

D 29 タンパク質が結合するDNAパターンをコンピュータで予測する

はたらくゲノム-タンパク質

展示責任者 丸山 修

展示責任者所属

九州大学大学院数理学研究院

タンパク質は、生物の生命維持活動において重要な役割を担っています。生成するタンパク質の選択や調整は、タンパク質の複合体(タンパク質が複数集まって形成された分子機械)がDNA配列の特定の領域に結合するというメカニズムに大きく依存しています。

このことを単純化して考えると、タンパク質複合体が選択的に特定の領域に結合することは、生物の多様性を作り出す「切り替えスイッチ」と捉えることが出来ます。

しかしながら、これら結合領域の特徴として、組み合わせの要素や曖昧さがあり、理解するにはなかなか手強い相手であり、また、現在、様々な生物種のゲノム配列が決定され、大量の配列がデータベースに蓄積されつつありますが、先ほどの結合領域を一つ一つ生物実験で決定することは可能ですが非常に手間がかかります。この現状への対応策として、コンピュータを用いて、結合領域のパターンを特定し結合領域の候補を絞り込む方法が考えられます。

会場では、コンピュータを用いた結合領域予測手法の意義と困難さについて紹介します。

会場では、コンピュータを用いた結合領域予測手法の意義と困難さについて紹介します。

DNA配列解析の例

Q左のグループにあって、右のグループにないものは何?

AC GTAGC TAGT	TGTC ATCGGAT
GCC TAGAGATC	GGCC TATAC GT
GATAGC TAAAA	AC CGCG TTACG
C GGATAG TTAC	TATACC TGGCT
C CC GGCC TAGA	AGC TAC C TATC
TTAGCC TAGC G	GGCG CGGTC TC
TTAAGC TAGC T	CC TACTC TACA
ATAGC GATC GT	GGC TATCGGAT
AG TAGGAC TGA	CT TATAC C TAT
TC AC TAGTC TG	AGGC GC AATC T

E 30 ゲノムを解読するには、どんな計算が必要なのでしょう?

ゲノム情報を解析する

展示責任者 森下真一 / 土井晃一郎 / 瀬々 潤

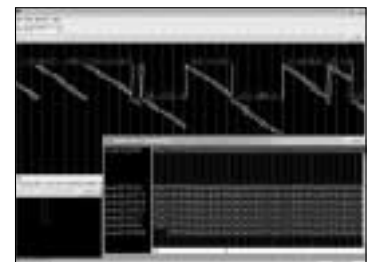
展示責任者所属

東京大学大学院情報生命科学専攻

ゲノムの解読や遺伝子の解析にはコンピュータソフトウェアが必要だといわれます。どのような場面で「計算」が必要なのでしょう? たとえばヒトゲノムは約30億個の文字から成り立っていますが、文字をすべて解読するには、ゲノムを細かい数千万個の断片にして、それから繋ぎ合わせてゆきました。実はこの繋ぎ合わせる作業は結構大変で、自動化するため、いろいろと計算方法を工夫します。たとえばグラフ構造という数学の手法が巧みに使われます。それから、長さが数千の遺伝子配列を、非常に長いゲノム配列の中を探すのも時間がかかる作業です。高速で見落としがない方法を考えようとする、確率論とアルゴリズムを組み合わせた設計

方法が必要になります。さらにゲノムから読み取られる遺伝子の振る舞いを一度に観測できる装置の出現で、数万個の遺伝子の様子も追跡できるようになりました。数万個パラメータ(遺伝子)を処理

できるデータマイニングという統計学の拡張が使われます。このようにゲノム解読には、さまざまな数学や計算の考え方が使われますが、基本的となる考え方をやさしく紹介しようと思います。



TCACCAT BBTGACTAARBRBCTCABCTBCTCTCTCCBCTTCACTGTCAGCTGT BCTTGGAGGABG BCTAGCTCCAGCAGCCGBCCTCAG
TCAGAAA AABCTATDT BTCAABTAAAGBBACABCBRTCAAGAT BCTCCCTTCAAAATCTGTGTGTCTACGTCTTAACTCTCAABGATAAGCBBC
CTCAAAA BBAAGTBCAGABGT BCAGACACBCTCCTT BCAGCCTCA BCTCCCTGGCTC BACBARTCTCTT BCTCABCTCCCAABT BCTTGGACCA
BAGTTC BTCCBACAT BCCCACT BATT TCGATT TTTTBTAG BAAAT BTGGTTCCACATATTGCTCA BBTGCTCTTAAAT BCTTBAACCAAGC
BCTTCCACCTTGGCTCTCAABT BCTCAGCATA BCTCCTG BCCBACTT BBAACCACTTAAAGAAAATAATATCTCCBCTTBAACCAAGC