



個人研究

健康社会

ゲノムデータの情報解析



ゲノム塩基配列データが大量に低コストで得られるようになり、様々な生物の多くのゲノムデータが収集されています。その大量のゲノムデータを疾患原因の解明などに活かすために、ゲノムデータ解析手法の研究、ゲノム解析ソフトウェアの開発、統計解析手法の研究をしています。

KEYWORDS バイオインフォマティクス、ゲノム解析、NGS データ解析

RESEARCHER

応用生物学部 准教授 土井晃一郎

<https://doi-lab.bs.teu.ac.jp/>

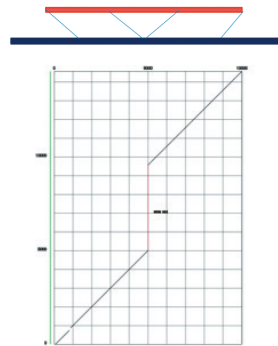


学会発表・論文・著書・社会活動

- [1] Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease, *Nature Genetics*, Vol.51, No.8 (2019) 1222-1232.
- [2] Expansions of intronic TTCA and TTTA repeats in benign adult familial myoclonic epilepsy, *Nature Genetics*, Vol.50, No.4 (2018) 581-590.
- [3] Centromere evolution and CpG methylation during vertebrate speciation, *Nature communications* Vol.8, No.1 (2017)

01 | ゲノム構造多型

個人のゲノムを解析するときには個人のゲノム間の変異を抽出することが必要です。挿入、削除、逆位などの構造多型の抽出の研究をしています。ゲノムの変異の中で一塩基多型 (SNP) と呼ばれる 1 塩基の違いはある程度正確に抽出できるようになってきていますが、構造多型と呼ばれる大きな変異を NGS の短い塩基長のデータから正確に抽出するのは困難です。NGS データから構造多型を抽出する手法、PacBio、Nanopore の長いリードの活用方法、抽出した構造多型の解析をしています。



02 | ゲノム解析のためのソフトウェア

ゲノム解析を行うには、解析項目が多いために、既存のソフトウェアを組み合わせるとともに、新たな解析ソフトウェアを開発していくことが必要です。現在、一部の神経疾患の原因になることが知られているタンデムリピート伸長を NGS のリード中のタンデムリピートのヒストグラムを利用して発見するソフトウェアを開発しています。このソフトウェアを利用して、疾患に関わる新たなタンデムリピート伸長が発見されています。更に新たな解析ソフトウェアの開発も行っていきます。

ACTAACAGCAGCAGCAGCAGCAGCAGCAGCTATCA
 ↓
 ACTAACAGCAGCAGCAGCAGCAGCAGCAGCAGCTATCA
 CAG repeat

