

第15回ゲノム・オミックス連携推進セミナー

「教師なし機械学習によるゲノムビッグデータからの 想定外の知識発見と社会的な重要課題への応用」

池村 淑道 先生

長浜バイオ大学 客員・名誉教授

国立遺伝学研究所・総合研究大学院大学 名誉教授

【日時】 10月4日（木） 15：30～17：00

【場所】 東北メディカル・メガバンク棟 3階大会議室

【セミナー要旨】

ゲノム配列を代表例とする、生命科学分野の多様なデータがビッグデータ化しており、どの様な興味深い新知識が潜んでいるのかを想像することが困難となってきた。この状況下では、モデルや仮説や予備知識なしにビッグデータを研究できる「教師なし機械学習」が威力を発揮する。ゲノム配列の連続塩基（例えば5連や6連、あるいは20連続塩基）の出現頻度に着目した機械学習を行うことで、想定外の多様な知識発見を可能してきた(1,2)。ヒトゲノムの核内配置やクロマチン相互作用に關与するシグナルに關する解析例(4)や、「エボラやインフルエンザウイルスの弱みを探る」(3,5,6)等の社会的な重要課題への応用例も紹介する。

(1) Informatics for unveiling hidden genome signatures. *Genome Res*, 13, 693-702, 2003. (2) Novel Phylogenetic Studies of Genomic Sequence Fragments Derived from Uncultured Microbe Mixtures in Environmental and Clinical Samples. *DNA Res*, 12, 281-290, 2005. (3) Prediction of directional changes of influenza A virus genome sequences with emphasis on pandemic H1N1/09 as a model case. *DNA Res*, 18, 125-136, 2011. (4) Notable clustering of transcription-factor-binding motifs in human pericentric regions and its biological significance. *Chromosome Res*, 21, 461-474, 2013. (5) Directional and reoccurring sequence change in zoonotic RNA virus genomes visualized by time-series word count. *Sci Rep*. 2016 Nov 3;6:36197. (6) Time-series oligonucleotide count to assign antiviral siRNAs with long utility fit in the big data era. *Gene Ther*, 24, 668-673, 2017.

連絡先 : [seminar\(at\)gpc.megabank.tohoku.ac.jp](mailto:seminar(at)gpc.megabank.tohoku.ac.jp) ←(at)を@に変えて下さい
(東北メディカル・メガバンク機構 ゲノムプラットフォーム連携センター)