

07 ゲノムデータ解析

尾崎 遼（バイオインフォマティクス研究室） <https://sites.google.com/view/ozakilab-jp>

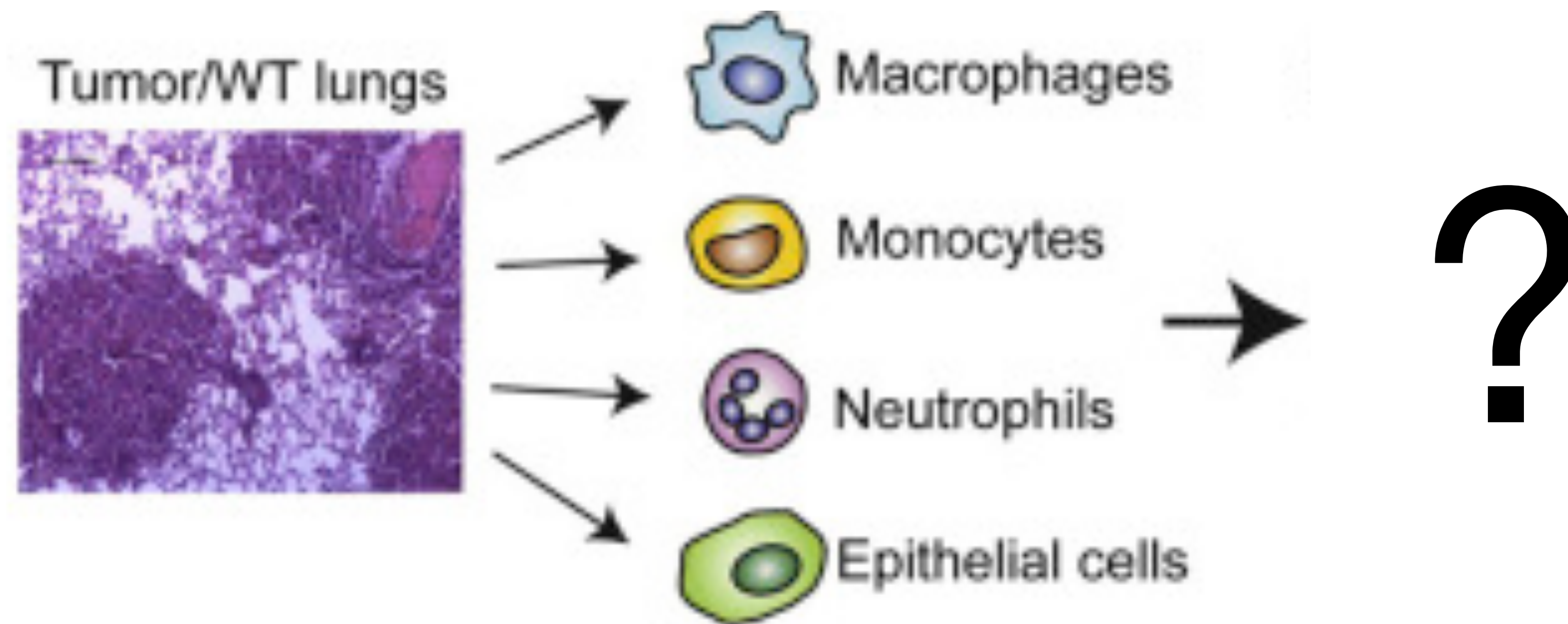
この講義の内容

- がんのRNA-seqデータを統合して、間質細胞から腫瘍細胞への細胞間相互作用についての仮説を導出する
-

がんのRNA-seqデータの統合

非小細胞肺癌 (non-small-cell lung cancer; NSCLC)

- 問い：NSCLCにおいて、腫瘍組織内に存在する間質細胞から腫瘍細胞にシグナルのクロストークが起こって腫瘍が活性化されるか？

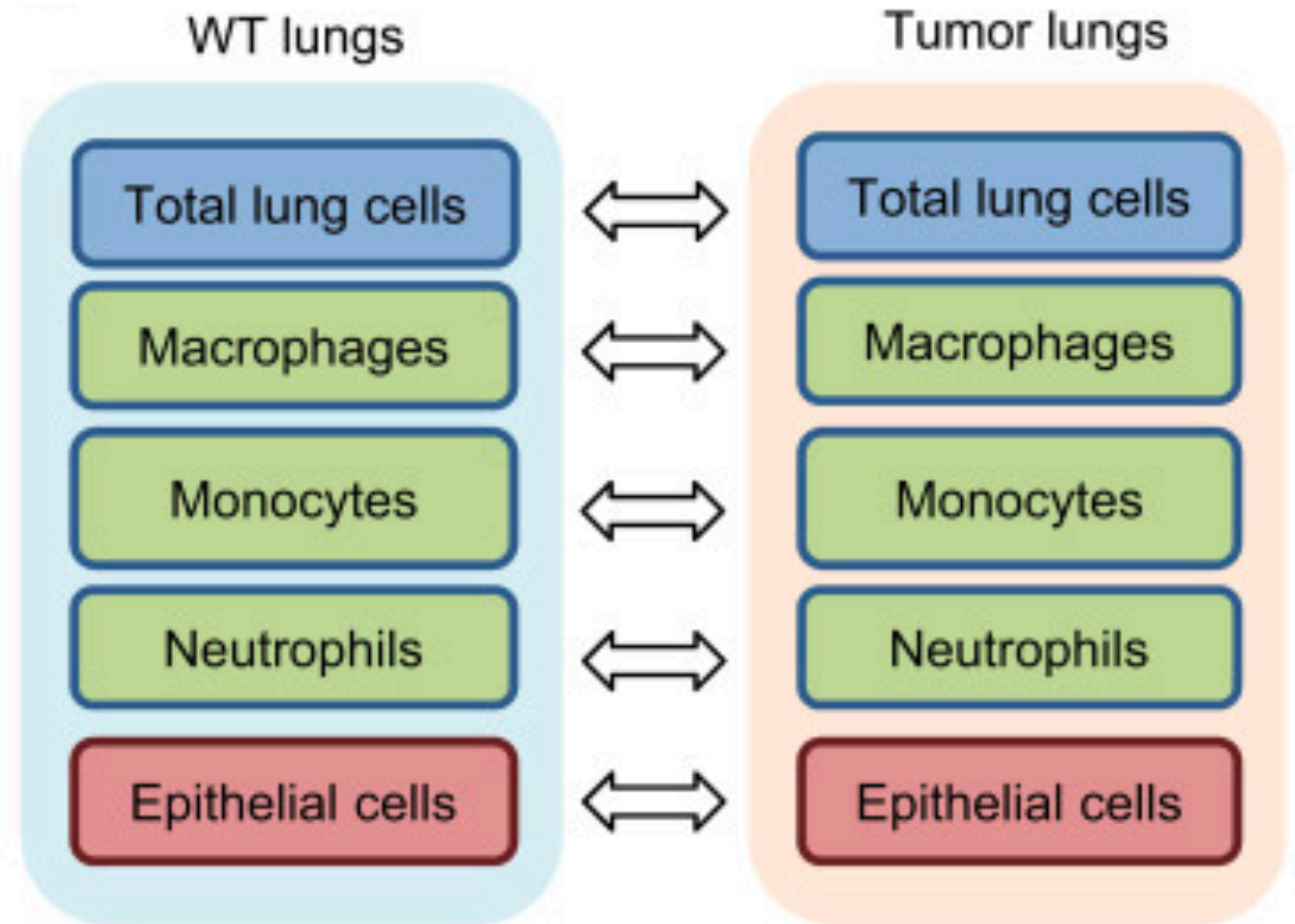


データの出どころ

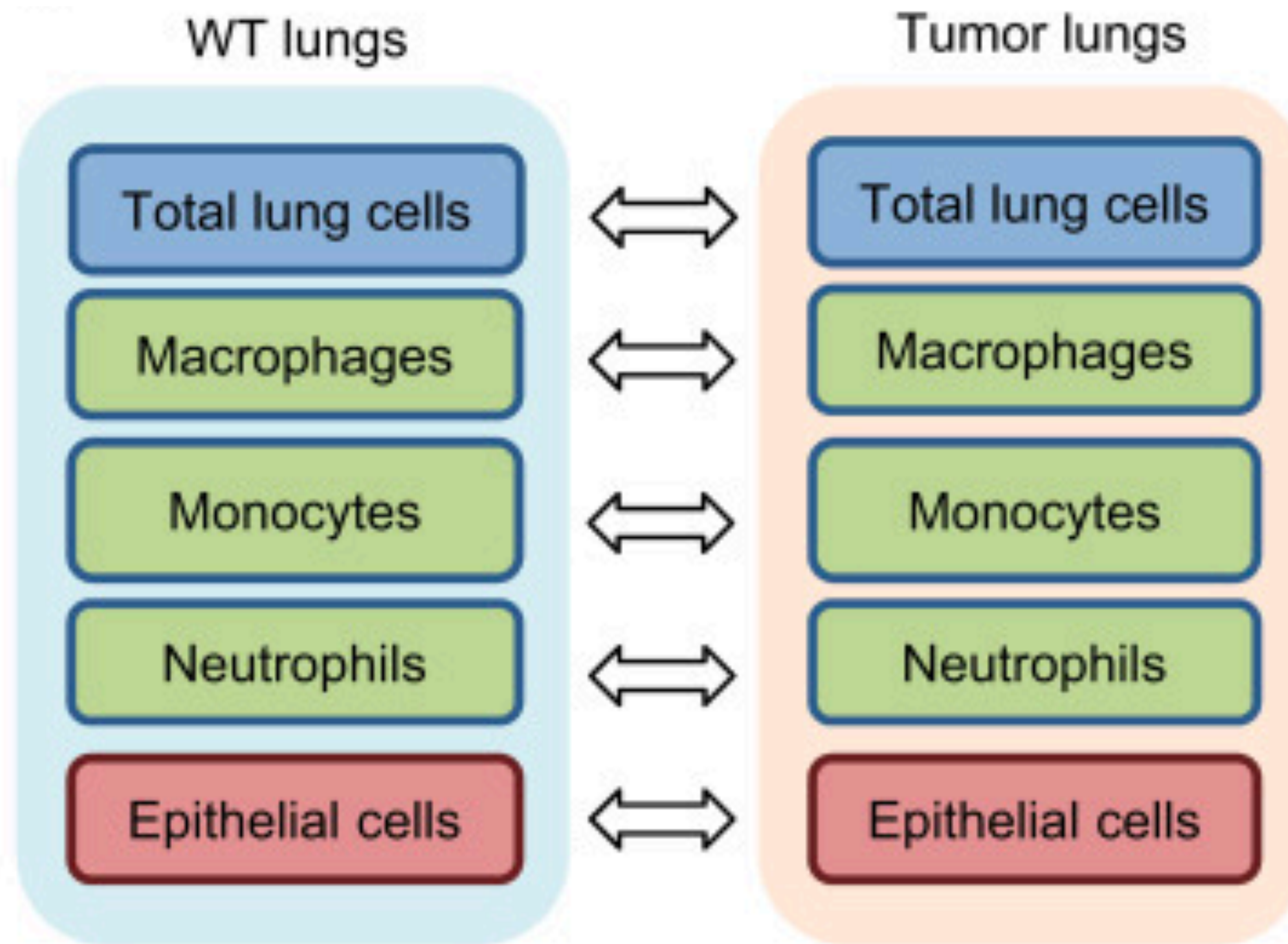
- Toi et al., Transcriptome Analysis of Individual Stromal Cell Populations Identifies Stroma-Tumor Crosstalk in Mouse Lung Cancer Model, Cell Reports (2015)
- <https://doi.org/10.1016/j.celrep.2015.01.040>

非小細胞肺癌 (non-small-cell lung cancer; NSCLC)

- 計測: RNA-seqデータ
- サンプル:
 - マクロファージ、単球細胞、好中球、上皮細胞 (NSCLCモデルマウスと野生型のマウスの肺からセルソーターで分けた)
 - Total lung cells
- 種: マウス



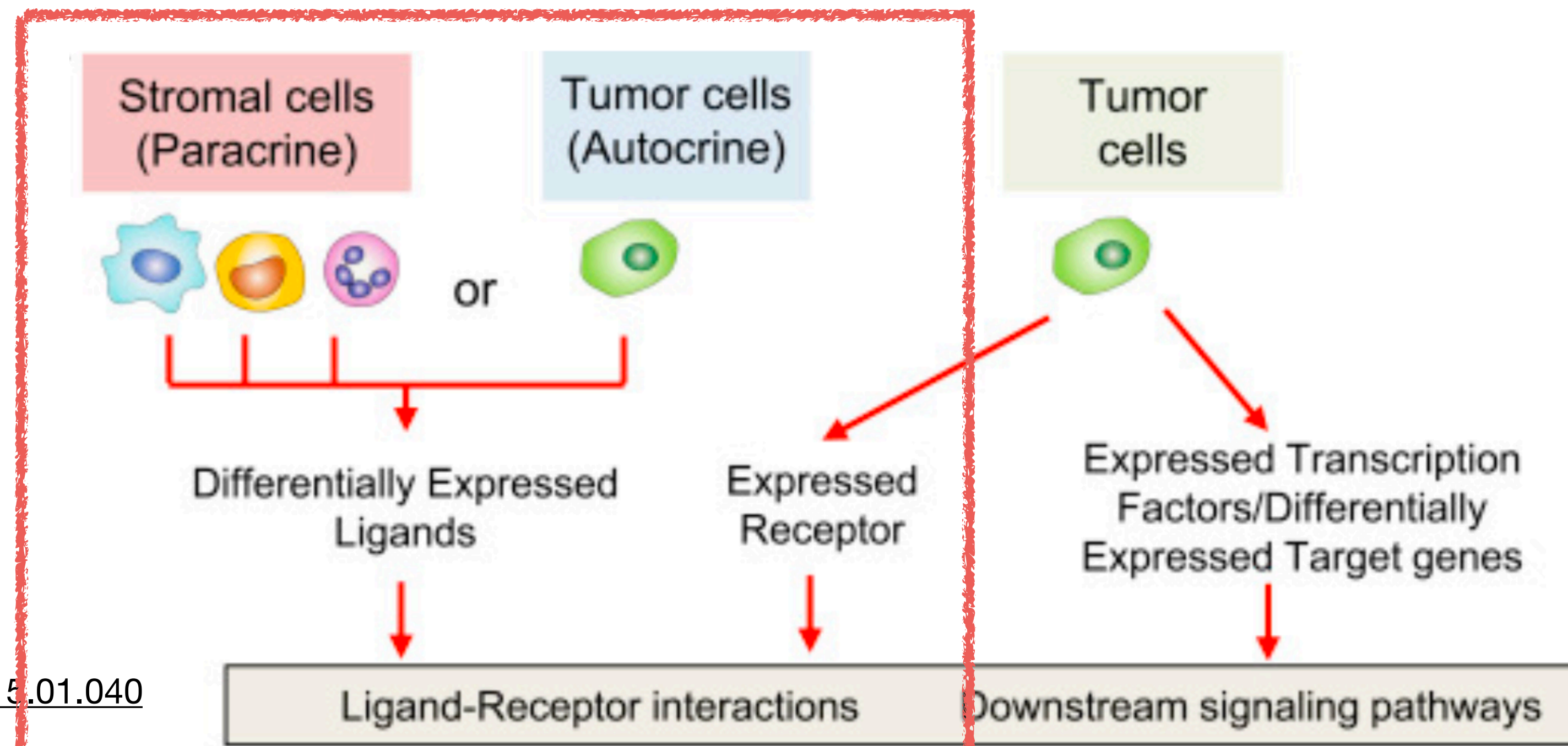
非小細胞肺癌 (non-small-cell lung cancer; NSCLC)



Cells	Surface markers
Total lung cells	-
Macrophages (Tumor)	CD11c+ CD11b+
Macrophages (WT)	CD11c+ CD11b-
Neutrophils	CD11b+ Ly6G+
Monocytes	CD11b+ Ly6C+
Epithelial cells	Epcam+ CD11c-

方針

- 1. 間質細胞において「正常組織由来のサンプル」に比べて「腫瘍組織由来のサンプル」で発現量が増加している遺伝子群を抽出する
- 2. 腫瘍組織において正常組織において発現量が増加している遺伝子群を抽出する
- 3. 「1.に含まれるリガンド」と「2.に含まれる受容体（レセプター）」をリガンド・受容体のペアのデータベースと比較する

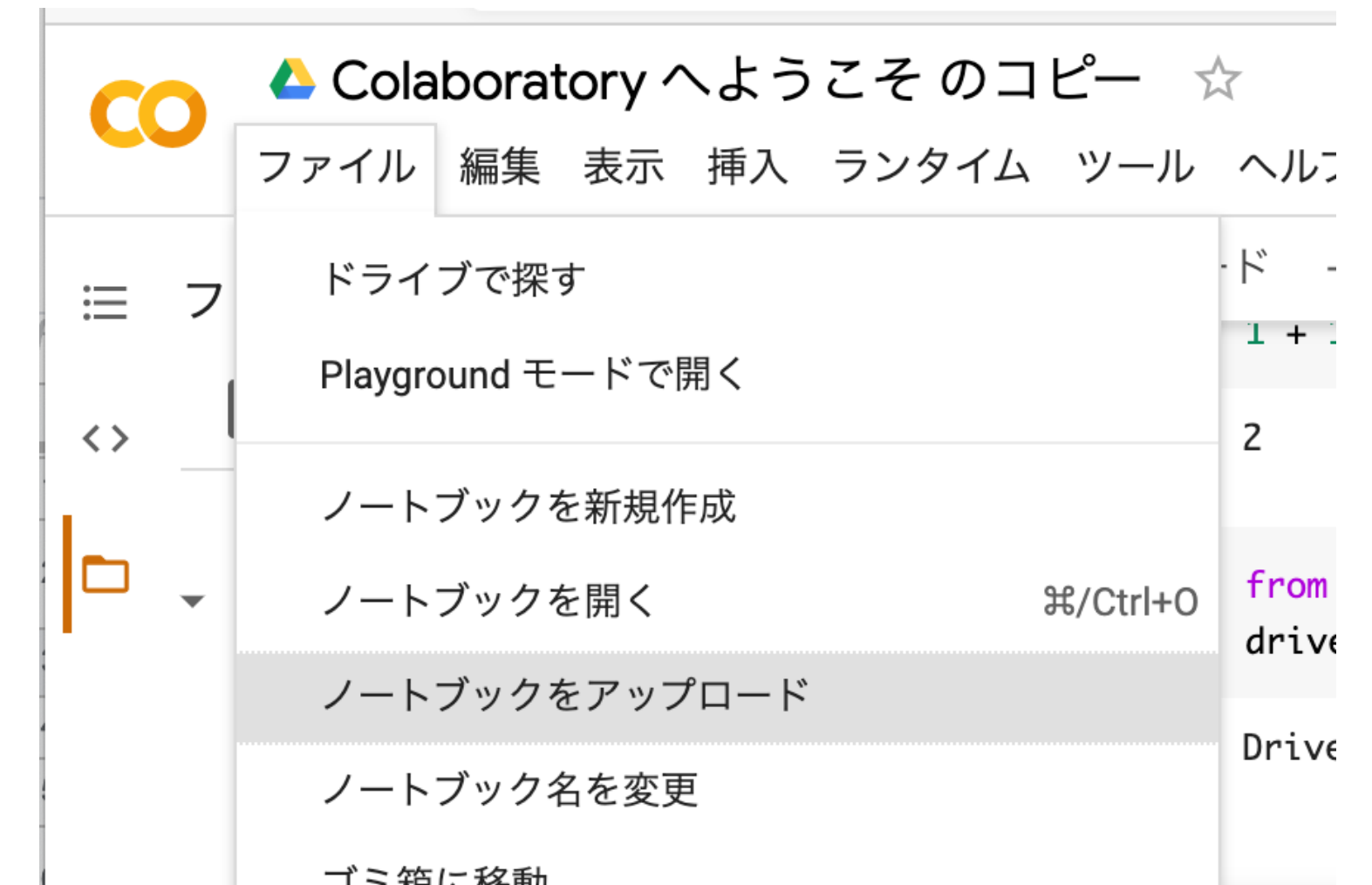


準備 (1/3)

- 07 のフォルダ <https://github.com/bioinfo-tsukuba/AdvancedCourse2020/tree/master/07>
- ファイルをダウンロードして Google Drive に置く
-

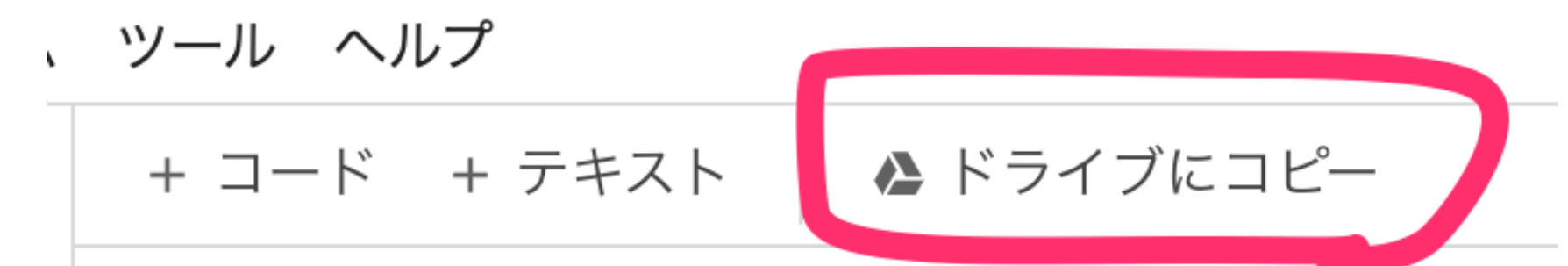
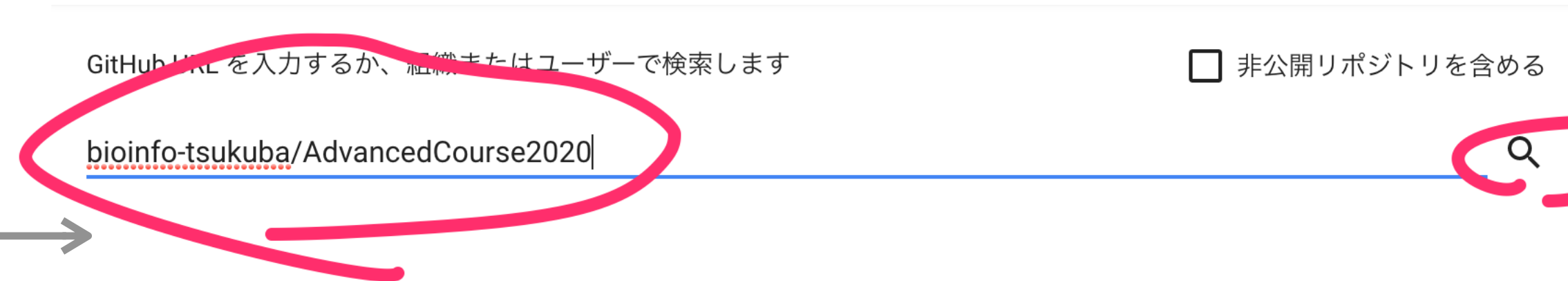
準備 (2/3) [Colab] GitHub にあるノートブックをコピーする (1/2)

- 「ファイル > ノートブックをアップロード」を選択します
- 出てきた画面にある「GitHub」タブをクリックします



準備 (3/3) [Colab] GitHub にあるノートブックをコピーする (2/2)

- 枠に入力し、🔍をクリック
 - bioinfo-tsukuba/AdvancedCourse2020
- **notebook_001_tumor_communication.ipynb** をクリックします
- 「ドライブにコピー」をクリックします
 - これでGitHubにあるノートブックが自分のGoogle Driveにコピーされました。



CSV と TSV について

- CSV

- comma separated values
- 区切り文字が コンマ
- `pd.read_csv()`

- TSV

- tab separated values
- 区切り文字が タブ
- `pd.read_csv()` だが、 `sep='\t'` というオプションが必要

使うファイルについて

- マクロファージのRNA-seqから遺伝子発現量に変動しているかを調べたデータ
 - 07/data/CCCExplorer/CD11CB_output/gene_exp.diff
- 上皮細胞（腫瘍細胞）での遺伝子発現量の表データ
 - 07/data/CCCExplorer/EP_output/genes.read_group_tracking
- Mouse Genome Informatics database から取得されたヒト-マウスオーソログリスト
 - data/CCCExplorer/HOM_MouseHumanSequence.rpt
- リガンド-受容体の対応表のデータ（`From` がリガンドで、`To` が受容体）
 - 07/data/CCCExplorer/LR_manual_revised.txt
- 遺伝子発現量のデータ
 - 07/data/GSE59831_processed_data_FPKM.txt