

# 22PO-am372

## 自然言語処理によるノックアウトマウス表現型予測の検討

○藤原 大<sup>1</sup>, Attayeb MOHSEN<sup>1</sup>, 伊藤 眞理<sup>1</sup>, 鈴木 治<sup>1</sup>, 水口 賢司<sup>1</sup> (<sup>1</sup>医薬基盤健康研)

【背景と目的】病態モデルを作成する上で、ノックアウトマウスは有用な手段の一つである。現在我々は、疾患モデルマウスの資源バンクを作成している。そのため、ノックアウトマウスの作成効率を高めることはコスト面でも重要である。しかしながら、対象とする遺伝子によっては、胚性致死となることはよく知られている。これまでは既存のデータベースを参照しながら、事前調査を実施していたが、今回、自然言語処理などの人工知能を活用して、ノックアウトマウスの表現型予測を実施したので報告する。

【方法】マウスゲノムインフォマティクス (<http://www.informatics.jax.org/>) のウェブサイトからフェノタイプ情報を取得した(12938例)。“Lethal”や“Death”の記載のあるフェノタイプを陽性としてラベルづけを実施した。それぞれのフェノタイプのテキスト記載に対して自然言語処理を実施し、ベクトル化を行った後、線形モデルを用いて、フェノタイプの予測を実施し、5分割交差検証を実施した。

【結果】陽性 50 例陰性 50 例を用いたデータセットにおいて、本モデルでは、曲線下面積 (Area Under Curve: AUC) が 0.992 という高い予測性能を得られた。

【考察】今後、マウスの系統差も考慮し、より実用的なモデルを開発予定である。