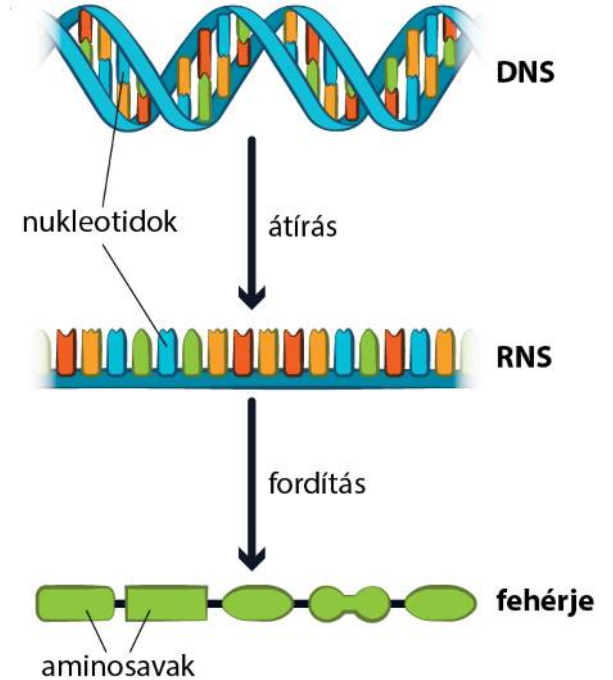
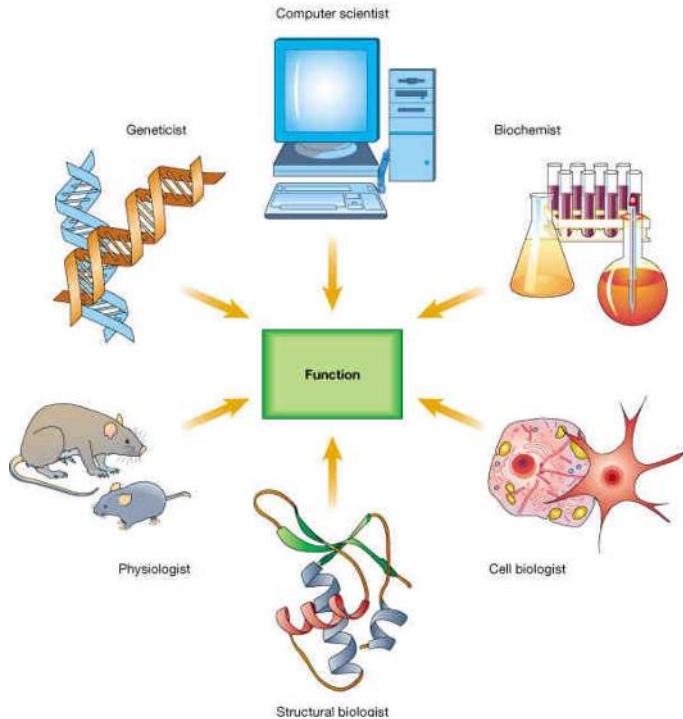


Kalandozások a bioinformatikában a Microsoft Copilot-tal

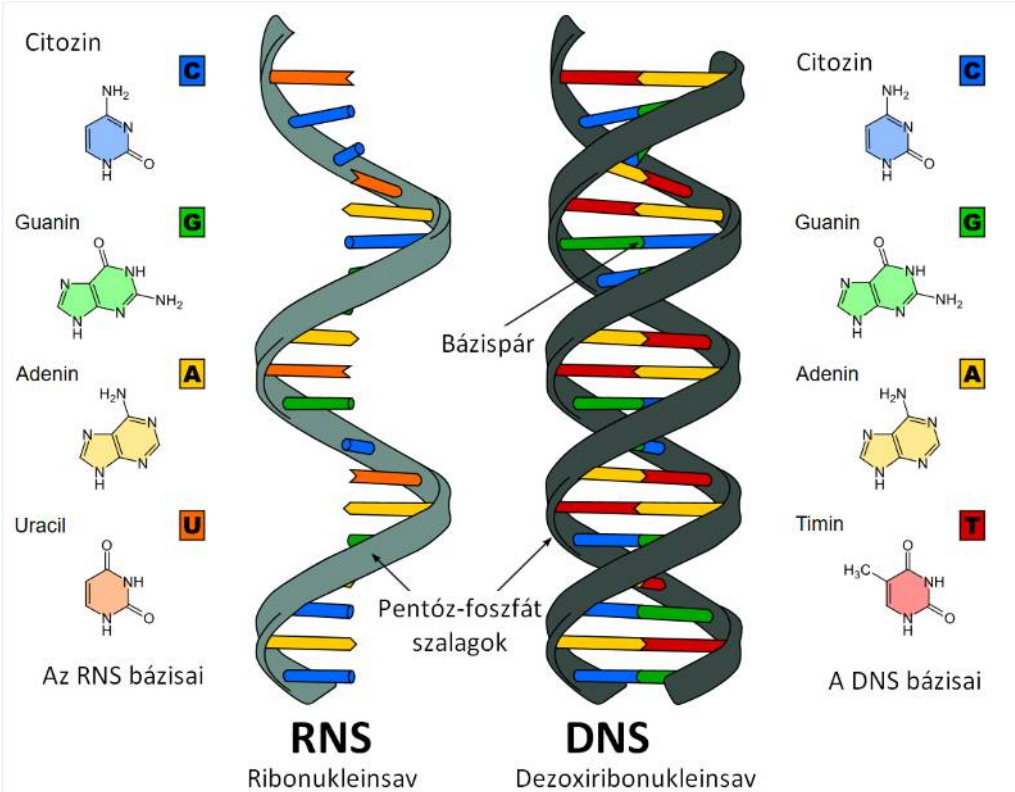


Te is lehetsz bioinformatikus!

Célkitűzés

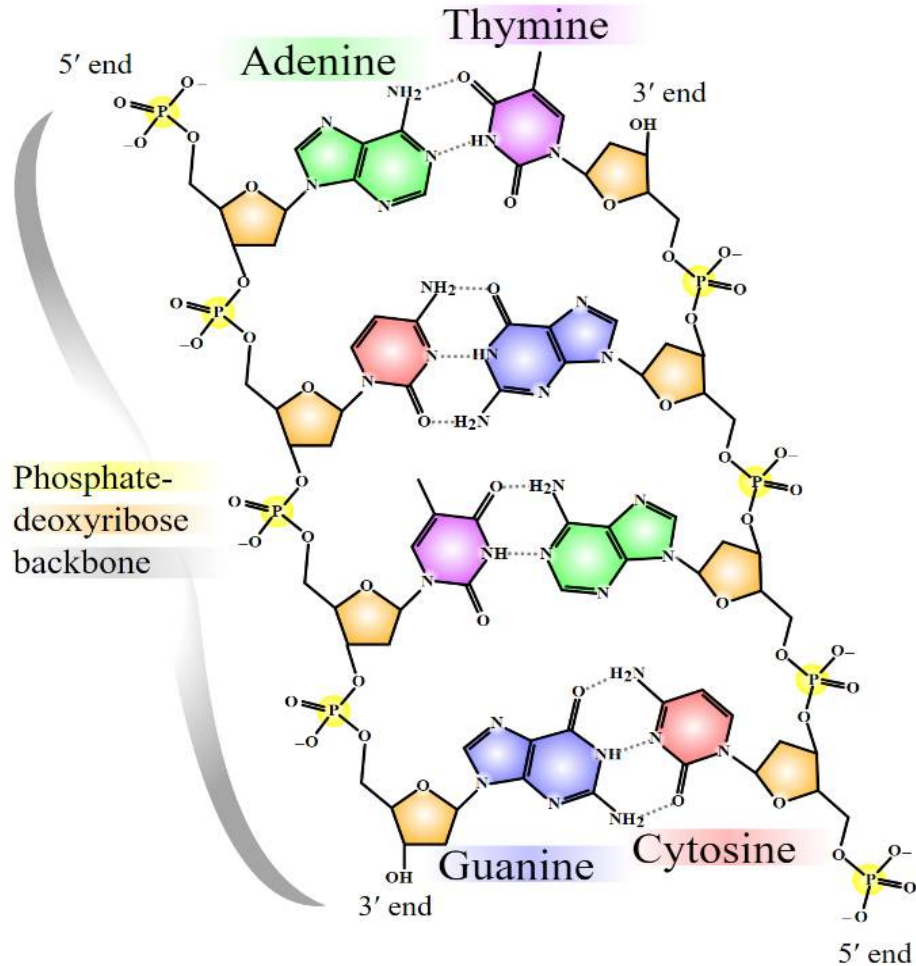
- ❖ Rákkutatáshoz kapcsolódó projektek nyilvános adatbázisainak kezelése
 - Adatok válogatása, szűrése
 - Rákos sejtvonalak genetikai tulajdonságainak elemzése
 - Rákos sejtvonalak genetikai tulajdonságainak összehasonlítása
- ❖ Felhasznált eszköz: az Edge böngészőben található Microsoft Copilot
 - Kérdésekkel irányított párbeszédés tanulás
 - Szövegesen specifikált programok elkészítése
- ❖ Programok:
 - Excel VBA script
 - Python program
 - MATLAB program

A DNS szerkezete

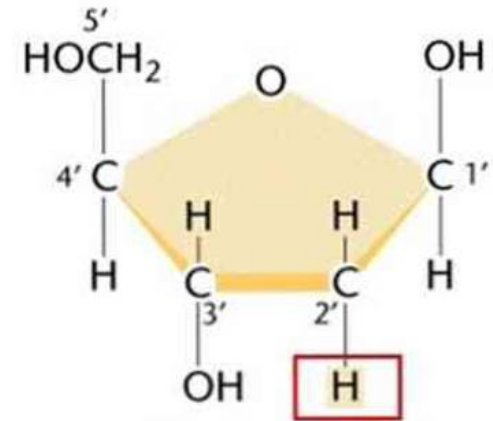


- ❖ a DNS ismétlődő nukleotid egységekből álló nagy méretű molekula (polimer)
- ❖ A DNS kettős hélix váza foszfodiészter kötéssel egymáshoz kapcsolódó dezoxiribózból áll (foszfát és cukor)
- ❖ A DNS változó része az egymást követő nukleotidok bázisainak a sorrendje (a bázisok is a kettős hélix váz dezoxiribóz komponenséhez kapcsolódnak)
- ❖ Adeninnel szemben csak timin, guaninnal szemben csak citozin állhat
- ❖ Az RNS csak egy hélixből áll, melyben dezoxiribóz helyett ribóz, timin helyett pedig uracil áll

A DNS szerkezete



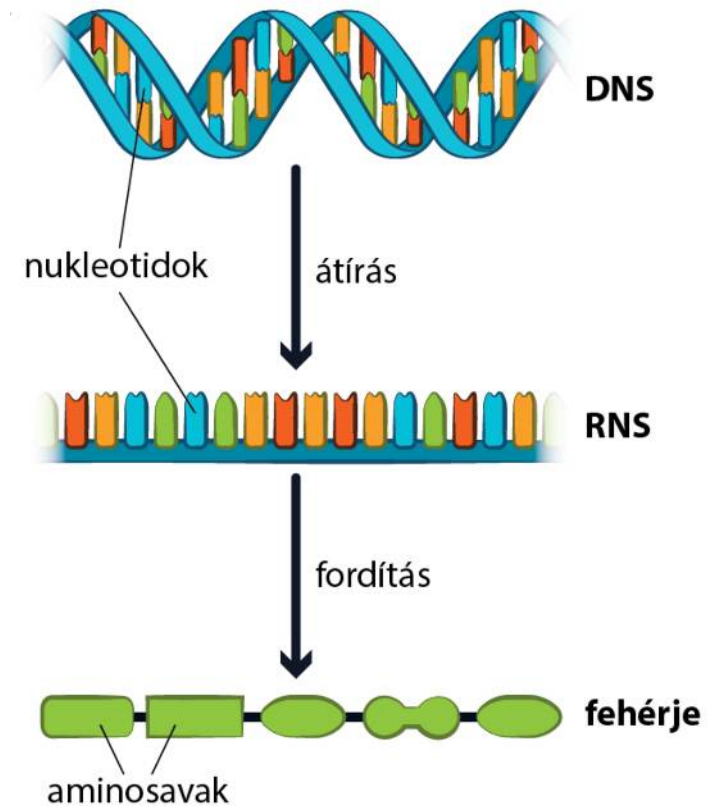
- ❖ A DNS-ben a dezoxiribóz molekula 3' és 5' szénatomjai játszanak szerepet a foszfodiészter kötések kialakításában
- ❖ A bázisok pedig a dezoxiribóz molekula 1' szénatomjához kapcsolódnak (glikozidos kötés)



2 - dezoxiribóz

Centrális dogma

- ❖ A molekuláris biológia alapja a **centrális dogma**, vagyis a genetikai információ-áramlás iránya: DNS → mRNS → fehérje
- ❖ A centrális dogma magába foglalja a transzkripciót (az átíródást: DNS → mRNS) és a transzlációt (az átfordítást: mRNS → fehérje)
- ❖ Az általunk vizsgált adatbázisban a rákos sejtvonalakat a DNS génjei átíratainak expressziós szintjeivel jellemzik, amelyeket TPM (Transcripts Per Million) egységben adnak meg. Az adatokat (hogy milyen intenzitással fordulnak elő átíratok) általában kutatók használják a különböző rákos sejtvonalak génexpressziós profiljainak összehasonlítására és elemzésére



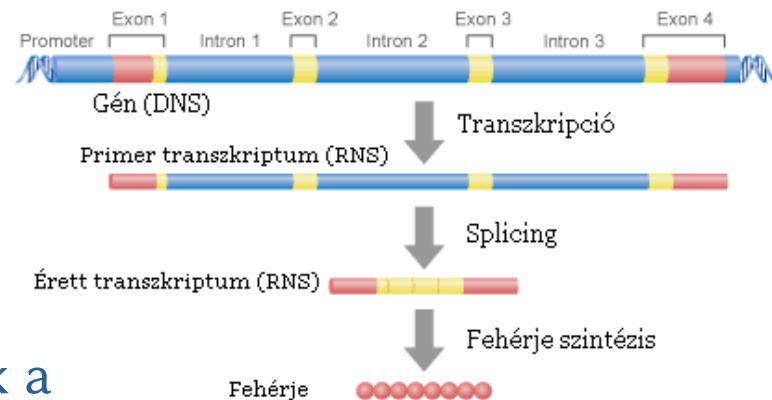
Kromoszómák és gének

- ❖ Az emberi DNS a sejtmagban található, és nem egyetlen hosszú molekulaként létezik, hanem több darabra van osztva, amelyeket **kromoszómáknak** nevezünk. Az emberi sejtekben 23 pár kromoszóma található, összesen 46 kromoszóma
- ❖ Molekuláris biológiai értelemben a **gén** a DNS olyan része, amelyet a sejt mRNS-be ír át, majd a fehérjeszintézis (transzláció) során, annak **egy részéből** (exonok) származó információ alapján, fehérjét készít
- ❖ **Genetikai jelek:** (a DNS-en belül jelzik a gének/génrészek határait)
 - **Promóter régiók:** olyan szekvenciák, amelyek a gén transzkripciójának megkezdéséhez szükségesek. Pl. a TATA-box egy gyakori promóter elem, amely általában így néz ki: 5'-TATAAA-3'.
 - **Exon-intron határok:** specifikus szekvenciák, amelyek segítenek azonosítani a gének különböző részeit. A gén kódoló részei az exonok, amelyek az mRNS-ben maradnak, az intronok nem kódoló szekvenciák, az mRNS éretté válása során eltávolításra kerülnek
 - **Stop kodonok:** A gének végén található stop kodonok jelzik a transzkripció befejezését, és segítenek meghatározni a gének végpontját

Génkifejeződés (génexpresszió)

- ❖ A génkifejeződés (génexpresszió) az a folyamat, amely során a génben tárolt információ aktívvá válik és funkcionális termékké, például fehérjévé vagy RNS-molekulává alakul
- ❖ A génkifejeződés dinamikus és adaptív folyamat, amelyet különböző tényezők befolyásolhatnak:
 - **Transzkripciós faktorok:** Fehérjék, amelyek a DNS-hez kötődve szabályozzák a gén transzkripcióját
 - **Epigenetikai módosítások:** DNS metiláció és hiszton módosítások, amelyek befolyásolják a gén hozzáférhetőségét.
 - **RNS interferencia:** Kis RNS molekulák, amelyek gátolják az mRNS transzlációját vagy stabilitását
- ❖ A génkifejeződést befolyásoló tényezők rákbetegség hatására is megváltozhatnak, akadályozva a gének normális kifejeződését és szabályozását

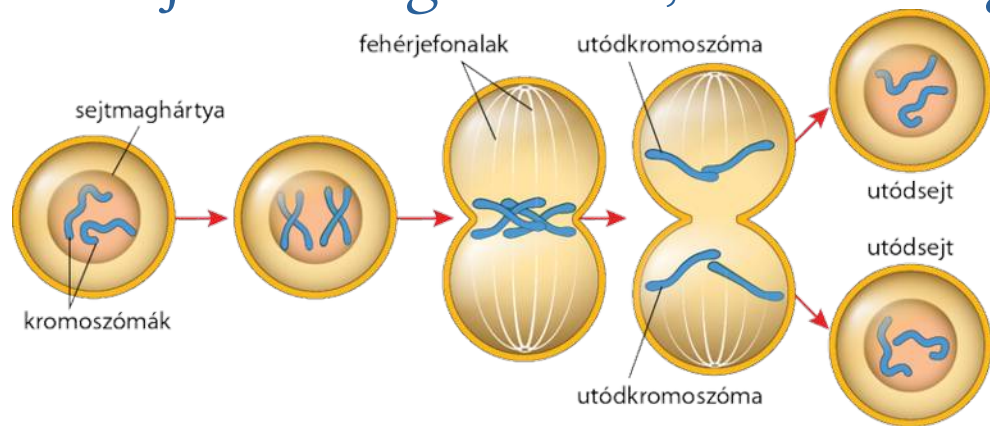
Génexpresszió



© Wellcome Trust

Sejtvonalak

- ❖ A **sejtvonal** olyan sejtek populációja, amelyeket laboratóriumi körülmények között tenyésztenek és fenntartanak (sejtklónozás).
- ❖ Ezek a sejtek képesek hosszú ideig osztódni és növekedni, így lehetővé teszik a kutatók számára, hogy különböző biológiai és orvosi vizsgálatokat végezzenek rajtuk.
- ❖ A sejtvonalak lehetnek normál sejtekből vagy rákos sejtekből származó sejtek, és gyakran használják őket genetikai, farmakológiai és biokémiai kutatásokban.



A CCLE projekt

- ❖ A **Cancer Cell Line Encyclopedia (CCLE) projekt** egy szoros együttműködés eredménye a Broad Intézet és a Novartis Intézetei között. A projekt célja a humán rákfajták sejtcsoportjainak részletes genetikai és farmakológiai jellemzése
- ❖ Az adatok nyilvánosan hozzáférhetőek
- ❖ **Főbb célkitűzések**
 - **Genetikai jellemzés:** A sejtcsoportok genetikai profiljainak elemzése, beleértve a DNS másolatokat és a mutációkat, több, mint 1000 rákfajtánál
 - **Farmakológiai tesztelés:** Az adott sejtcsoportok reakcióit különböző gyógyszerekre és mérgekre tesztelik.
 - **Prediktív modellezés:** A genetikai elemzések segítenek azonosítani a különböző farmakológiai sebezhetőségeket a genom mintázatokhoz.
 - **Kutatási eszköz:** Az adatok segítenek a rák sejtekben végzett kutatásokban és a gyógyszertervezésben.

Python bővítmények telepítése

❖ Az itt bemutatni kívánt **Python** programokhoz a **Python 3.12.7** verzióját használtam és a szükséges bővítményeket (könyvtárak) az alábbi parancsokkal telepítettem:

- `pip install pandas`
- `pip install matplotlib`
- `pip install numpy`
- `pip install openpyxl`
- `pip install seaborn`

Első próbálkozás

- ❖ **Kérdés:** Ismered a **CCLE_RNAseq_rsem_genes_tpm_20180929.txt.gz** állományt?
- ❖ **Válasz:** Igen, a **CLE_RNAseq_rsem_genes_tpm_20180929.txt.gz** egy adatállomány, amely a Cancer Cell Line Encyclopedia (CCLE) projekt részeként készült. Ez az állomány tartalmazza a különböző rákos sejtvonalak génexpressziós adatait, amelyeket **RNA-seq** technológiával mértek és **RSEM** szoftverrel elemeztek. Az adatok **TPM** (Transcripts Per Million) formátumban vannak, ami egy normalizált érték a génexpresszió mérésére
- ❖ Az állományban található oszlopok általában a következő információkat tartalmazzák:
 - **gene_id:** Az Ensembl génazonosítója (pl. ENSG000000003.10).
 - **transcript_ids:** Az adott génhez tartozó transzkriptumok azonosítói (pl. ENST00000373020.4).
 - **TPM értékek:** Az egyes sejtvonalakban (ebben a fájlban 1019 szerepel) mért TPM értékek, amelyek a génexpresszió szintjét mutatják
- ❖ Az adatokat általában kutatók használják a különböző rákos sejtvonalak génexpressziós profiljainak összehasonlítására és elemzésére.

Adatválogatás egy sejtvonalon belül

- ❖ Melyek azok a gének, amelyek pl. az **SNU475_LIVER** sejtvonalonban a legmagasabb a TPM értékkel szerepelnek? Szűrjük ki az **SNU475_LIVER** fejlécű oszlopból a 10 legmagasabb értéket, és listázzuk ki az első oszlopból a hozzá tartozó **gene_id**-t. Bónuszként a hozzá tartozó **transcript_ids**-t is kiírjuk a második oszlopból

- ❖ Eredmény:

```
TPM érték: 49636.0, Gene ID: ENSG00000228253.1, Transcript ID: ENST00000361851.1
TPM érték: 28412.0, Gene ID: ENSG00000087086.9, Transcript ID: ENST00000331825.6
TPM érték: 20955.0, Gene ID: ENSG00000198712.1, Transcript ID: ENST00000361739.1
TPM érték: 15475.0, Gene ID: ENSG00000198804.2, Transcript ID: ENST00000361624.2
TPM érték: 14877.0, Gene ID: ENSG00000198938.2, Transcript ID: ENST00000362079.2
TPM érték: 12571.0, Gene ID: ENSG00000198899.2, Transcript ID: ENST00000361899.2
TPM érték: 10061.0, Gene ID: ENSG00000212907.2, Transcript ID: ENST00000361335.1
TPM érték: 9686.7, Gene ID: ENSG00000198886.2, Transcript ID: ENST00000361381.2
TPM érték: 9496.3, Gene ID: ENSG00000210082.2, Transcript ID: ENST00000387347.2
TPM érték: 6726.6, Gene ID: ENSG00000166710.13, Transcript ID: ENST00000349264.6,
ENST00000544417.1,ENST00000557901.1,ENST00000558401.1,ENST00000559220.1,
ENST00000559720.1,ENST00000559907.1,ENST00000559916.1,ENST00000560556.1,
ENST00000560681.1,ENST00000561139.1,ENST00000561424.1
```

SNU475_LIVER.py

- ❖ Az adott sejtvonalonban megkeressük és kiíratjuk a 10 legmagasabb TPM értéket

```
import pandas as pd

# Fájlból beolvasása
file_path = 'CCLE_RNAseq_rsem_genes_tpm_20180929.txt.gz'
data = pd.read_csv(file_path, compression='gzip', sep='\t')

# Az első két oszlop kiválasztása
gene_ids = data.iloc[:, 0]
transcript_ids = data.iloc[:, 1]

# Az SNU475_LIVER oszlop kiválasztása
snu475_liver = data['SNU475_LIVER']

# A 10 legmagasabb TPM érték megtalálása
top_10_indices = snu475_liver.nlargest(10).index

# Eredmények kiírása
for idx in top_10_indices:
    tpm_value = snu475_liver[idx]
    gene_id = gene_ids[idx]
    transcript_id = transcript_ids[idx]
    print(f"TPM érték: {tpm_value}, Gene ID: {gene_id}, Transcript ID: {transcript_id}")
```

TAB-bal szeparált,
gzip-pel tömörített
szövegfájlt olvasunk
be

Egy sejtvonala kiemelése a nagy táblázatból

- ❖ **Kérdés:** Redukáljuk a letöltött `CCLE_RNAseq_rsem_genes_tpm_20180929.txt.gz` fájl eredeti táblázatát úgy, hogy csak az 1., 2. és az `SNU475_LIVER` fejlécű oszlopot hagyjuk meg és elhagyjuk belőle a nulla TPM értéket tartalmazó sorokat is. Cseréljük fel a sejtvonala és a `transcripts_ids` oszlopok sorrendjét, majd írjuk ki egy fájlba a redukált táblázatot Excel (.XLSX) formátumban!
- ❖ **Válasz:**
 - Az `SNU475_LIVER_all.py` nevű Python program a `pd.read_csv` függvénnyel beolvassa a tömörített fájlt
 - Kiválaszta a `gene_id`, `transcript_id` és `SNU475_LIVER` oszlopokat.
 - Eltávolítja azokat a sorokat, amelyekben a `SNU475_LIVER` oszlop értéke nulla
 - Felcseréli a `SNU475_LIVER` és a `transcript_id` oszlopokat
 - Az `output_file` változóban megadott fájlnevet használva mentjük az eredményt .XLSX formátumban az `openpyxl` modul segítségével
- ❖ Az eredmény a `reduced_data.xlsx` táblázatban található

SNU475_LIVER_all.py

- ❖ Az eredmény csak a `gene_id`, `SNU475_LIVER` és a `transcript_ids` oszlopokat fogja tartalmazni, s csak azokat a sorokat, amelyekben nem nulla TPM értékek vannak

```
import pandas as pd

# Fájl beolvasása
file_path = 'CCLE_RNAseq_rsem_genes_tpm_20180929.txt.gz'
data = pd.read_csv(file_path, compression='gzip', sep='\t')

# Az 1., 2. és az SNU475_LIVER oszlopok kiválasztása
reduced_data = data[['gene_id', 'transcript_ids', 'SNU475_LIVER']]

# Nullák eltávolítása
reduced_data = reduced_data[reduced_data['SNU475_LIVER'] != 0]

# Oszlopok felcserélése
reduced_data = reduced_data[['gene_id', 'SNU475_LIVER', 'transcript_ids']]

# Eredmény mentése XLS formátumban
output_file = 'reduced_data.xlsx'
reduced_data.to_excel(output_file, index=False)

print(f"A redukált táblázat mentve lett az {output_file} fájlba.")
```

Az összes májjal kapcsolatos sejtvonala kiválasztása

- ❖ A letöltött `CCLE_RNAseq_rsem_genes_tpm_20180929.txt.gz` fájl beolvasott táblázatból hagyjuk meg az 1. és a 2. oszlopot, valamint az összes olyan sejtvonala, amelynek fejlécében szerepel a 'LIVER' szó. Az eredményt mentjük el a `liver_cancers.xlsx` fájlba (excel formátum) úgy, hogy a `transcript_ids` fejlécű oszlopot tegyük át a második oszlopból az utolsó helyre!
- ❖ A `liver_cancers.py` program beolvassa a tömörített fájlt, kiválasztja a `gene_id`, `transcript_ids` és az összes olyan oszlopot, amelyben szerepel a "LIVER" szó
- ❖ Eltávolítja a sorokat, amelyekben az összes "LIVER" oszlop TPM értéke nulla
- ❖ Áthelyezzük a `transcript_ids` oszlopot, hogy az utolsó helyre kerüljön
- ❖ Az `output_file` változóban megadott fájlnevet használva (`reduced_data.xlsx`), az `openpyxl` modul segítségével elmentjük az eredményt XLSX formátumban
- ❖ Megjegyzés: az `openpyxl` modult nem kell importálni, a `pandas` automatikusan felismeri

liver_cancers.py

- ❖ Az eredmény csak a `gene_id`, az `xxx_LIVER` és a `transcript_ids` oszlopokat fogja tartalmazni, s csak azokat a sorokat, amelyekben nem nulla TPM értékek vannak

```
import pandas as pd
# Fájlból beolvasása
file_path = 'CCLE_RNAseq_rsem_genes_tpm_20180929.txt.gz'
data = pd.read_csv(file_path, compression='gzip', sep='\t')

# Az 1., 2. és az összes "LIVER" szót tartalmazó oszlop kiválasztása
columns_to_keep = ['gene_id', 'transcript_ids'] + [col for col in data.columns if 'LIVER' in col]
reduced_data = data[columns_to_keep]

# Nullák eltávolítása
reduced_data = reduced_data[(reduced_data.iloc[:, 2:] != 0).any(axis=1)]

# Oszlopok felcserélése
reduced_data = reduced_data[['gene_id'] + [col for col in reduced_data.columns if col !=
'transcript_ids'] + ['transcript_ids']]

# Eredmény mentése XLS formátumban
output_file = 'liver_cancers.xlsx'
reduced_data.to_excel(output_file, index=False, engine='openpyxl')
print(f"A redukált táblázat mentve lett az {output_file} fájlba.")
```

A májjal kapcsolatos sejtvonalak összehasonlítása

- ❖ Kérdés: Nézzük pl. az `xxx_LIVER` (ahol `xxx` tetszőleges szöveg) oszlopokat, hogyan tudnánk megállapítani, hogy melyek között található nagyobb fokú egyezés, vagy hasonlóság?
- ❖ Válasz: Az egyik lehetséges megközelítés a **korrelációs mátrix** kiszámítása, amely megmutatja, hogy az egyes oszlopok (sejtvonalak) mennyire hasonlítanak egymásra. A korrelációs együttható értéke -1 és 1 között mozog, ahol 1 tökéletes pozitív korrelációt, -1 tökéletes negatív korrelációt, és a 0 a korreláció hiányát jelenti. A korrelációs mátrixot a **liver_cancers_correlations.py** nevű programban egy ún. hő térképen (heat map) mutatjuk be
- ❖ A korrelációs mátrix hő térképét a **seaborn** könyvtár **heatmap** függvényével jelenítjük meg. A skálát a minimális és maximális értékekhez igazítjuk, az annotációkat (számokat) kikapcsolva, az oszlopneveket 45 fokban elforgatva, és elmentjük az ábrát **liver_cancer_corr.png** néven

liver_cancers_correlations.py – 2/1.

- ❖ A program első felében az előző programhoz hasonlóan kiválogatjuk a megfelelő oszlopokat és a legalább egy nem nulla TPM értéket tartalmazó sorokat

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Fájl beolvasása
file_path = 'CCLE_RNAseq_rsem_genes_tpm_20180929.txt.gz'
data = pd.read_csv(file_path, compression='gzip', sep='\t')

# Az 1., 2. és az összes "LIVER" szót tartalmazó oszlop kiválasztása
columns_to_keep = ['gene_id', 'transcript_ids'] + [col for col in data.columns if 'LIVER' in col]
reduced_data = data[columns_to_keep]

# A csupa nullákat tartalmazó sorok eltávolítása
reduced_data = reduced_data[(reduced_data.iloc[:, 2:] != 0).any(axis=1)]

# Csak a "LIVER" oszlopok kiválasztása a korrelációs mátrixhoz
liver_columns = [col for col in reduced_data.columns if 'LIVER' in col]
liver_data = reduced_data[liver_columns]
```

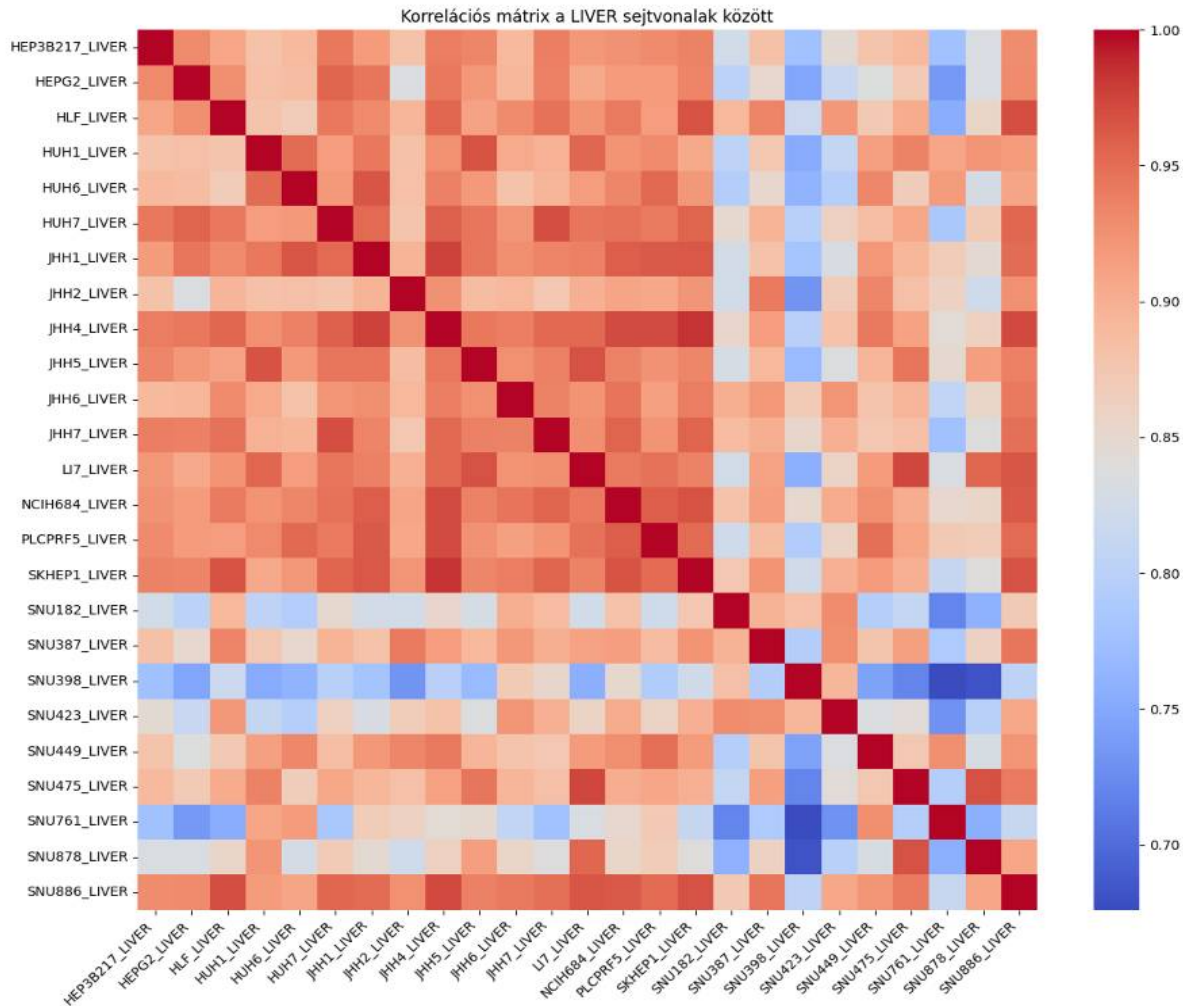
liver_cancers_correlations.py – 2/2.

```
# Korrelációs mátrix létrehozása
correlation_matrix = liver_data.corr()

# Korrelációk maximális és minimális értékeinek meghatározása
min_corr = correlation_matrix.min().min()
max_corr = correlation_matrix.max().max()

# Korrelációs mátrix megjelenítése hő térképen és mentése
plt.figure(figsize=(14, 12))
sns.heatmap(correlation_matrix, annot=False, cmap='coolwarm', vmin=min_corr,
vmax=max_corr)
plt.title('Korrelációs mátrix a LIVER sejt vonalak között')
plt.xticks(rotation=45, ha='right')
plt.yticks(rotation=0)
plt.tight_layout(pad=5.0)
plt.savefig('liver_cancer_corr.png')
plt.show()

# Korrelációk maximális és minimális értékeinek kiírása
print(f"Korrelációk minimális értéke: {min_corr}")
print(f"Korrelációk maximális értéke: {max_corr}")
```



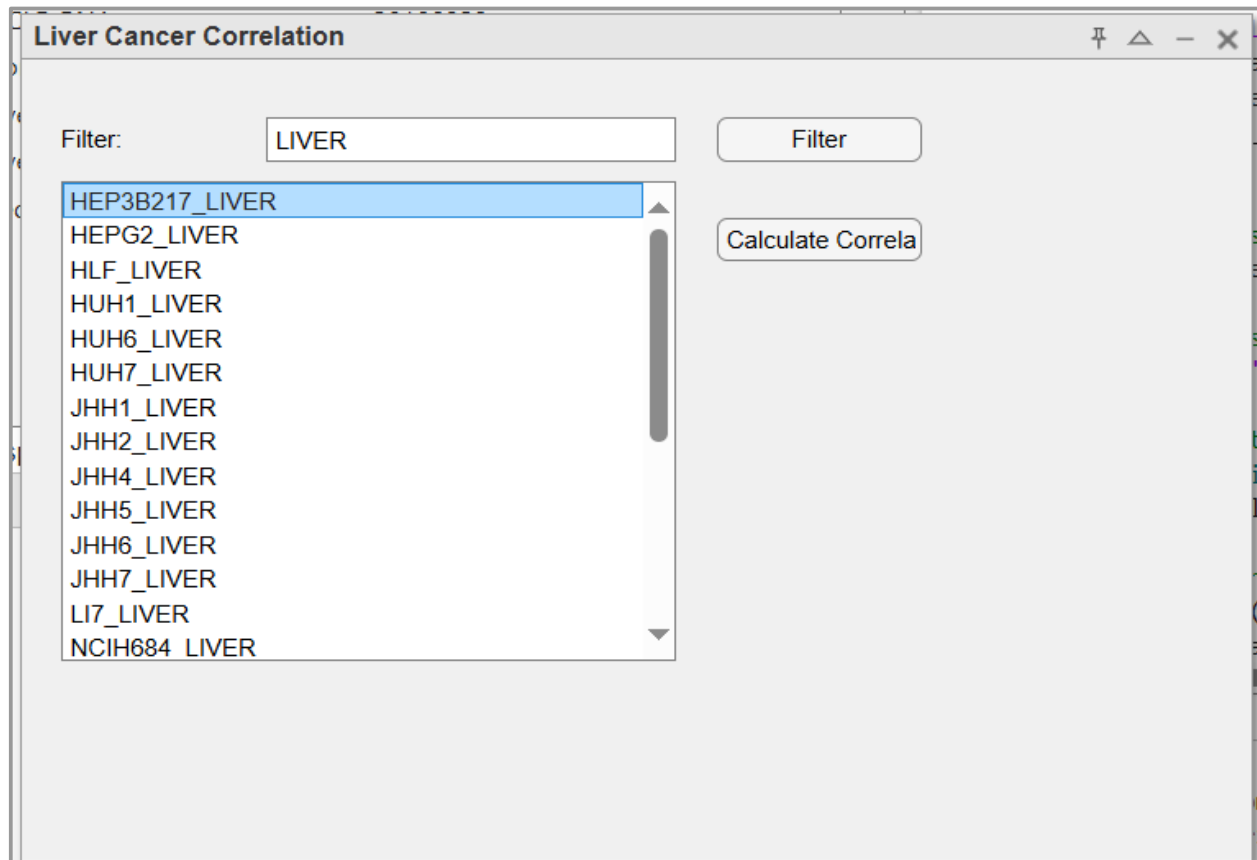
❖ A hőtérképen a kék érték a minimumhoz közeli, a barnáspiros a maximumhoz közeli korrelációs értéket jelöli

Interaktív adatválogatás MATLAB-ban

- ❖ Az előző programot írassuk át MATLAB-ra és a kiválogatást szervezzük meg interaktív módon!
- ❖ Az online MATLAB fejlesztői felületet fogjuk használni, ami ingyenes **MathWorks** regisztrációval használható (a Thingspeak fiókhoz kapcsolódó **MathWorks** regisztráció is megfelel)
- ❖ Az online felületre belépve hozzunk létre új scriptet (ne Live Scriptet!) és abba másoljuk bele a Copilot által írt programot
- ❖ A **.text.gz** kiterjesztésű állományt beolvasáskor ki kell tömöríteni a **gunzip()** függvénnyel, de azt csak akkor kell elvégezni, ha a kicsomagolt fájl még nem létezik

liver_cancer_corr_gui2.m

- ❖ A beíró ablakban egy szűrőfeltételt adhatunk meg, amit a **Filter** gombbal aktiválhatunk
- ❖ A legördülő listából egy vagy több elemet választhatunk ki (utóbbi esetben a **Ctrl** gombbal kombinált kattintással), de tartományt is kijelölhetünk a **Shift** gombbal kombinált kattintással
- ❖ A **Calculate Correlations** gomb a számolást és ábrázolást indítja el



liver_cancer_corr_gui2.m – 3/1.

```
function liver_cancer_corr_gui
% Fájlf kicsomagolása, ha szükséges
file_path = 'CCLE_RNAseq_rsem_genes_tpm_20180929.txt.gz';
unzipped_file_path = 'CCLE_RNAseq_rsem_genes_tpm_20180929.txt';
if exist(unzipped_file_path, 'file') ~= 2
    gunzip(file_path);
end

% Fájlf beolvasása
data = readtable(unzipped_file_path, 'FileType', 'text', 'Delimiter', '\t',
'ReadVariableNames', true);

% GUI létrehozása
fig = uifigure('Name', 'Liver Cancer Correlation', 'Position', [100, 100, 600, 400]);

% Beíró mező létrehozása
lbl = uilabel(fig, 'Position', [20, 350, 100, 22], 'Text', 'Filter:');
txt = uieditfield(fig, 'text', 'Position', [120, 350, 200, 22]);
```

liver_cancer_corr_gui2.m – 3/2.

```
% Listbox létrehozása
lst = uilistbox(fig, 'Position', [20, 100, 300, 240], 'Multiselect', 'on');
lst.Items = data.Properties.VariableNames;

% Szűrés gomb létrehozása
btnFilter = uibutton(fig, 'Position', [340, 350, 100, 22], 'Text', 'Filter', ...
    'ButtonPushedFcn', @(btn, event) filterColumns(txt, lst, data));

% Korrelációs gomb létrehozása
btnCorr = uibutton(fig, 'Position', [340, 300, 100, 22], 'Text', 'Calculate
Correlation', ...
    'ButtonPushedFcn', @(btn, event) calculateCorrelation(lst, data));

% Szűrés funkció
function filterColumns(txt, lst, data)
    filterText = txt.Value;
    filteredItems =
data.Properties.VariableNames(contains(data.Properties.VariableNames, filterText,
'IgnoreCase', true));
    lst.Items = filteredItems;
end
```

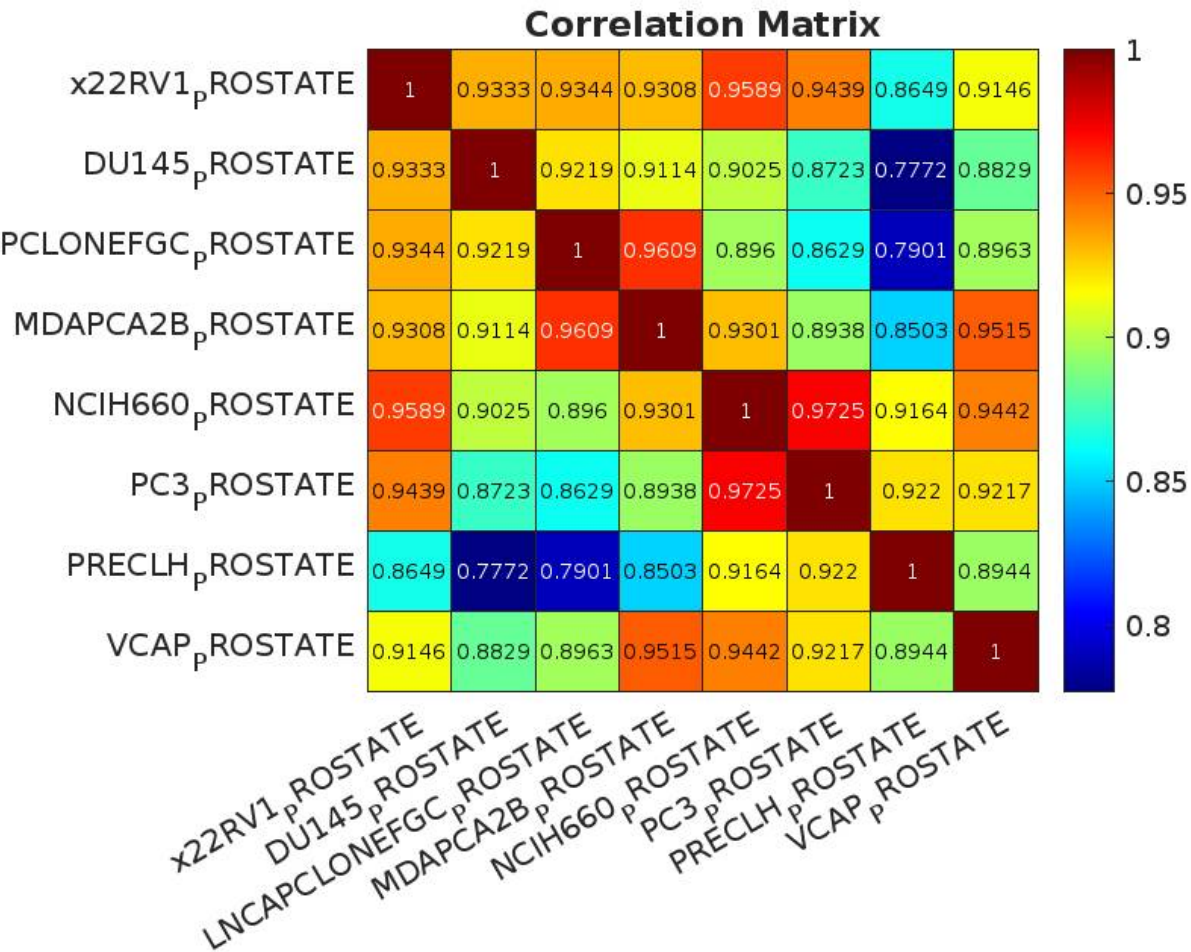
liver_cancer_corr_gui2.m – 3/3.

```
function calculateCorrelation(lst, data)    % Korrelációs számítás funkció
    selectedColumns = lst.Value;
    if isempty(selectedColumns)
        uialert(fig, 'Please select at least one column.', 'Error');
        return;
    end
    selectedData = data(:, selectedColumns); % A kijelölt oszlopok kiválasztása
    % Redukált táblázat mentése Excel fájlba
    reducedTable = [data(:, {'gene_id'}), selectedData, data(:, {'transcript_ids'})];
    writetable(reducedTable, 'reduced_data.xlsx');

    % Korrelációs mátrix létrehozása
    correlationMatrix = corr(table2array(selectedData), 'Rows', 'complete');
    figure;
    h = heatmap(selectedColumns, selectedColumns, correlationMatrix, 'Colormap', jet,
'ColorbarVisible', 'on');
    title('Correlation Matrix');
    saveas(gcf, 'correlation_matrix.png');
end
end
```

liver_cancer_corr_gui2.m: eredmény

- ❖ Az alábbi ábrához a 'PROSTATE' szöveget tartalmazó fejlécű oszlopokat válogattuk ki, és ezek korrelációs mátrixát láthatjuk a hőtérképen
- ❖ A legszorosabb korreláció itt az X22RV1_PROSTATE és a NCIH660_PROSTATE sejtvonalak között mutatkozik



Adatok válogatása Excel-ben

❖ Adatok előkészítése:

- Nyisd meg az Excel fájlt, amely tartalmazza a **TAB** szeparált **CCLE_RNAseq_rsem_genes_tpm_20180929.txt.gz** adatokat.

❖ Adatok szűrése:

- Jelöld ki az adatokat tartalmazó táblázatot.
- Menj a "**Data**" fülre, és válaszd a "**Filter**" opciót. Ez hozzáad egy legördülő menüt minden oszlop fejlécéhez.

❖ Szűrő beállítása:

- Az oszlopok fejlécében található legördülő menük segítségével szűrheted az adatokat. Az Excel lehetőséget biztosít arra, hogy szöveges szűrőket alkalmazz, például "Contains" (Tartalmaz) vagy "Begins With" (Kezdődik).

Adatok válogatása Excel-ben

❖ Kezelői felület létrehozása:

- Hozz létre egy új munkalapot a kezelői felület számára.
- Adj hozzá két legördülő menüt, egyet a gének és egyet a sejtvonalak számára. Ezt a "Data Validation" funkcióval teheted meg:
- Menj a "Data" fülre, és válaszd a "Data Validation" opciót.
- A "Settings" fülön válaszd a "List" opciót, és add meg a gének vagy sejtvonalak listáját.

❖ Szűrő beállítása VBA makróval:

- Nyomd meg az "Alt + F11" billentyűkombinációt a VBA szerkesztő megnyitásához.
- Hozz létre egy új modult, és másold be a következő oldalon bemutatott kódot:

Adatok válogatása Excel-ben

```
Sub FilterData()  
    Dim ws As Worksheet  
    Set ws = ThisWorkbook.Sheets("Adatok") ' Az adatok munkalap neve  
    Dim gene As String  
    Dim cellLine As String  
    ' A gének és sejtvonalak kiválasztása a kezelői felületről  
    ' A gének legördülő menü cellája  
    gene = ThisWorkbook.Sheets("KezelőiFelület").Range("B2").Value  
    ' A sejtvonalak legördülő menü cellája  
    cellLine = ThisWorkbook.Sheets("KezelőiFelület").Range("B3").Value  
    ' Szűrés alkalmazása  
    ws.Range("A1").AutoFilter Field:=1, Criteria1:="*" & gene & "*"  
    ws.Range("A1").AutoFilter Field:=2, Criteria1:="*" & cellLine & "*"  
End Sub
```

- ❖ **Menj vissza az Excel munkalapra, és adj hozzá egy gombot a kezelői felülethez:**
 - **Menj a "Developer" fülre, és válaszd az "Insert" opciót, majd válaszd a "Button" (Gomb) opciót.**
 - **Helyezd el a gombot a munkalapon, és rendeld hozzá a "FilterData" makrót.**