バイオインフォマティクス
（第1回）

慶應義塾大学生命情報学科
榊原康文

講義の形式

榊原：前中半

- オンライン形式（オンデマンド形式の動画配信）と対面形式のハイブリッド
- 講義内容は動画配信
- 演習は対面
- 演習は，５月１０日（月）と6月７日（月）を予定（あくまで予定）
- それまでに講義を行った内容に関して，演習問題を解く
- 演習は，出席点も兼ねる

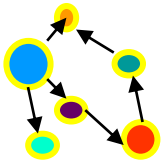
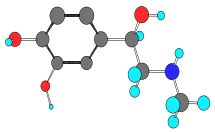
講義の予定

榊原：前中半

- ゲノム配列のアセンブリと決定法
- ペアワイズアライメントと動的計画法の仕組みと理論
- マルチプルアライメント（プロファイル，モチーフなど）
- データベース検索（BLASTなど）
- クラスタリング（マイクロアレイ解析への応用）
- 系統樹解析，比較ゲノム解析
- アドバンスな研究課題，医科学・生命科学ビッグデータ解析，など

佐藤健吾：後半

医学・生命科学ビッグデータ時代の到来

ゲノム配列解読 【Genome Online DB】 （125,504ゲノム）	遺伝子ネットワーク 【KEGG PATHWAY】 （39万エントリー）	遺伝子発現 【NCBI GEO】 （175万細胞サンプル）	化合物データベース 【PubChem】 （2億1895万化合物）
ゲノム配列 GATTTAAGATT	遺伝子ネットワーク 	発現プロファイル 	化合物 

- オミックスデータベース
  - ゲノム配列，遺伝子発現プロファイル，エピゲノム，化合物データベース
- バイオバンク
  - 検体の収集・管理，．．．

- 個人ゲノム
- コホート
  - 個人ゲノム，SNPs，環境要因，生活習慣

バイオインフォマティクスという分野

「なぜ生命科学の解析にコンピュータが必要なのか？」

↓

① 近年の生命科学データの爆発的な増大！

② コンピュータを用いる2つのアプローチ：

i. 網羅的（ハイスループット）なウエット実験から生成される大量データの処理

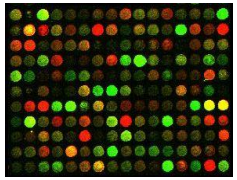
ii. 実験前にコンピュータによる情報処理を徹底的に行うことにより，ウエット実験を絞り込み効率化する ⇒インシリコスクリーニング

③<u>バイオインフォマティシャンの育成は喫緊の課題</u>
（データサイエンティストの育成）

ハイスループット実験装置と情報処理

【マイクロアレイ】

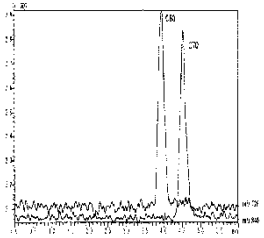
（GenePix）



＋ プロファイル解析プログラム

【質量分析装置】

（島津製作所）



＋ マススペクトル解析プログラム

【次世代シークエンサー】

（Illumina）

ATGCATATGC
CATGCATAGC
ATTGTTGCTA
ATGCATATGC

＋ アセンブリ，マッピングプログラム

次世代シークエンサーの登場

	処理能力 Mbp / 日	リード長 bp
従来のシークエンサー（ABI 社　3730xl）	～2	～800
次世代シークエンサー（Illumina社 HiSEQ4000）	～400,000	～150

20万倍

従来のゲノム解読
数ヵ国のプロジェクトが協力
数億円の規模の莫大な費用

個人ゲノムを1000ドルで読む時代へ

新しい実験プロトコルの定義

【従来の実験プロトコル】

ウエット実験

サンプル抽出，PCR，電気泳動，…

結果：
1 or 0

【網羅的研究アプローチ】
個々の遺伝子をクローニングするのではなく，
ゲノムを丸ごと読んでしまう

装置

データ

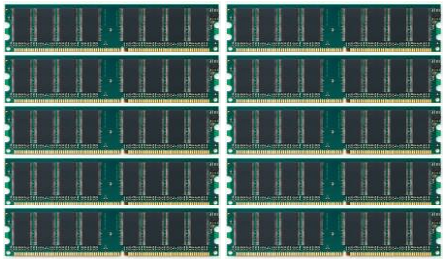
大量のハイスループットデータを解析するためには，
コンピュータを用いた大規模計算が不可欠

データ解析を支えるコンピュータシステム

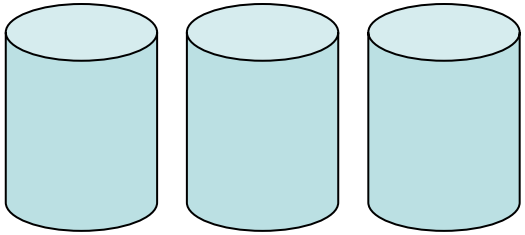
クラスターマシン

① 超並列・分散計算
多数のCPU

② 大容量メモリ
ギガバイト，テラバイト



③ 大容量ディスク
テラバイト，ペタバイト



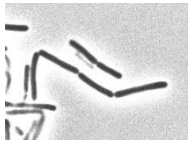
微生物

真核・
高等生物

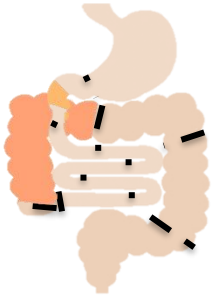
ゲノム解読

さまざまなゲノムを
読んで解析する

納豆菌



メタゲノム
（環境細菌叢）
腸内細菌叢

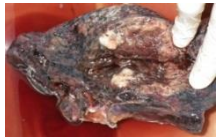


ヤコウタケ

マーモセット

アサガオ

肺腺がん
ゲノム解析



メダカ

RNA解析

発がんゲノム
悪性化のゲノム
変化

講義webページと参考書，単位など

http://www.dna.bio.keio.ac.jp/lecture/bioinfo/

連絡先（メールアドレス）：yasu@bio.keio.ac.jp

教科書なし

参考書：（いろいろ出版されているが．．．）

例えば，「バイオインフォマティクス入門」

日本バイオインフォマティクス学会 (編集)
慶應義塾大学出版会 （2,916円）

単位：期末試験 （＋数回の演習）