

## 臨床シーケンスでの人工知能技術の応用 Utilizing Artificial Intelligence for Clinical Sequencing

井元 清哉  
Seiya Imoto

東京大学医科学研究所ヘルスインテリジェンスセンター  
Health Intelligence Center, Institute of Medical Science, University of Tokyo

### 1. 人工知能を臨床シーケンスに活用するための 医科学研究所の取り組み

東京大学医科学研究所では、2011年からヒトゲノム解析センターのスーパーコンピュータシステムを活用し、がんを対象とした全ゲノムシーケンスに基づく臨床シーケンス体制を構築してきた。国際がんゲノムコンソーシアムなど数々のがんゲノム研究で実績のある全ゲノムシーケンスデータや RNA シーケンスデータの解析パイプラインをはじめ、ゲノム解析のさまざまなプログラムを整備し、医科研ヒトゲノム解析センタースパコンの一部をインターネットから切り離れた上に実装し、高セキュリティなデータ解析環境を構築している。また、Laboratory Information Management System (Clarity LIMS) によるシーケンス解析・データ解析マネジメントシステム、生体認証によるセキュリティ管理、網羅的多地点カメラによる安全・データ事故管理などの一連のシステムを構築し運用している。

シーケンスデータ解析を行えば、患者さん個人に生じた数千から数万のゲノム変異を知ることが出来る。このシーケンス情報を治療方針の立案に活用する臨床シーケンスのボトルネックは、膨大な数発見される変異の臨床翻訳にある。現在行われている多くの臨床シーケンスは、数十から百程度の既知のがん関連遺伝子を標的としたパネルシーケンスに基づく。しかしながら、変異の臨床翻訳のための基本情報である科学論文では、いまや全ゲノム解析が当たり前、全ゲノム解析でなければ論文が通りにくいという声も聞く。そのような科学論文の情報を蓄積している生物科学分野文献データベース PubMed には、2000 万件を超える論文が登録され、がん・ゲノム分野だけでも年に 20 万件を超える勢いで増えている。パネルシーケンスの枠を超え、この全ゲノム情報を臨床シーケンスに活用するた

めの研究が必須である。そこで、医科研では、IBM の Watson Genomic Analytics Early Adaptor Program に採択され、2015 年 7 月 1 日に IBM Watson Genomic Analytics (現在は、Watson for Genomics と名称変更) が医科研に研究用として導入され、臨床シーケンスにおける活用について研究を行っている。

2015 年度からは、これまでの大腸がんのマルチリージョナル全ゲノム解析だけでなく、Myeloid パネルを使った血液腫瘍の臨床シーケンスも始まった。時系列で採取されたゲノム情報の重要性、パネルシーケンスの限界、マルチリージョンでゲノム、RNA の統合的な解析の必要性、1日以下でシーケンス解析を行い、変異を解釈できるスピードが多くの患者さんの命を救う可能性などさまざまな成果を得てきた。現在も隔週でシーケンス解析や Watson の解釈について評価する検討会を開催し、臨床シーケンスの体制整備、シーケンス解析環境の整備、同定したゲノム変異の臨床翻訳についての研究、および運用のための経費のことなど、医師(含 臨床遺伝専門医)、生物学者、情報科学者、遺伝学者、生命倫理研究者が一堂に会して議論を行いさまざまなノウハウがこの5年ほどの間に蓄積している。

本講演においては、東京大学医科学研究所にて人工知能を臨床シーケンスに活用するために行ったこれまでの研究の成果について紹介する。

連絡先: 井元清哉, 東京大学医科学研究所ヘルスインテリジェンスセンター健康医療データサイエンス分野, 〒108-8639 東京都港区白金台 4-6-1, Tel 03-5449-5611, Fax 03-5449-5442, E-mail: imoto@ims.u-tokyo.ac.jp