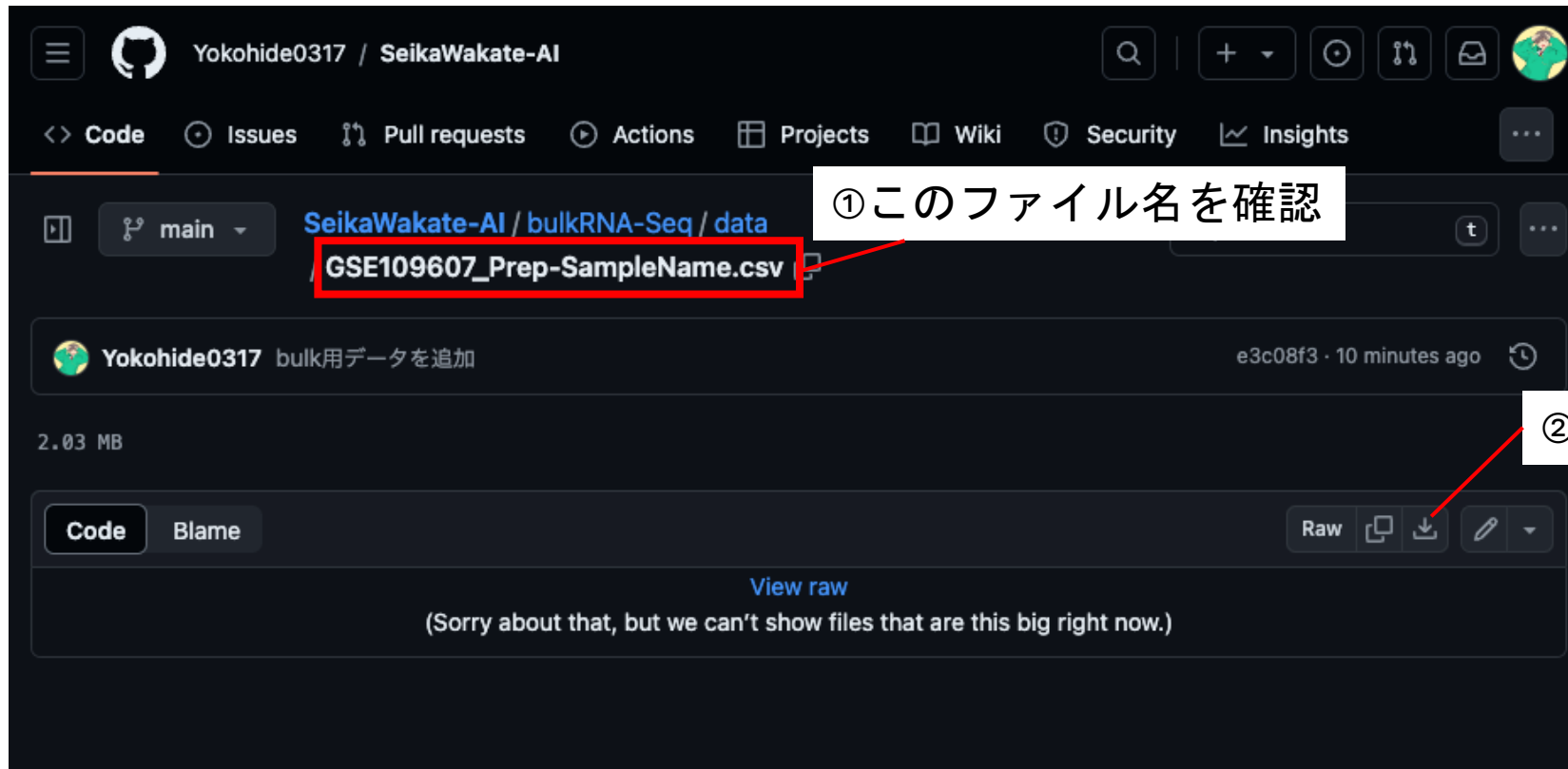


## 2. ChatGPTを活用！Pythonでデータ解析

### データのダウンロード方法



[https://github.com/Yokohide0317/SeikaWakate-AI/blob/main/bulkRNA-Seq/data/GSE109607\\_Prep-SampleName.csv](https://github.com/Yokohide0317/SeikaWakate-AI/blob/main/bulkRNA-Seq/data/GSE109607_Prep-SampleName.csv)

# Google Colabの使い方 - Google Driveに入る

Google Drive

 or <https://drive.google.com/drive/u/0/my-drive>

ドライブ

+ 新規

マイドライブ

パソコン

共有アイテム

最近使用したアイテム

スター付き

スパム

ゴミ箱

保存容量

15 GB 中 60.2 MB を使用

保存容量を増やす

ドライブで検索

マイドライブ

ファイル形式 ユーザー 最終更新

さまざまなファイルを保存できます

Google ドキュメント、スプレッドシート、スライド、

Microsoft Office など数々のファイル

ファイルやフォルダをドライブに直接ドラッグできます

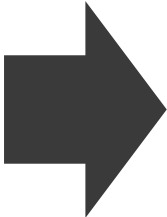
?

⚙

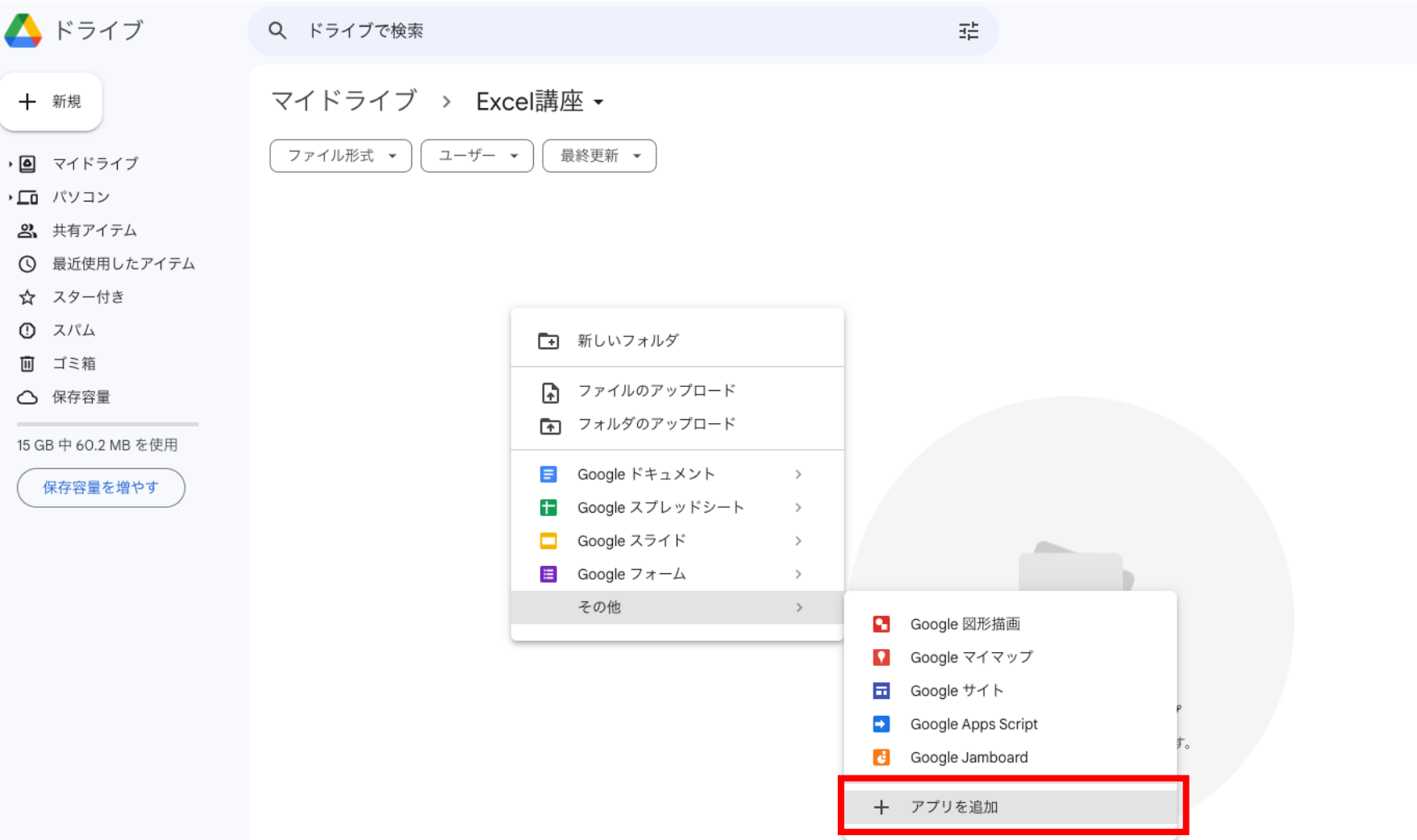
⋮

英裕

# Google Colabの使い方 - 作業フォルダの作成



# Google Colabの使い方 - Google Colabのインストール



「アプリを追加」をクリック

# Google Colabの使い方 - Google Colabのインストール

Google Workspace Marketplace

colab  
Colaboratory

すべてのフィルタ (1)

Google ドライブに対応

料金

どこでも利用可能な仕事用アプリ

これらのアプリは、組織の生産性向上に役立ちます。

miro

Miro

Miro ★ 4.5 407万+

Docu

DocuSign eSignature for Googl...

DocuSign ★ 3.8 320万+

Signeasy eSignatures for Googl...

Signeasy ★ 4.4 54万+

ビジネスの基本アプリ

これらのアプリは、ワークフローの生産性向上に役立ちます。

AUTOMAGICAL FORMS

Fax Your Document

colab

と入力&検索

Google Workspace Marketplace

colab

すべてのフィルタ (1)

Google ドライブに対応

料金

検索結果: colab

レビューや評価を Google が確認することはありません。

CO

Colaboratory

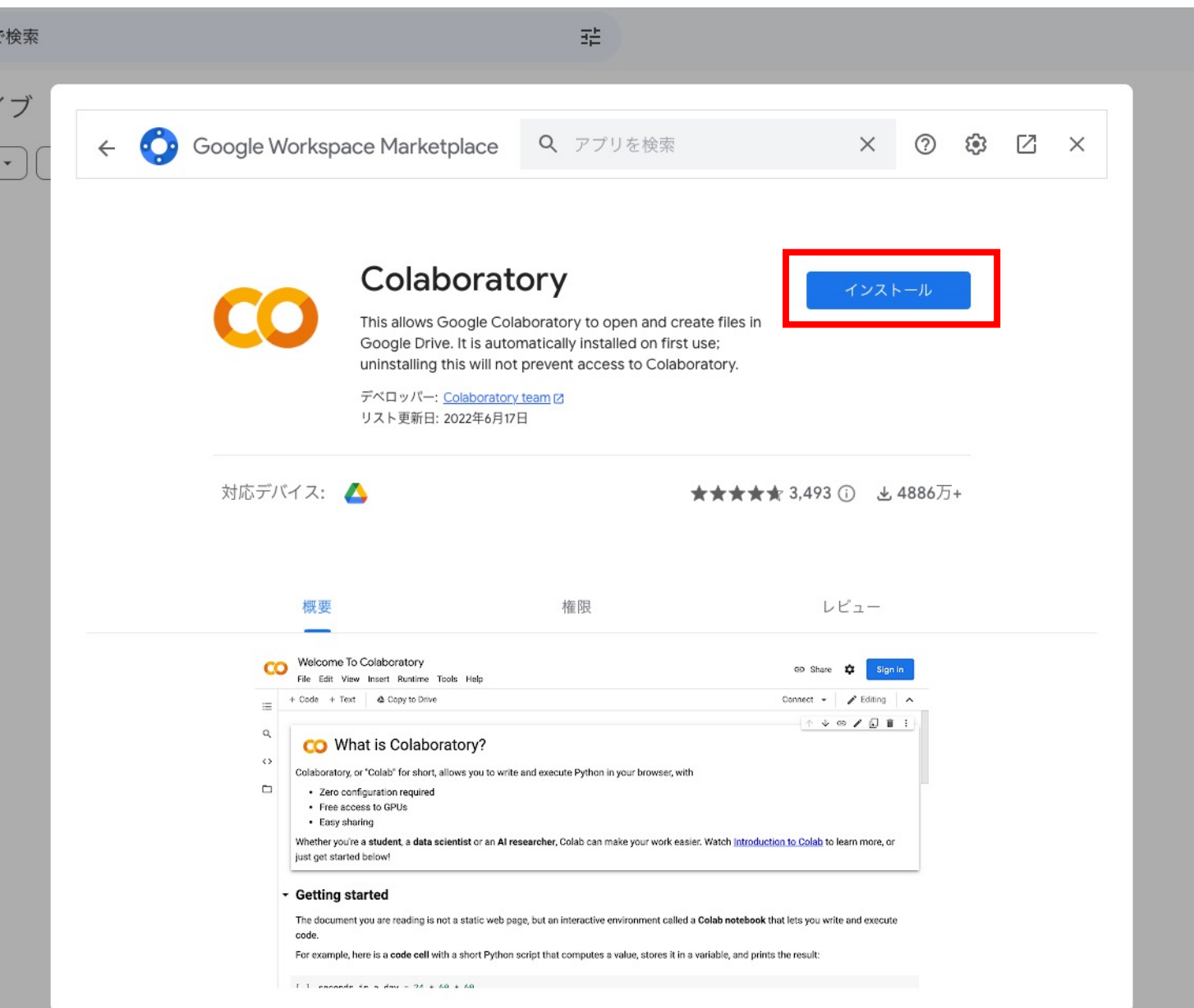
Colaboratory team

This allows Google Colaboratory to open and create files in Google Drive. It i...

★ 4.7 4886万+

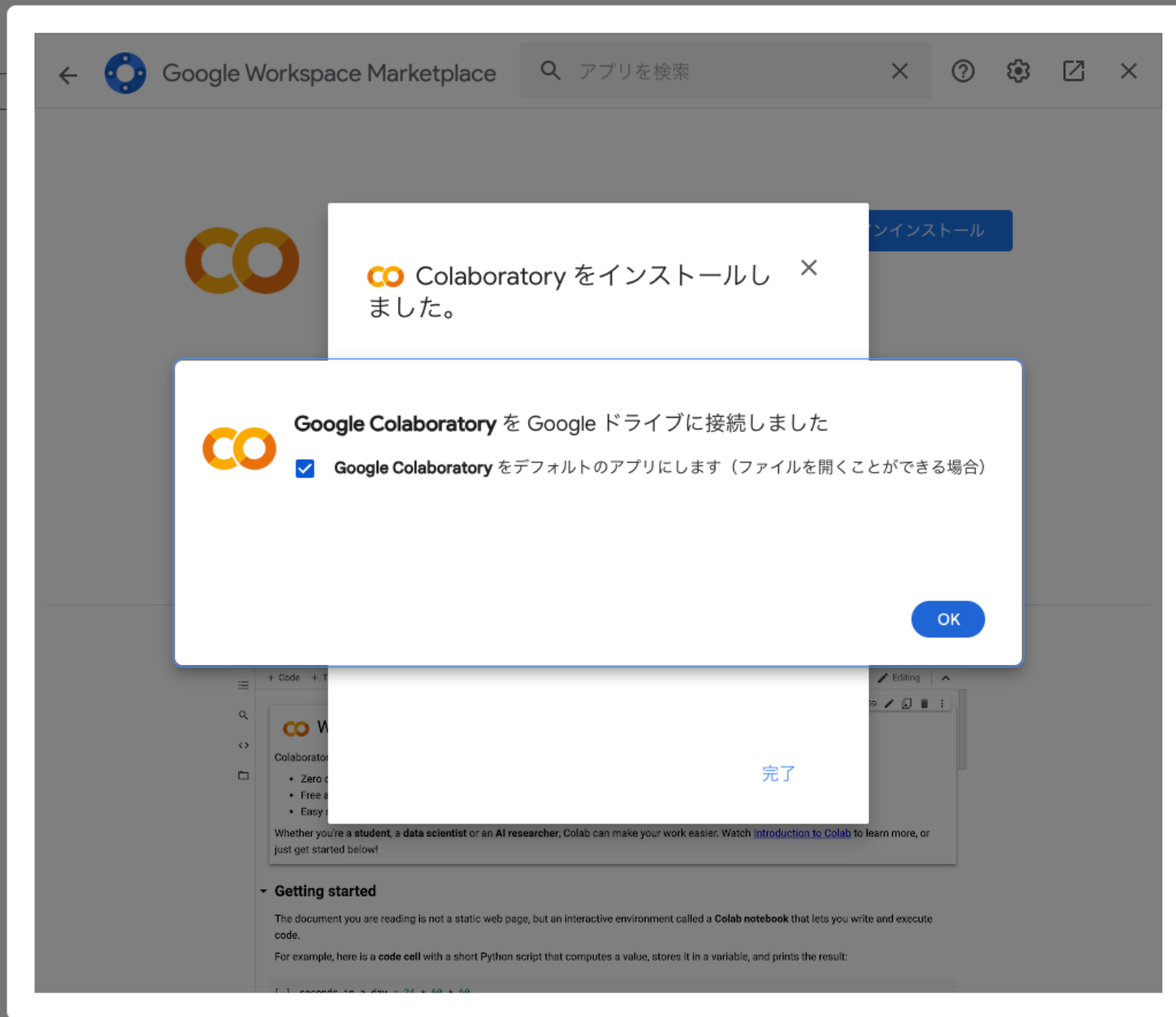
検索結果

# Google Colabの使い方 - Google Colabのインストール



インストールする。

※Google アカウントの認証が必要かもしれません。



「OK」や「完了」  
を押して、進む。

Google Driveの画面まで戻る。

# Google Colabの使い方 - インストールの確認 & ノートブックの作成

Q ドライブで検索

≡

マイドライブ > Excel講座 ▾

ファイル形式 ▾

ユーザー ▾

最終更新 ▾

📁 新しいフォルダ

📁 ファイルのアップロード

📁 フォルダのアップロード

📄 Google ドキュメント >

📊 Google スプレッドシート >

📽 Google スライド >

📄 Google フォーム >

その他 >

1. 作成したフォルダの中にあるのを確認し、

➡ 2. あいているところを右クリック

3. 「その他」にカーソルをあわせて、

📄 Google 図形描画

📍 Google マイマップ

🌐 Google サイト

🔗 Google Apps Script

🔗 Google Colaboratory

🗺 Google Jamboard

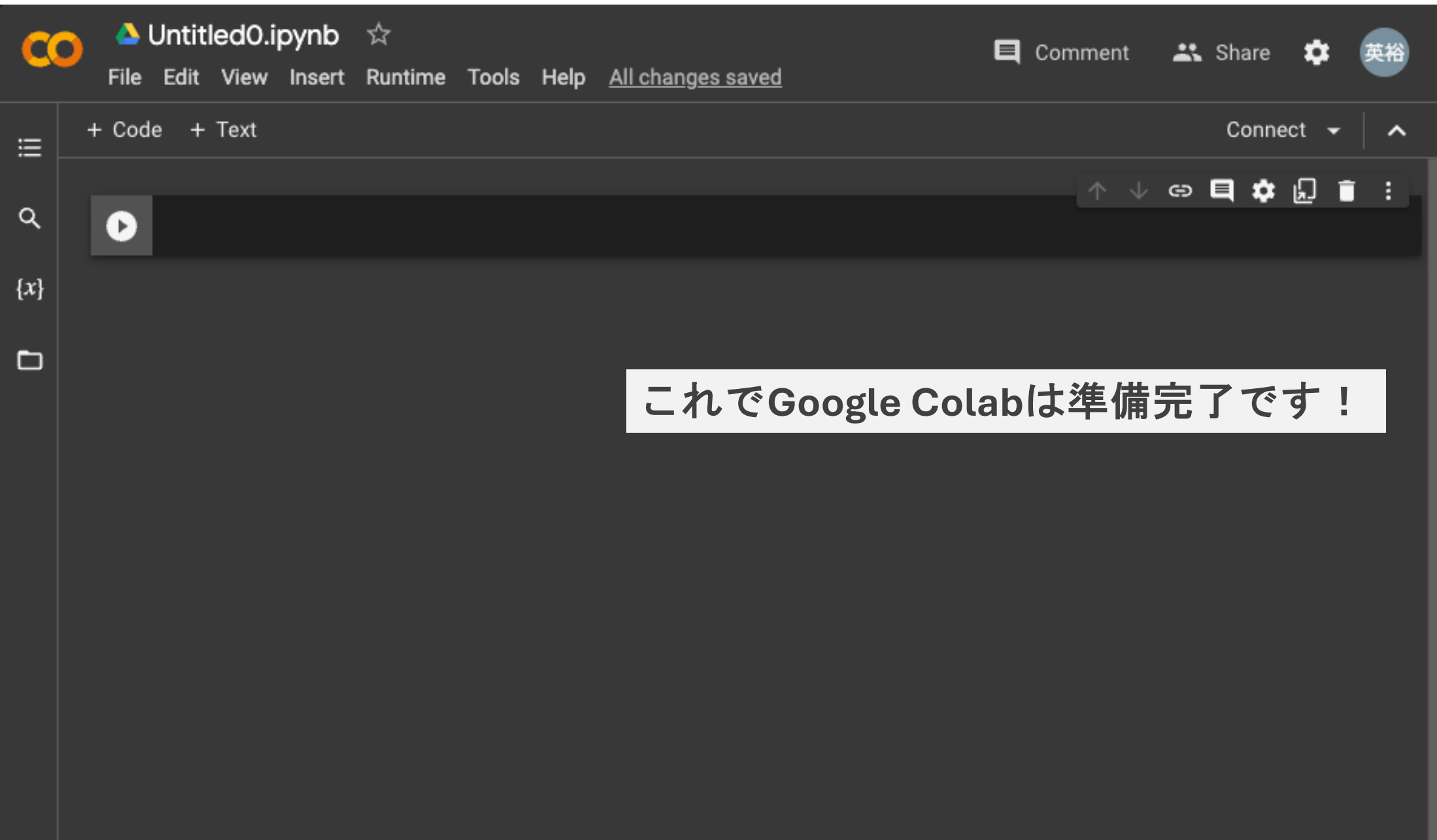
+ アプリを追加

4. 「Google Colaboratory」が候補に出てきたら、  
準備完了です！

クリックしてください。



# Google Colabの使い方 - ノートブックの作成



The screenshot displays the Google Colab web interface. At the top, the Google Colab logo is on the left, followed by the file name 'Untitled0.ipynb' and a star icon. To the right are buttons for 'Comment', 'Share', and a settings gear, along with a user profile icon labeled '英裕'. Below this is a menu bar with 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help', followed by the status 'All changes saved'. The main workspace has a toolbar with '+ Code' and '+ Text' buttons, and a 'Connect' button with a dropdown arrow. On the left sidebar, there are icons for a menu, search, and a file explorer showing a folder named '{x}'. The central area contains a code cell with a play button icon and a text box with the message 'これでGoogle Colabは準備完了です！' (Google Colab is now ready!).

Google Colabの使い方 - ノートブックの作成

Untitled0.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

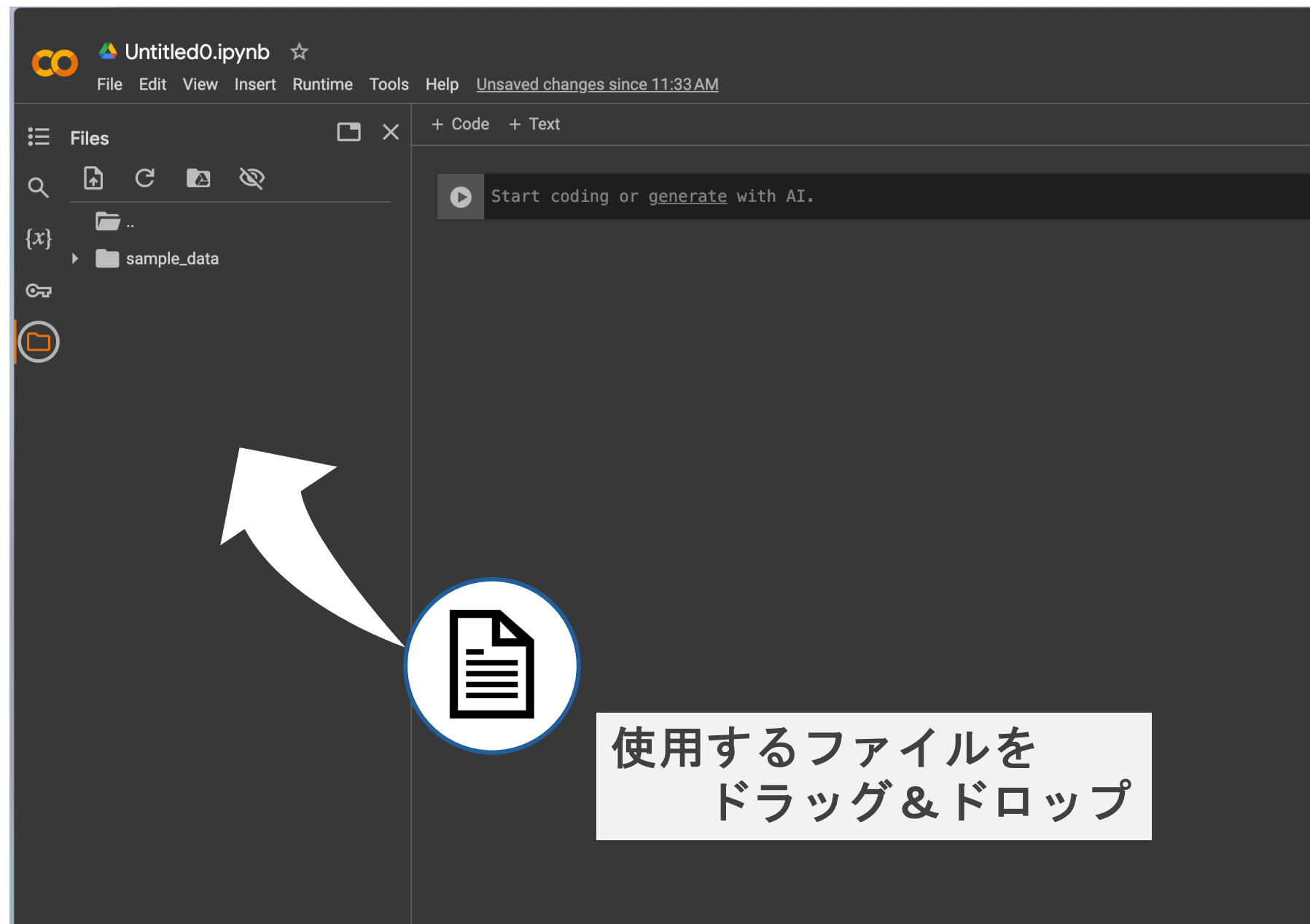
Comment Share 英裕

+ Code + Text Connect ^

↑ ↓ ↻ 💬 ⚙️ 📄 🗑️ ⋮

これでGoogle Colabは準備完了です！

# Google Colabの使い方 - ファイルのアップロード方法



The screenshot displays the Google Colab web interface. On the left, the 'Files' sidebar is visible, showing a file explorer with a folder named 'sample\_data'. A large white arrow points from a circular icon containing a document symbol to the 'sample\_data' folder. Below this icon, a text box contains the instruction: 使用するファイルを  
ドラッグ&ドロップ (Drag & drop the files you want to use). The main area of the interface shows a code editor with a prompt: 'Start coding or generate with AI.'

Google Colabの使い方 - ファイルのアップロード方法

使用するファイルを  
ドラッグ&ドロップ

# Google Colabの使い方 - ファイルのアップロード方法

Files

📁 ..

📁 sample\_data

📄 GSE109607\_TPM-prep-HT.csv

+ Code

+ Text

✓  
0s

[4] `import pandas as pd`  
  
`df = pd.read_csv("/content/GSE109607_TPM-prep-HT.csv", index_col=0)`  
`df`

🔄

	HT-C1	HT-C2	HT-OPC1	HT-OPC2	HT-GSE1	HT-GSE2
SFN	109.900	136.300	186.600	175.7000	193.500	194.7000
CDKN1A	15.990	20.810	53.710	46.1200	45.990	55.9600
MAD1L1	11.980	13.080	14.210	12.1300	10.920	11.7600
CCNE2	6.862	7.282	6.255	6.9070	4.911	4.4270
E2F1	39.310	44.130	40.060	38.9500	28.100	30.4900
TGFB3	2.144	1.361	1.295	0.8241	1.075	0.6942

📊

📈

✎

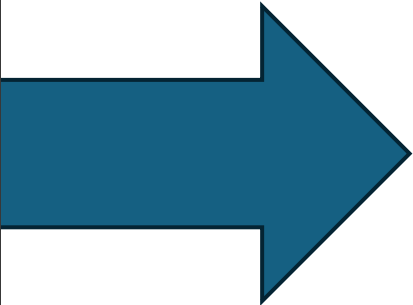
“/content/<ファイル名>”で指定することができます。

## 2. ChatGPTを活用！Pythonでデータ解析

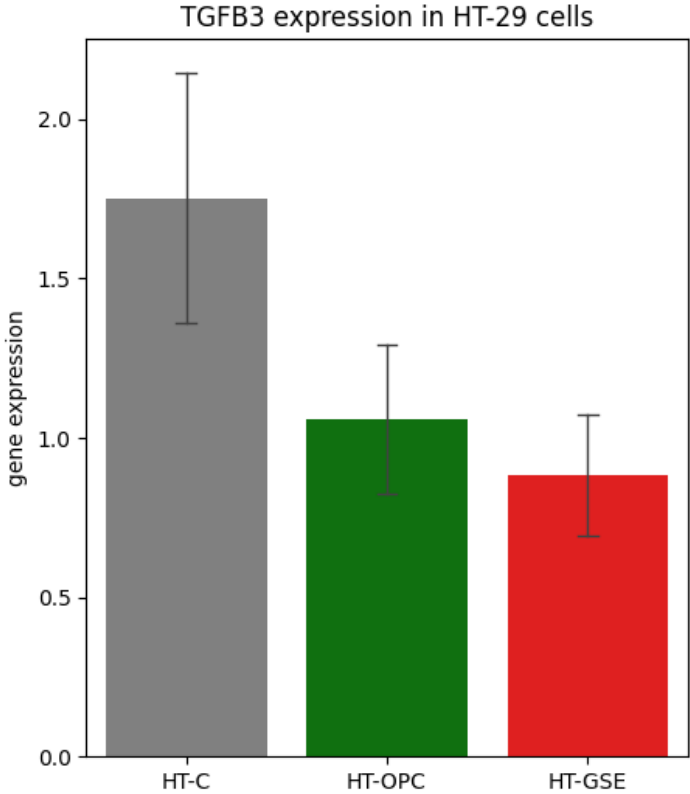
### 1. bulk RNA-Seq（チュートリアル）

スタート  
(表データ)

	HT-C1	HT-C2	HT-OPC1	HT-OPC2	HT-GSE1	HT-GSE2
SFN	109.900	136.300	186.600	175.7000	193.500	194.7000
CDKN1A	15.990	20.810	53.710	46.1200	45.990	55.9600
MAD1L1	11.980	13.080	14.210	12.1300	10.920	11.7600
CCNE2	6.862	7.282	6.255	6.9070	4.911	4.4270
E2F1	39.310	44.130	40.060	38.9500	28.100	30.4900
TGFB3	2.144	1.361	1.295	0.8241	1.075	0.6942



最終ゴール  
(棒グラフ)



## 2. ChatGPTを活用！Pythonでデータ解析

### 1. bulk RNA-Seq（チュートリアル）

#### ① 「Group行」を追加

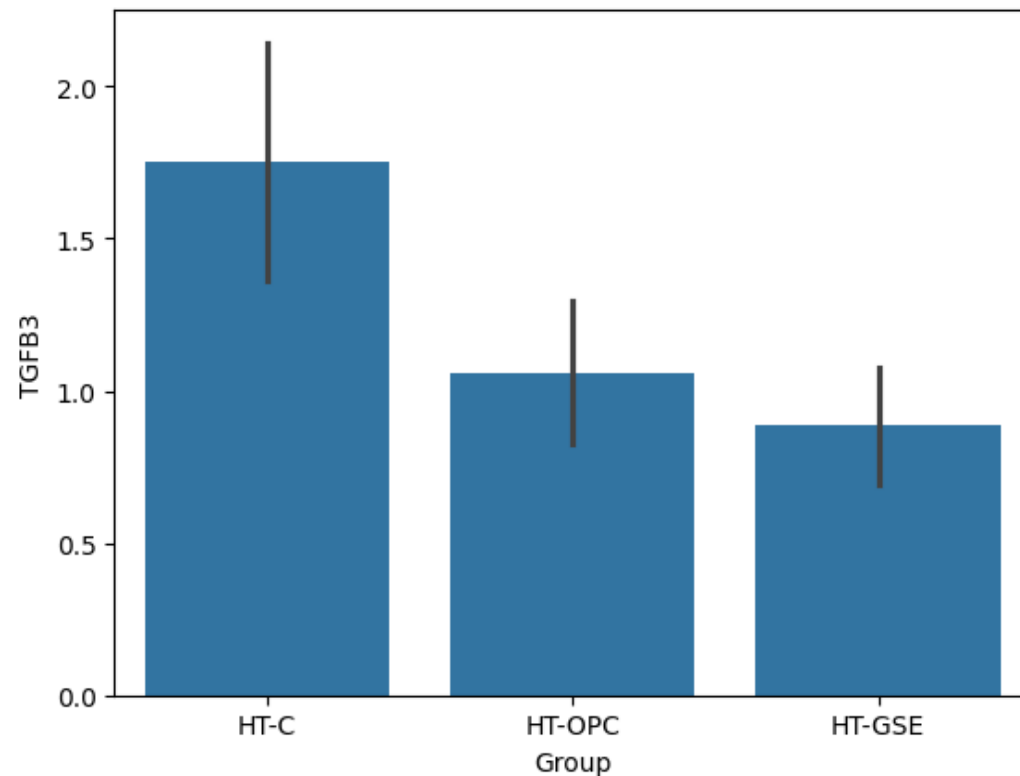
	SFN	CDKN1A	MAD1L1	CCNE2	E2F1	TGFB3	Group
HT-C1	109.9	15.99	11.98	6.862	39.31	2.1440	HT-C
HT-C2	136.3	20.81	13.08	7.282	44.13	1.3610	HT-C
HT-OPC1	186.6	53.71	14.21	6.255	40.06	1.2950	HT-OPC
HT-OPC2	175.7	46.12	12.13	6.907	38.95	0.8241	HT-OPC
HT-GSE1	193.5	45.99	10.92	4.911	28.10	1.0750	HT-GSE
HT-GSE2	194.7	55.96	11.76	4.427	30.49	0.6942	HT-GSE

入力例：「Pythonで、下のようなデータがあります。csvから読み込み、`df` という関数に格納されています。  
Groupという行に、HT-C、HT-C、HT-OPC、HT-OPC、HT-GSE、 HT-GSE」を入れたいです。」

## 2. ChatGPTを活用！Pythonでデータ解析

### 1. bulk RNA-Seq（チュートリアル）

#### ② TGFB3を棒グラフに



入力例：「先程のデータを棒グラフで可視化したいです。TGFB3行 y 軸に、Group行 x 軸に使用してください。」

## 2. ChatGPTを活用！Pythonでデータ解析

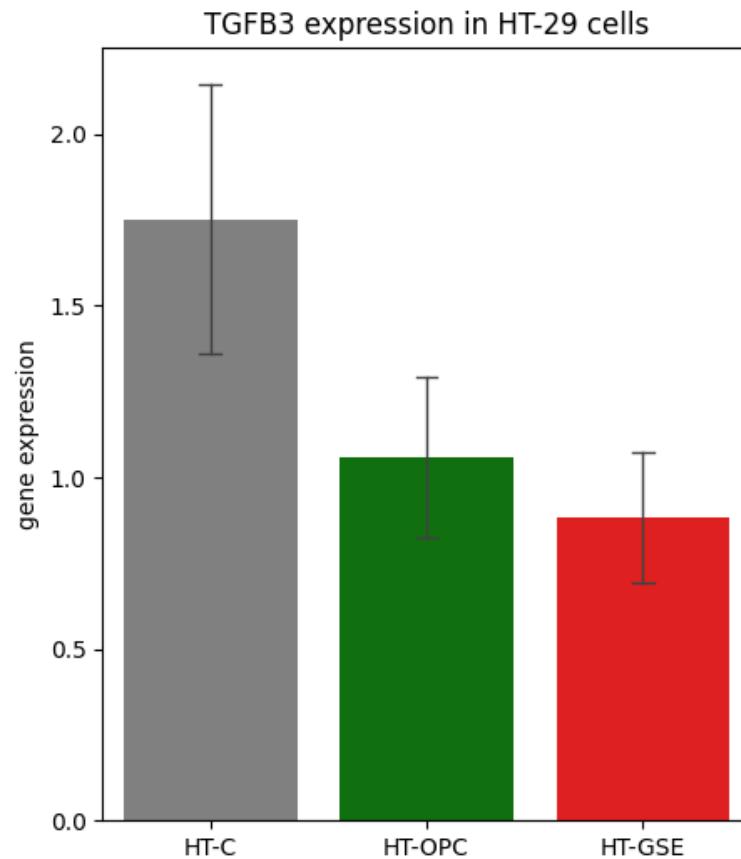
### 1. bulk RNA-Seq（チュートリアル）

#### ③論文中的のようにカスタマイズ

入力例：「先程の棒グラフを次のようにカスタマイズしてください。

- ・ 描写範囲を縦長に
- ・ 色を、HT-Cは灰色、HT-OPCが緑、HT-GSEが赤
- ・ エラーバーを追加
- ・ グラフのタイトルを追加
- ・ y軸のラベルを変更

」



# 2. ChatGPTを活用！Pythonでデータ解析

## 2. Microbiome解析

d__Bacteria;p__Campylobacterota;c__Campylobacteri;o__Campylobacterales;f__Campylobacteraceae	
SampleID	d__Bacteria;p__
CTRL_feces_028	0
CTRL_feces_099	0
CTRL_feces_100	0
CTRL_feces_115	0
CTRL_feces_118	0
CTRL_feces_144	0
CTRL_feces_161	0
CTRL_feces_184	0
CTRL_feces_207	0
CTRL_feces_231	0
PDAC_feces_001	0
PDAC_feces_003	0
PDAC_feces_005	0
PDAC_feces_010	
PDAC_feces_015	
PDAC_feces_020	
PDAC_feces_021	
PDAC_feces_023	
PDAC_feces_028	8
PDAC_feces_044	229

- ・ 膵臓がん患者 x10人
- ・ コントロール x10人

の腸内細菌(科レベル 81群)の観測リード数

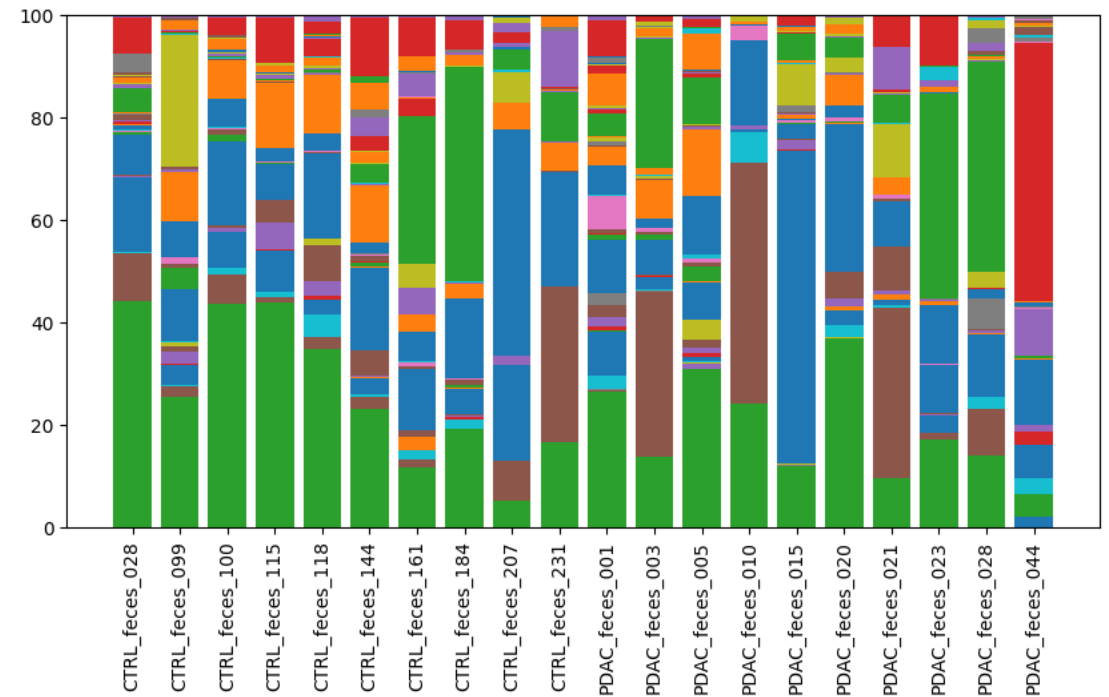
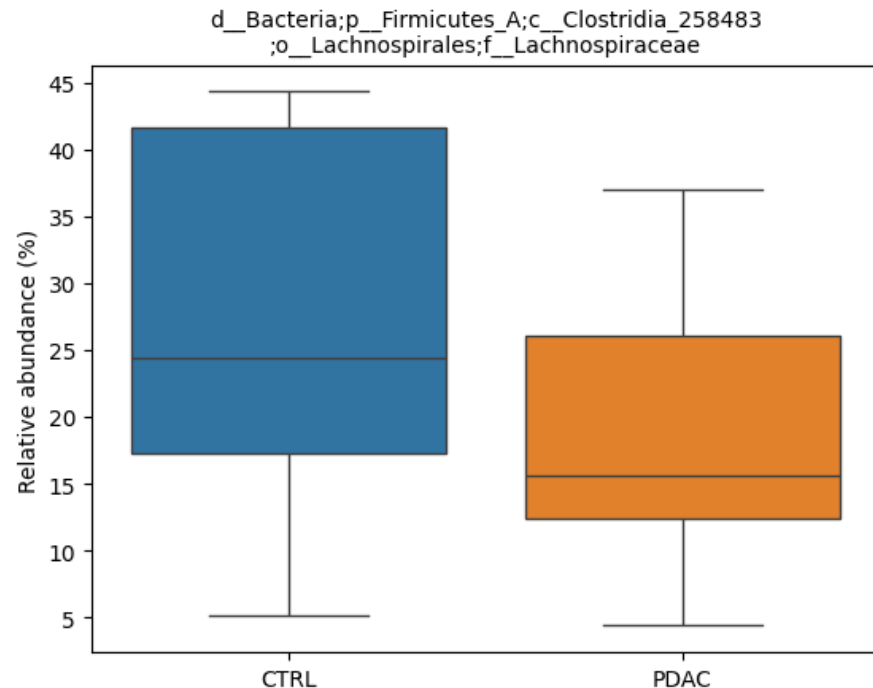
20 rows x 81 columns



## 2. ChatGPTを活用！Pythonでデータ解析

### 2. Microbiome解析

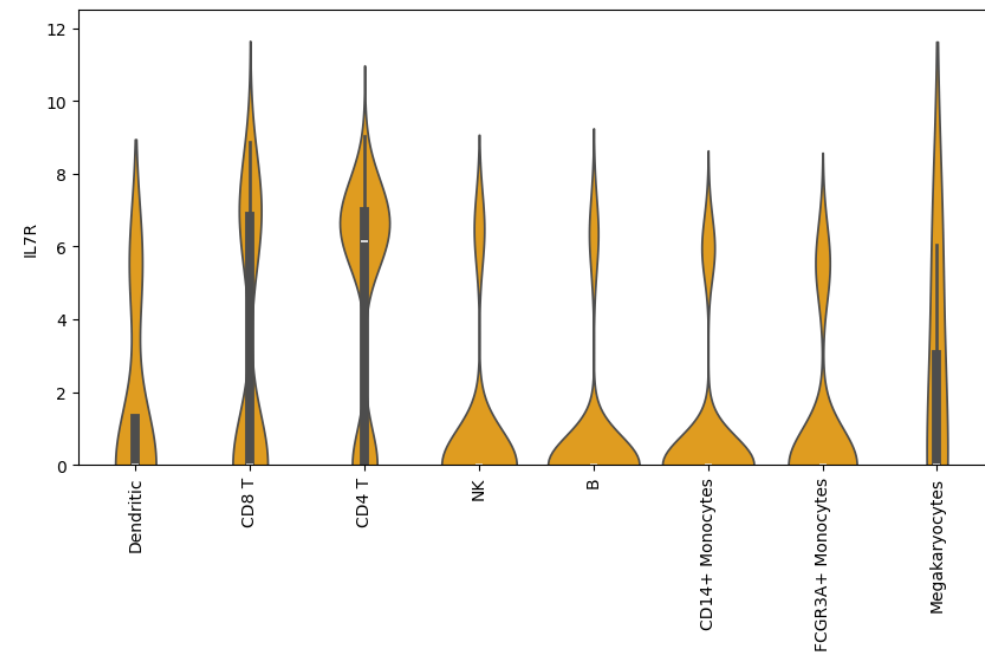
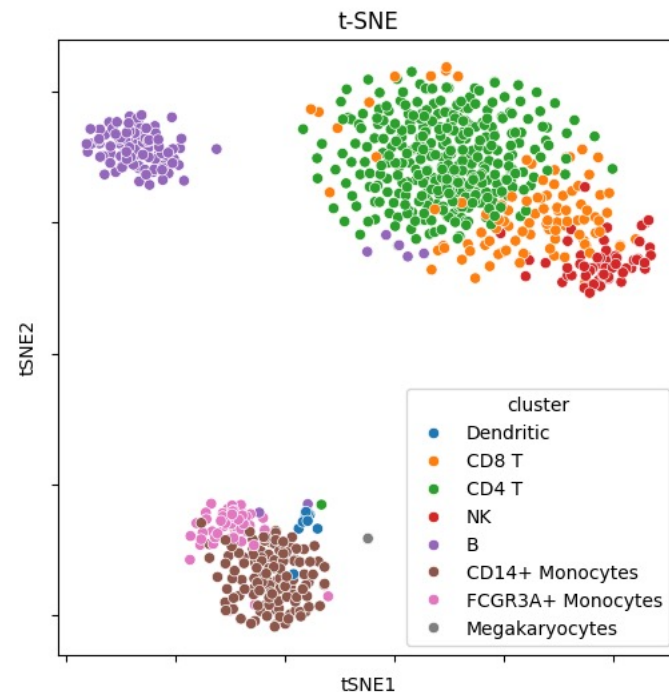
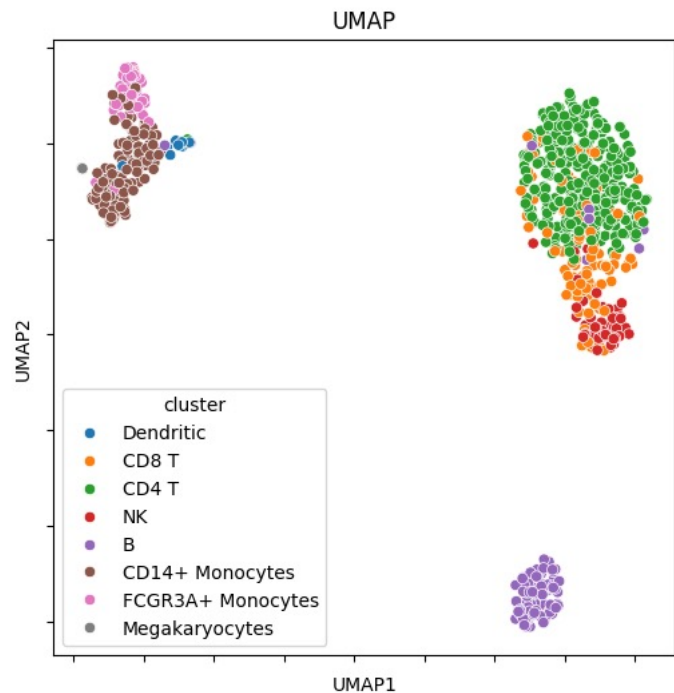
- ① データの読み込み
- ② サンプルごとに百分率化（合計値を合わせる）
- ③ `Group` 行の追加
- ④ `d\_\_Bacteria;p\_\_Firmicutes\_D;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Lactobacillaceae` をグループごとに箱ひげ図で可視化
- ⑤ 積み上げ棒グラフで、サンプルごとの組成を可視化



## 2. ChatGPTを活用！Pythonでデータ解析

### 3. Single cell RNA-Seq解析

- ① データの読み込み x2
- ② Cellごとに発現量を全マッピング数で補正（CPM補正）
- ③  $\log_{1p}$  で対数変換
- ④, ⑤ UMAPで次元削除 & 散布図で可視化
- ⑥, ⑦ t-SNEで次元削除 & 散布図で可視化
- ⑧ 細胞種ごとに`IL7R`の発現量を`violinplot`で可視化



### 3. AlphaFoldで遊ぶ！タンパク質の構造予測