

実験家のためのデータ駆動科学オンラインセミナー
(20200717)

ゲノムワイド関連解析をやってみよう

田宮 元

東北大学・東北メディカルメガバンク機構・リスク統計解析室
理化学研究所革新知能統合研究センター・遺伝統計学チーム



ゲノム統計学・ビッグデータ統計学講義に先立って

- 田宮 元
- 東北大学東北メディカル・メガバンク機構・ゲノム遺伝統計学分野
=ゲノムコホート研究
- 理化学研究所革新知能統合研究センター・遺伝統計学チーム
=人工知能研究
- 専門分野：遺伝統計学、ゲノム医学
- ミッション：ゲノムコホートのビッグデータを機械学習・人工知能技術で柔軟に解析する

登山と
極地探検



本日の内容

- 1) GWASをやってみよう
 - ped形式の理解
 - plinkを使う
- 2) 本質的な理解の入り口とする

GWASデータの入手

- GWASデータとは
 - 1) 何らかの表現型値（疾患の有無や血圧など）
 - 2) ゲノムワイドのSNP遺伝子型
 - 3) 交絡因子の値（男女や年齢、集団構造など）
- 入手先
 - HapMapや1kgpなど：表現型は自分で作る
 - WTCCCやADNIなど：研究計画書を書いて申し込む

plink

- Purcellらが作成
Purcell S, et al.
PLINK: a tool set for whole-genome association and population-based linkage analyses.
Am J Hum Genet. 2007;81:559-75.
- ○リンクではなくプリンクと呼ぼう
- Population-based linkageの略
- 様々な統計手法が実装されており、ファイルのフォーマット変換やデータ抽出用のコマンドも多い
- v1.07→v1.9 (plink2) : 高速化されたがバグが多い

求められている人材

IT・情報処理一般
OS (Unix/Linux)
ネットワーキング
プログラミング
データベース



人工知能
確率論・統計学

数学一般

疫学 遺伝疫学

遺伝統計学

量的遺伝学

集団遺伝学

人類遺伝学

遺伝学一般

(メンデル遺伝学)



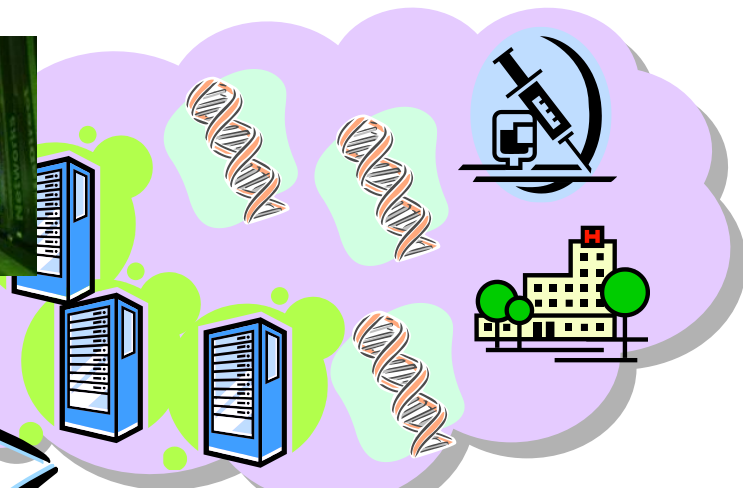
ゲノム科学

分子生物学

遺伝子工学

生化学

医学・生物学一般



大規模実験・データベース
超高速計算機



どうやって学ぶか(まとめ)

- (講義を受ける)
- 中心となる教科書をひたすら通読する
C&K, Introduction to Population Genetics. 1970
Three books by Nei, 1975, 1987, 2000
Kimura, The Neutral Theory of Molecular Evolution. 1983
- 原典をひたすら読解する
- データで探索的に分析を経験してみる
- データ分析を手伝ってみる
- 散歩する
- 数理スタッフと議論する
- 以上を数年間、欠かさず続ける

