

細胞レベルの機能・表現型と遺伝子発現を関連付ける「Cell IO」データベースの開発

https://cell-io.riken.jp/
β版公開中



尾崎遼^{1,2}、仮屋山博文¹、井尻 遥士^{1,2}、山田 涼太³、田原（新井）悠也^{1,2}、松澤 亮輔^{1,2}
¹ 理化学研究所、² 筑波大学、³ Science Aid 株式会社

01 背景と課題

シングルセル RNA-seq で遺伝子発現や細胞クラスターは得られるものの、細胞の機能や表現型の解釈はいまだ難しい。

レポジトリ・2次 DB

- ・遺伝子リスト等はみられる
- ・カタログ（詳しくは自分で解析）

データを報告した原著論文

- ・部分的な報告
- ・データと紐づかない

データ解析

- ・遺伝子発現がわかってもどんな細胞かわからない
- ・GO 解析でも解釈しづらい

細胞の機能・表現型に関する知識は論文中の自然言語に埋もれており検索・再利用がしづらい。

PubMed 等

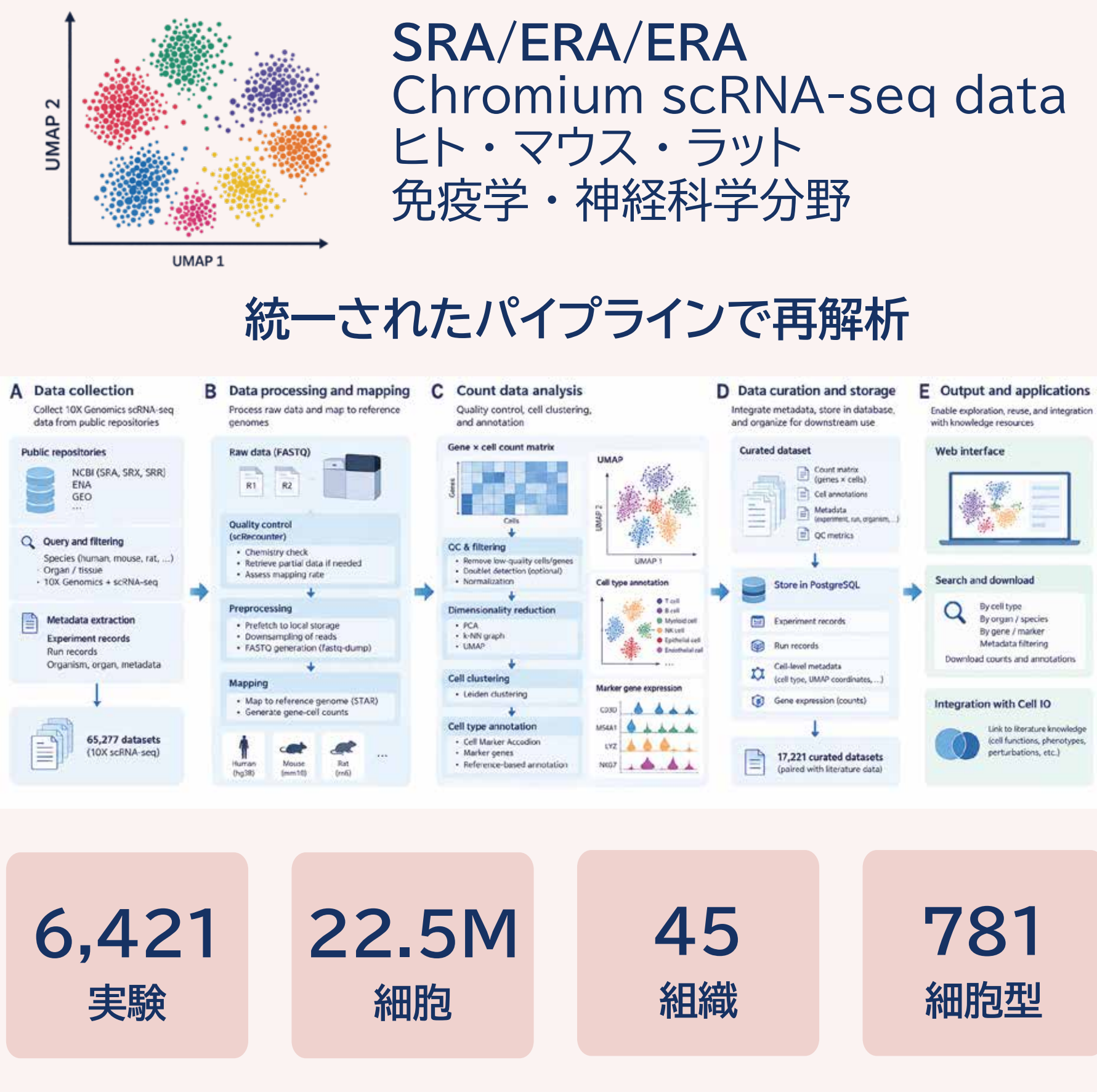
- ・細胞型名が曖昧で検索しづらい
- ・文脈依存で意味解釈が難しい
- ・読み込むにはドメイン知識が必要

LLM サービス

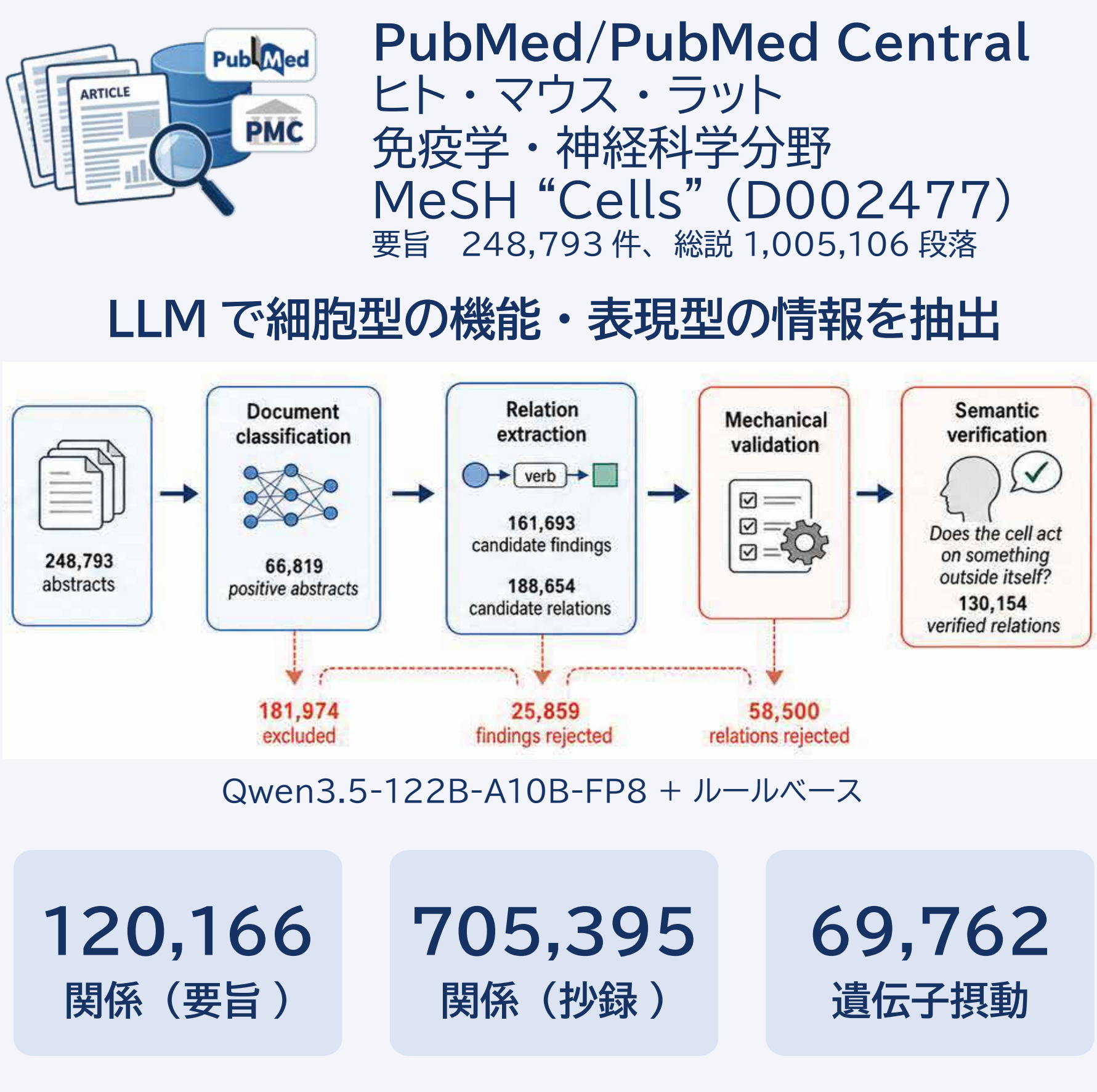
- ・ハルシネーション
- ・ドメイン知識がないと合っているかわからない

02 Cell IO データベース

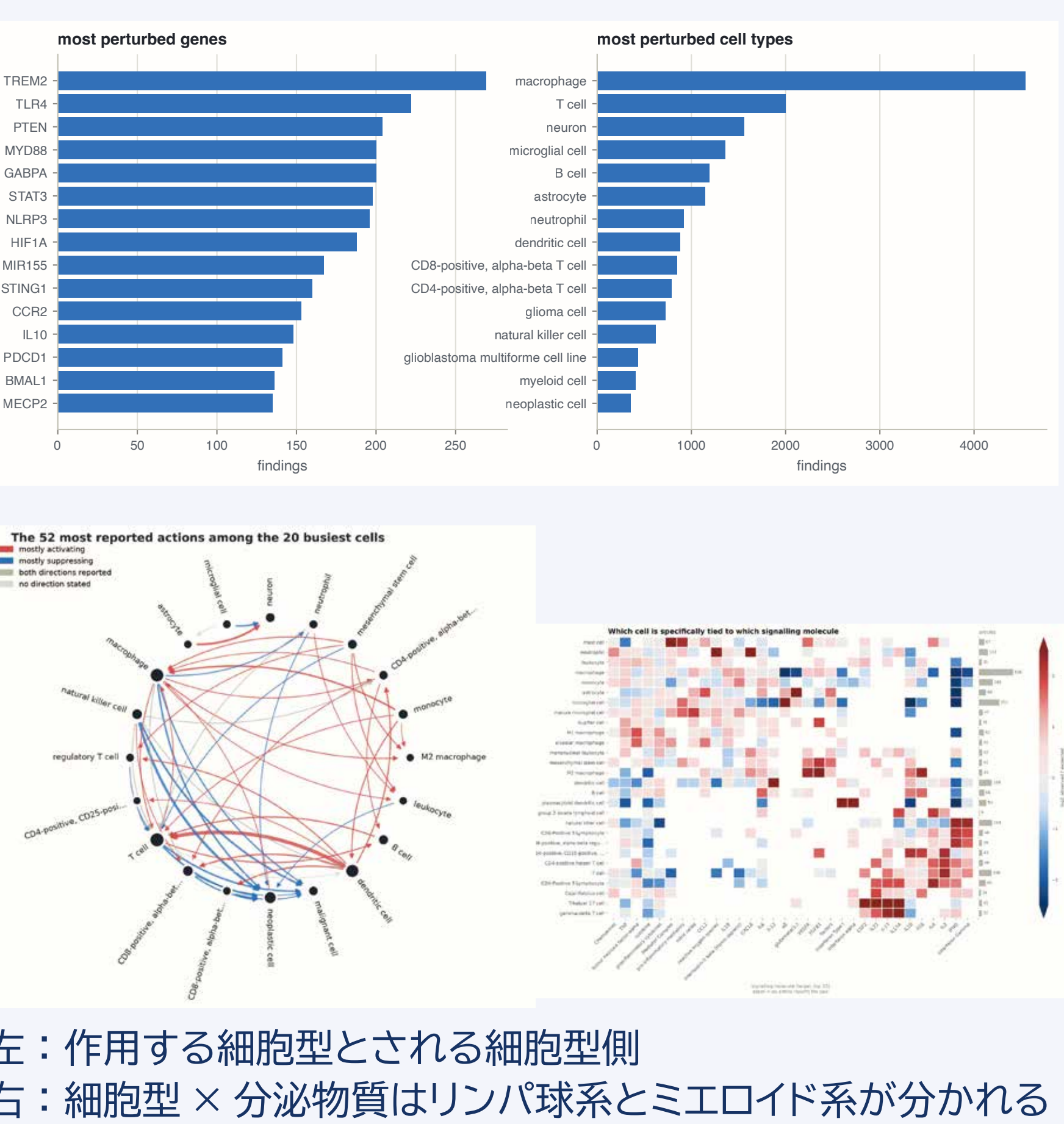
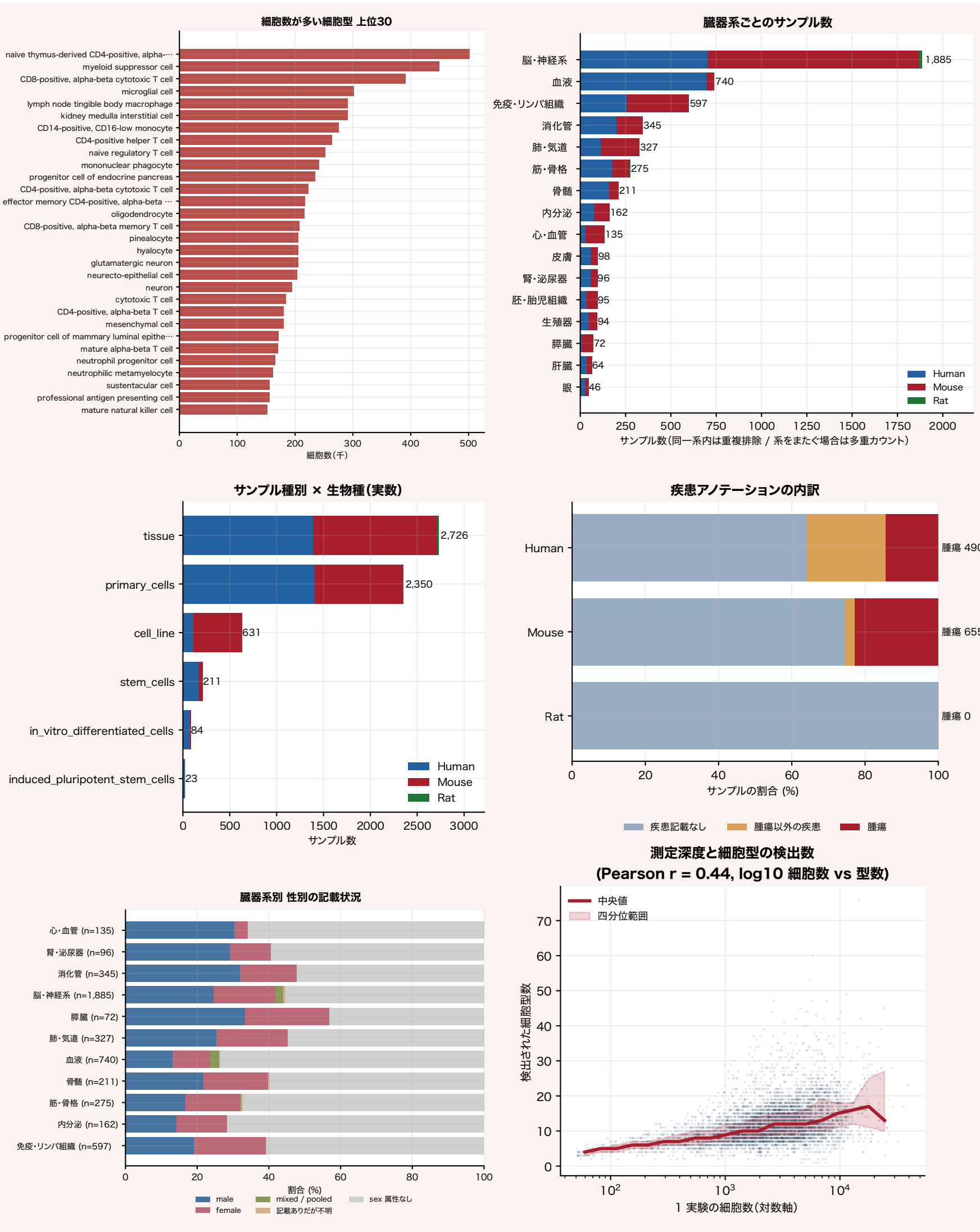
シングルセル RNA-seq データ



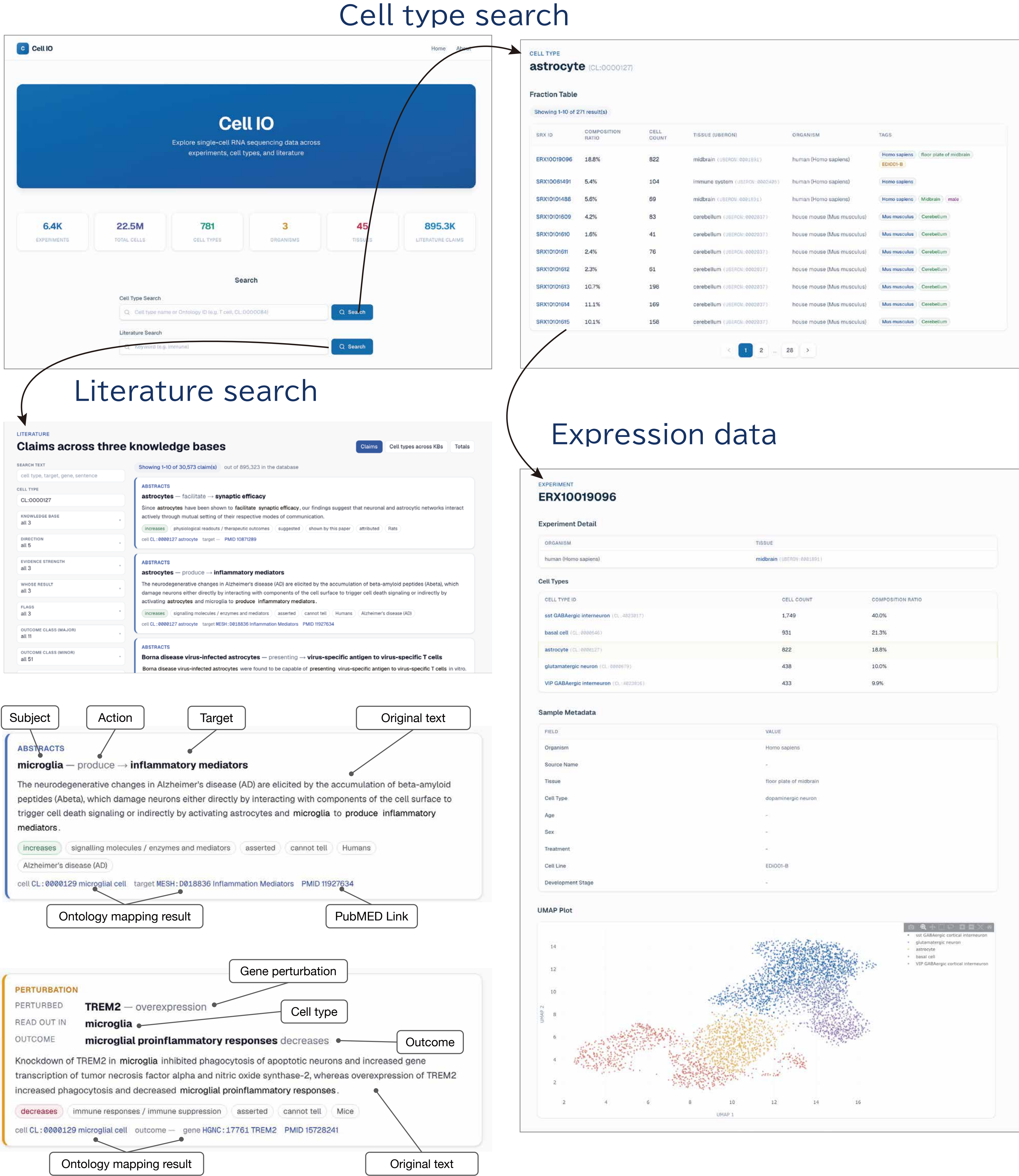
文献データ（非オミクス実験も含む）



04 Stats



03 細胞型名を介した遺伝子発現と文献情報の行き来



Acknowledgements This work was supported by JST NBDC JPMJND2402. Supercomputer systems of RIKEN, HOKUSAI SS and Rikyu, were used.