

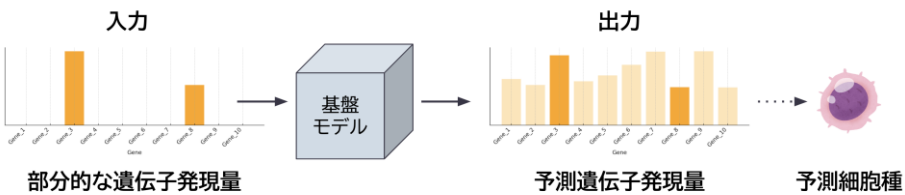
創薬を加速する遺伝子発現量の基盤モデル開発・事業成果概要

実施者

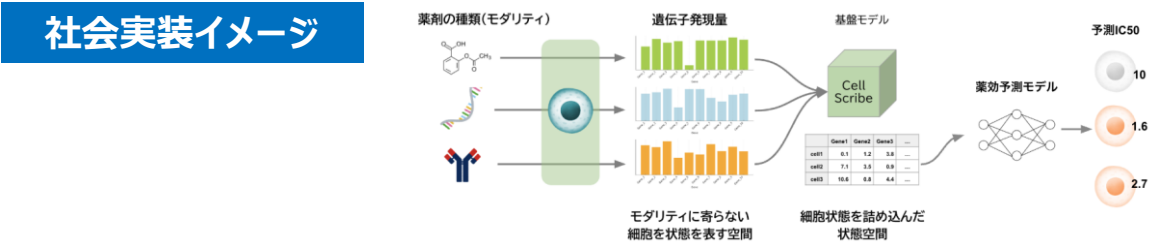
株式会社ヒューノーム研究所

事業概要

医薬品の薬効予測精度向上や創薬における臨床試験の効率化を目指し、多様な細胞状態を包括的に表現した地図となる基盤モデルを開発します。



創薬の効率化では医薬品の薬効評価や臨床試験結果の予測精度向上が重要です。本事業ではこれらの課題解決を目指し、公共データベースからシングルセル遺伝子発現量を大規模に収集し、細胞状態を包括的に表す基盤モデルCellScribeを開発します。



CellScribeは遺伝子実験結果の解析を効率化します。CellScribe上に構築された細胞状態と得られたデータの比較により、高精度かつ詳細な解析結果を提供します。製薬企業等では社内データによるファインチューニングで、さらなる薬効予測の精度向上を実現します。

事業成果

約9億細胞分のシングルセル遺伝子発現量データと、それに付随する器官、疾患、実験条件などのメタデータを収集しました。一定の品質基準を上回るデータを選別し、最終的に約3億細胞からなる高品質なデータセットを構築しました。このデータセットを用いて、非対称エンコーダ・デコーダ型のモデル構造を有する、3億パラメータの基盤モデル・CellScribeを学習しました。CellScribeは、遺伝子発現量基盤モデルの性能指標として広く用いられている平均二乗誤差（MSE）において2025年4月19日時点で0.295を達成しました。これは最先端モデルscFoundation（MSE 0.32）と比較し、予測誤差を約7.8%低減したことに相当します。

