CD226 DNAM1	336	Homo sapiens (Human)	38,614	reviewed	[37 108, 152 222]	[32, 83, 90, 97, 147, 186, 198, 231]	...
IAN	Chondroadherin (Cartilage leucine-rich protein)	CHAD SLRR4A	359	Homo sapiens (Human)	40,476	reviewed	[23 38, 304 346, 306 326]	[]	...
JMAN	Disintegrin and metalloproteinase domain-conta...	ADAM2 FTNB	735	Homo sapiens (Human)	82,457	reviewed	[287 370, 329 354, 331 336, 445 465, 616 627, ...	[122, 220, 353, 459, 566]	...
IAN	B-cell receptor CD22 (B-lymphocyte cell adhesi...	CD22 SIGLEC2	847	Homo sapiens (Human)	95,348	reviewed	[39 167, 44 102, 161 219, 265 309, 353 396, 44...	[67, 101, 112, 135, 164, 231, 363, 445, 479, 5...	...
IAN	Artemin (Enovin) (Neublastin)	ARTN EVN	220	Homo sapiens (Human)	22,878	reviewed	[123 188, 150 216, 154 218, 187 187]	[202]	...
IAN	Leucine-rich alpha-2-glycoprotein (LRG)	LRG1 LRG	347	Homo sapiens (Human)	38,178	reviewed	[43 56, 303 329]	[79, 186, 269, 325]	...
IAN	HLA class II histocompatibility antigen, DRB1-...	HLA-DRB1	266	Homo sapiens (Human)	30,030	reviewed	[146 202]	[48]	...
MAN	A disintegrin and metalloproteinase with throm...	ADAMTS20	1910	Homo sapiens (Human)	214,721	reviewed	[334 387, 363 369, 381 462, 419 446, 489 511, ...	[92, 191, 445, 702, 717, 728, 809, 870, 1061, ...	...
MAN	Beta-1,4-galactosyltransferase 4 (Beta-1,4-Gal...	B4GALT4 UNQ552/PRO1109	344	Homo sapiens (Human)	40,041	reviewed	[77 118, 189 208]	[220, 335]	...
				H			[310 395, 351 379,		

JMAN	Disintegrin and metalloproteinase domain-conta...	ADAM8 MS2	824	Homo sapiens (Human)	88,771	reviewed	353 362, 435 457, 448 454, ...	[67, 91, 436, 612]	...
MAN	Arylsulfatase A (ASA) (EC 3.1.6.8) (Cerebrosid...	ARSA	507	Homo sapiens (Human)	53,588	reviewed	[156 172, 161 168, 300 414, 488 500, 489 502, ...]	[158, 184, 350]	...
	...	...	...	...	...	...	...	...	...
MAN	Collagen alpha-2(V) chain	COL5A2	1499	Homo sapiens (Human)	144,910	reviewed	[1296 1328, 1336 1497, 1405 1450]	[1262, 1400]	... 8
JMAN	Adhesion G protein-coupled receptor A3 (G-prot...	ADGRA3 GPR125 UNQ556/PRO1113	1321	Homo sapiens (Human)	146,151	reviewed	[264 324]	[81, 98, 159, 206, 301, 332, 433, 453, 592, 65...]	... C
MAN	Glypican-1 [Cleaved into: Secreted glypican-1]	GPC1	558	Homo sapiens (Human)	61,680	reviewed	[32 68, 62 256, 69 259, 191 343, 246 279, 268 ...]	[79, 116]	...
MAN	Glutamate receptor ionotropic, kainate 2 (GluK...	GRIK2 GLUR6	908	Homo sapiens (Human)	102,583	reviewed	[96 347]	[67, 73, 275, 378, 412, 423, 430, 546]	... C
N	Coagulation factor V (Activated protein C cofa...	F5	2224	Homo sapiens (Human)	251,703	reviewed	[167 193, 248 329, 500 526, 603 684, 1725 1751...]	[51, 55, 239, 297, 382, 460, 468, 554, 741, 75...]	... 5
MAN	G-protein coupled estrogen receptor 1 (Chemoat...	GPER1 CEPR CMKRL2 DRY12 GPER GPR30	375	Homo sapiens (Human)	42,248	reviewed	[130 207]	[25, 32, 44]	... C
IAN	Frizzled-9 (Fz-9) (hFz9) (FzE6) (CD antigen CD...	FZD9 FZD3	591	Homo sapiens (Human)	64,466	reviewed	[39 100, 47 93, 84 122, 111 152, 115 139]	[53, 158]	...
MAN	Dopamine beta-hydroxylase (EC 1.14.17.1) (Dopa...	DBH	617	Homo sapiens (Human)	69,065	reviewed	[154 596, 232 283, 269 295, 390 503, 394 565, ...]	[64, 184, 566]	...
MAN	Kallikrein-13 (EC 3.4.21.-) (Kallikrein-like p...	KLK13 KLKL4	277	Homo sapiens (Human)	30,570	reviewed	[42 178, 61 77, 157 224, 189 203, 214 239]	[30, 225]	...
				Homo			[47 63,		

MAN	Kallikrein-15 (EC 3.4.21.-) (ACO protease)	KLK15	256	Homo sapiens (Human)	28,087	reviewed	138 215, 180 194, 205 230]	[171, 232]	...
MAN	Collagen alpha-1(XVIII) chain [Cleaved into: E...	COL18A1	1754	Homo sapiens (Human)	178,188	reviewed	[334 397, 344 390, 381 419, 408 443, 412 432, ...	[68, 129, 164, 926]	...
MAN	Collagen alpha-3(IV) chain (Goodpasture antige...	COL4A3	1670	Homo sapiens (Human)	161,813	reviewed	[1460 1551, 1493 1548, 1505 1511, 1570 1665, 1...	[253]	...
MAN	Protein FAM3B (Cytokine-like protein 2-21) (Pa...	FAM3B C21orf11 C21orf76 PRED44 UNQ320/PRO365	235	Homo sapiens (Human)	25,982	reviewed	[63 91, 69 229]	[120, 208]	...
MAN	HLA class II histocompatibility antigen, DQ al...	HLA-DQA1	254	Homo sapiens (Human)	27,805	reviewed	[132 188]	[103, 143]	...
MAN	Endothelial protein C receptor (Activated prot...	PROCR EPCR	238	Homo sapiens (Human)	26,671	reviewed	[118 186]	[47, 64, 136, 172]	...
AN	Exostosin-1 (EC 2.4.1.224) (EC 2.4.1.225) (Glu...	EXT1	746	Homo sapiens (Human)	86,255	reviewed	[652 704]	[89, 330]	...
MAN	Contactin-associated protein-like 2 (Cell reco...	CNTNAP2 CASPR2 KIAA0868	1331	Homo sapiens (Human)	148,167	reviewed	[35 181, 336 368, 520 552, 558 569, 563 578, 5...	[289, 346, 363, 379, 436, 506, 507, 546, 630, ...	...
MAN	C-type lectin domain family 2 member D (Lectin...	CLEC2D CLAX LLT1 OCIL	191	Homo sapiens (Human)	21,849	reviewed	[75 86, 103 184]	[95, 147]	...
MAN	Folate receptor beta (FR-beta) (Folate recepto...	FOLR2	255	Homo sapiens (Human)	29,280	reviewed	[31 59, 51 99, 60 103, 83 169, 90 140, 129 203...	[115, 195]	...
MAN	Metabotropic glutamate receptor 6 (mGluR6)	GRM6 GPRC1F MGLUR6	877	Homo sapiens (Human)	95,468	reviewed	[57 99, 244 536, 367 383, 423 430, 518 537, 52...	[296, 451, 479, 567]	...
AN	Carboxylesterase 3 (EC 3.1.1.1) (Liver carboxy...	CES3 UNQ869/PRO1887	571	Homo sapiens (Human)	62,282	reviewed	[97 124, 281 292]	[105]	...
MAN	Carboxylesterase 5A (EC 3.1.1.1) (Carboxyleste...	CES5A CES7	575	Homo sapiens (Human)	63,926	reviewed	[94 121]	[281, 363, 513, 524]	...

MAN	Polypeptide N-acetylgalactosaminyltransferase ...	GALNT13 KIAA1918	556	Homo sapiens (Human)	64,051	reviewed	[105 338, 329 407, 441 458, 481 496, 522 539]	[94, 116, 551]	...
IMAN	Endoplasmic reticulum metallopeptidase 1 (EC 3...	ERMP1 FXNA KIAA1815	904	Homo sapiens (Human)	100,231	reviewed	[204 222]	[182, 730]	...
MAN	Mammalian ependymin-related protein 1 (MERP-1)...	EPDR1 MERP1 UCC1	224	Homo sapiens (Human)	25,437	reviewed	[42 172, 88 222, 113 210]	[130, 182]	...
AN	Coagulation factor XIII B chain (Fibrin-stabil...	F13B	661	Homo sapiens (Human)	75,511	reviewed	[25 76, 59 87, 91 135, 118 146, 153 197, 180 2...	[162, 545]	...
MAN	EEF1AKMT4-ECE2 readthrough transcript protein ...	EEF1AKMT4-ECE2	883	Homo sapiens (Human)	99,773	reviewed	[212 217, 235 868, 243 828, 299 548, 757 880]	[279, 283, 324, 384, 429, 496, 652, 745, 753]	...
IAN	Protein delta homolog 2 (DLK-2) (Epidermal gro...	DLK2 EGFL9 UNQ2903/PRO28633	383	Homo sapiens (Human)	40,548	reviewed	[29 40, 33 46, 48 57, 66 71, 79 88, 95 107, 10...	[157]	...
MAN	Furin (EC 3.4.21.75) (Dibasic-processing enzym...	FURIN FUR PACE PCSK3	794	Homo sapiens (Human)	86,678	reviewed	[211 360, 303 333, 450 474]	[387, 440, 553]	...
MAN	Epidermal growth factor-like protein 6 (EGF-li...	EGFL6 MAEG PP648 UNQ281/PRO320	553	Homo sapiens (Human)	61,317	reviewed	[98 109, 105 118, 120 132, 178 191, 185 200, 2...	[397]	...

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [134]:
```

```
In [ ]:
```

```
In [31]:
```

```
Out[31]: 0        0
         1       57
         2        0
         3        0
         4        0
         5       28
         6        1
         7       22
```

```
8        7
9       74
10       68
11        1
12        0
13        0
14        1
15        0
16        0
17        3
18        0
19        2
20        0
21        4
22        0
23       16
24       26
25        0
26       13
27        0
28        8
29       16
        ...
2747      161
2748       60
2749       12
2750      251
2751       14
2752        0
2753       13
2754      110
2755       25
2756       77
2757        0
2758        0
2759       40
2760       56
2761       17
2762        0
2763        2
2764       81
2765        7
2766       13
2767       27
2768        0
2769      233
2770        0
2771        4
2772       10
2773        1
2774       18
2775        0
2776        0
Name: avg_Inside_Pairs_Length, Length: 2777, dtype: int64
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: