

予防・個別医療に向けた医療・生命 時系列データのデータ駆動型モデ リング

理化学研究所

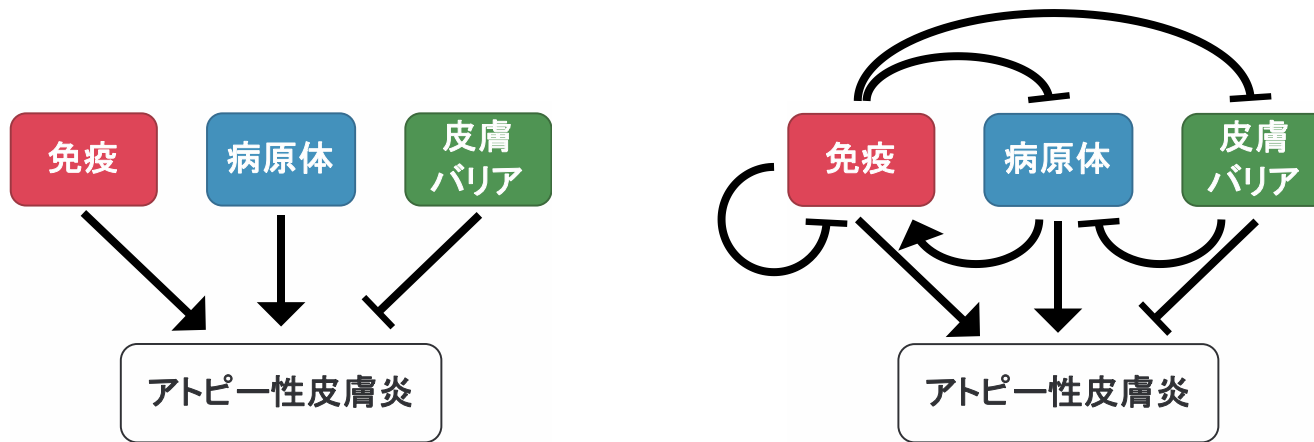
医科学イノベーションハブ推進プログラム

疾患機序研究グループ

川上 英良、五島 祐樹

1

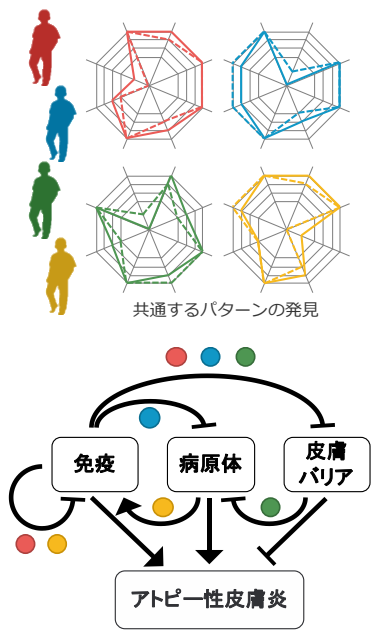
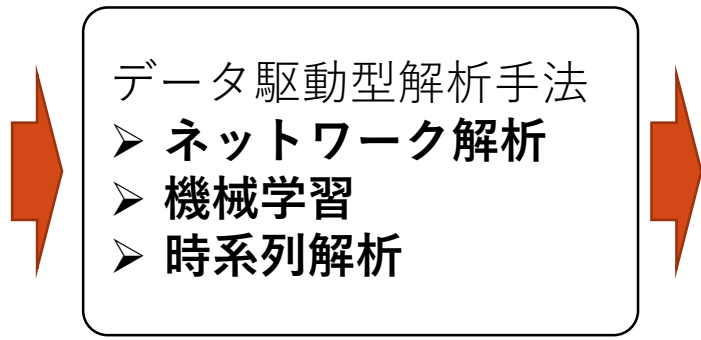
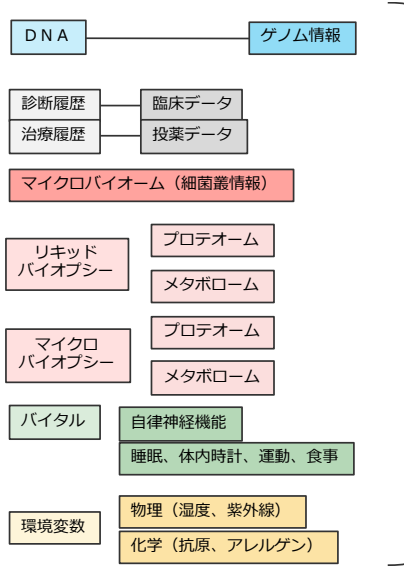
生命をシステムとして理解する



少数の因子の個別の影響の総和で複雑な現象の総体を説明することは困難
⇒因子同士の**時間的・空間的相互作用を考慮したモデル化**が必要

データ駆動型モデル化の流れ

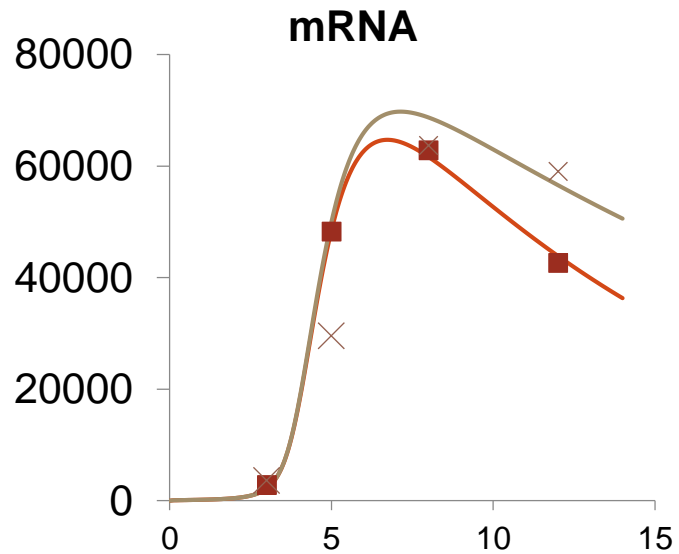
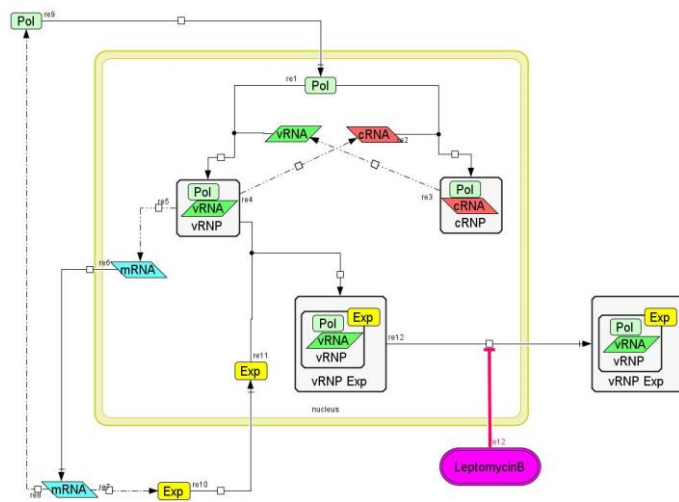
高次元・多階層 網羅的測定データ



個別性、非線形性を
考慮した生命・疾患モデル

従来の時系列データ解析の枠組み

少数因子の詳細な計測データと微分方程式モデルに基づくシミュレーション



Kawakami et al J Virol Methods 2011
のデータに基づく

医学・生物学データを時系列として扱う上での問題

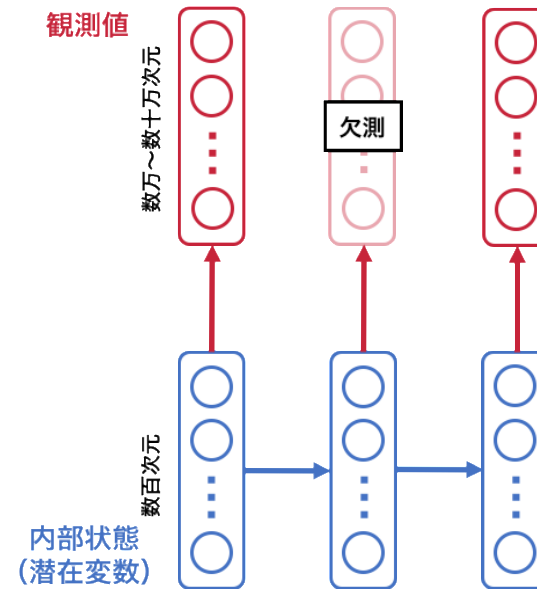
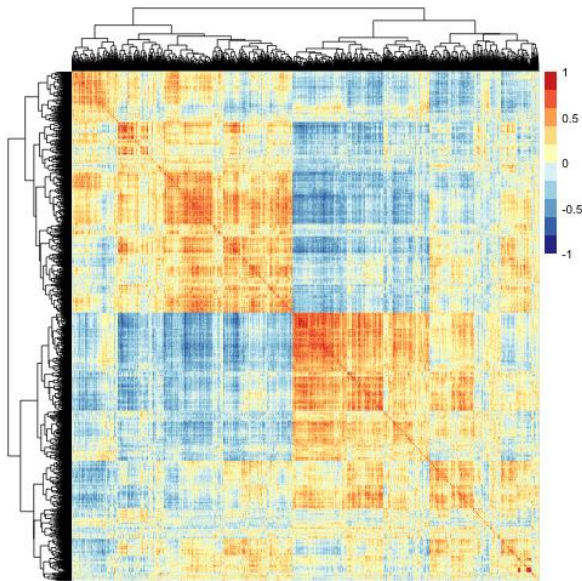
1. 測定項目の次元数が非常に多い
2. 因子間の相互作用が非線形
3. 計測時刻が散発的で不規則、開始時刻が一定でない
4. 内部因子だけでモデルが完結しない

一般的なシミュレーションモデルベースの時系列解析の枠組みがそのままでは使えない

本研究の目的

高次元で散発的な時系列測定データから状態遷移の規則・秩序を抽出する汎用的な手法を開発する

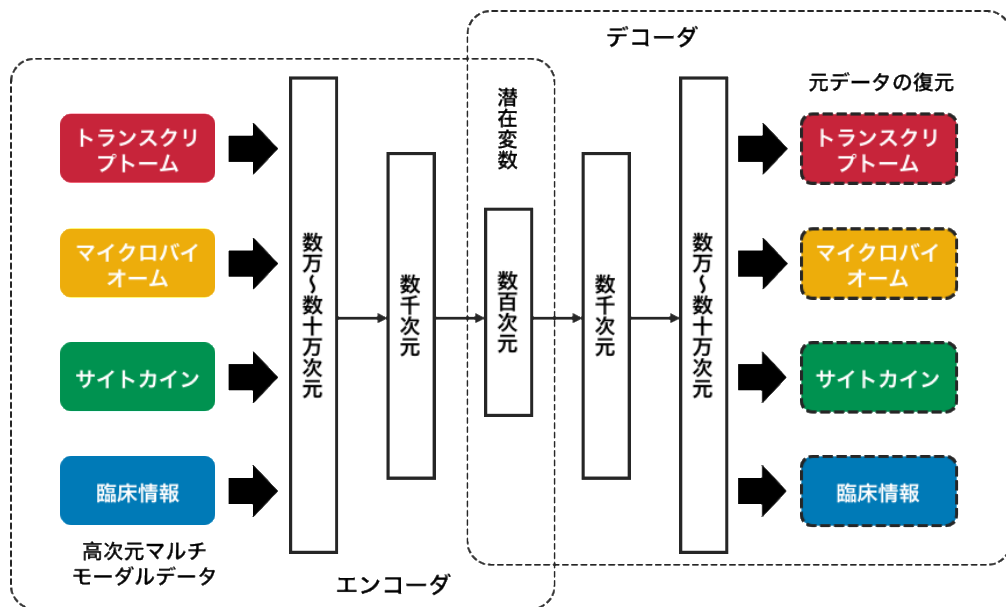
状態空間モデルに基づく生命現象の理解



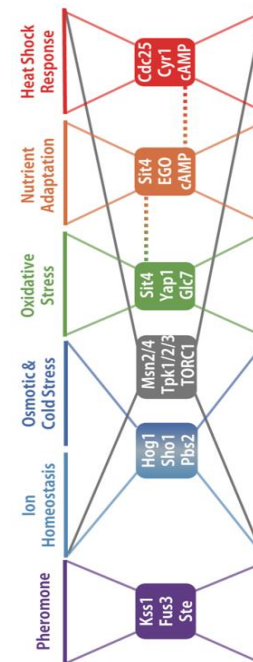
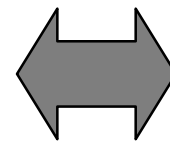
大腸菌4297遺伝子間に見られる相関
⇒少数の**制御因子**（**潜在変数**）の存在

状態空間モデルに基づく
生命現象の状態遷移モデル

次元削減・圧縮

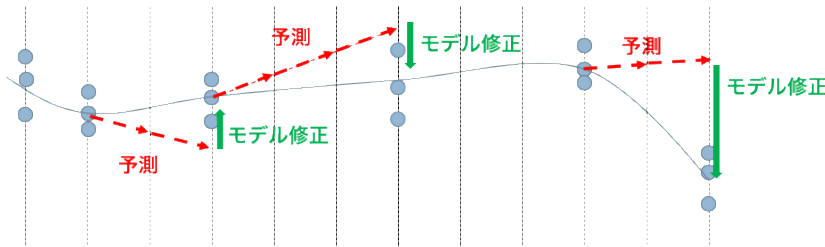


ディープラーニングによる
非線形相関を考慮した生命データの次元圧縮



実際の生物の制御ネットワークとの比較
による潜在変数の意味付け

逐次予測と観測データによるモデル修正



ただし、線形のVARモデルは「発散」「収束」「周期振動」のいずれかの挙動に限られる



係数行列 A の部分をニューラルネットワークで置き換えた、非線形VARモデル

多変量自己回帰 (VAR) モデル

$$X_t = C + A X_{t-1} + E_t$$

次の状態ベクトル

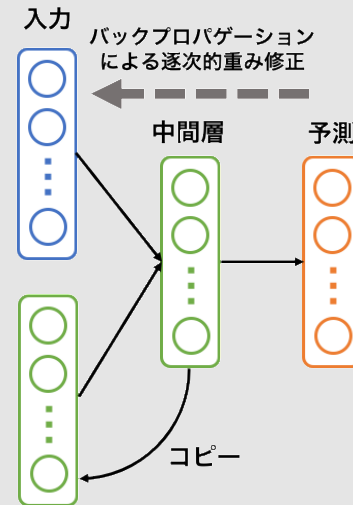
ベースライン

係数行列

1つ前の状態ベクトル

ノイズ

1つ前の状態によって次の状態が決まる変数間の影響をネットワークモデル化可能



大腸菌トランスクリプトームデータを用いた手法の検証

D866–D870 *Nucleic Acids Research*, 2008, Vol. 36, Database issue
doi:10.1093/nar/gkm815

Published online 11 October 2007

Many Microbe Microarrays Database: uniformly normalized Affymetrix compendia with structured experimental metadata

Jeremiah J. Faith¹, Michael E. Driscoll¹, Vincent A. Fusaro¹, Elissa J. Cosgrove², Boris Hayete¹, Frank S. Juhn², Stephen J. Schneider² and Timothy S. Gardner^{1,2,*}

¹Program in Bioinformatics, Boston University, 24 Cummington St. and ²Department of Biomedical Engineering, Boston University, 44 Cummington St., Boston, Massachusetts, 02215, USA

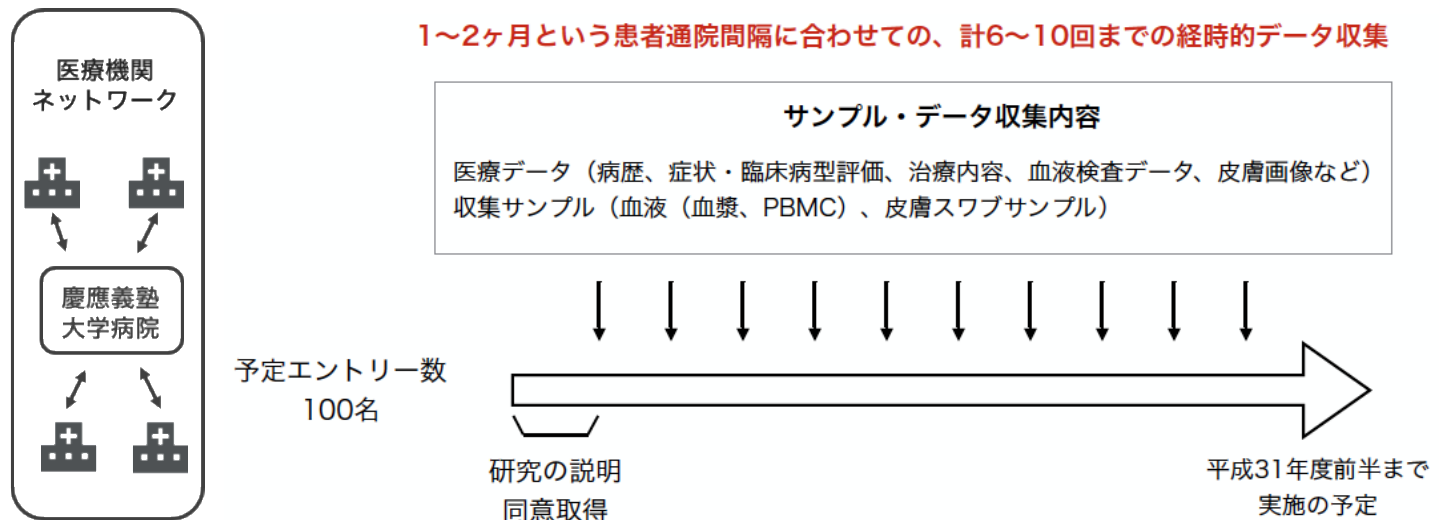
- 1,000以上の様々な刺激、栄養条件下の大腸菌のトランスクリプトームデータ
- 約300時点の時系列トランスクリプトームデータ

検証課題

1. ディープラーニングによる次元圧縮は大腸菌の制御ネットワークを再現できるか
2. 時系列モデルの学習には何例のデータが必要か
3. 学習によって得られた予測モデルは残りのデータを予測できるか

臨床マルチモーダルデータの経時的取得

慶應義塾大学病院を中心とする医療機関ネットワークを通じた
アトピー性皮膚炎患者の臨床マルチモーダルデータ取得体制



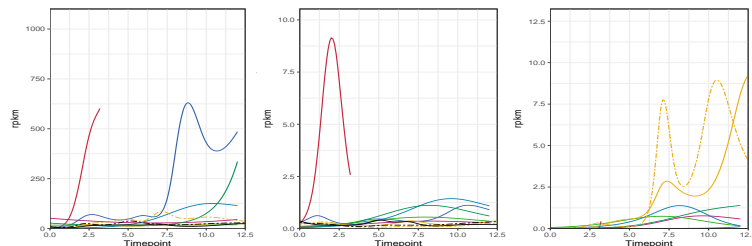
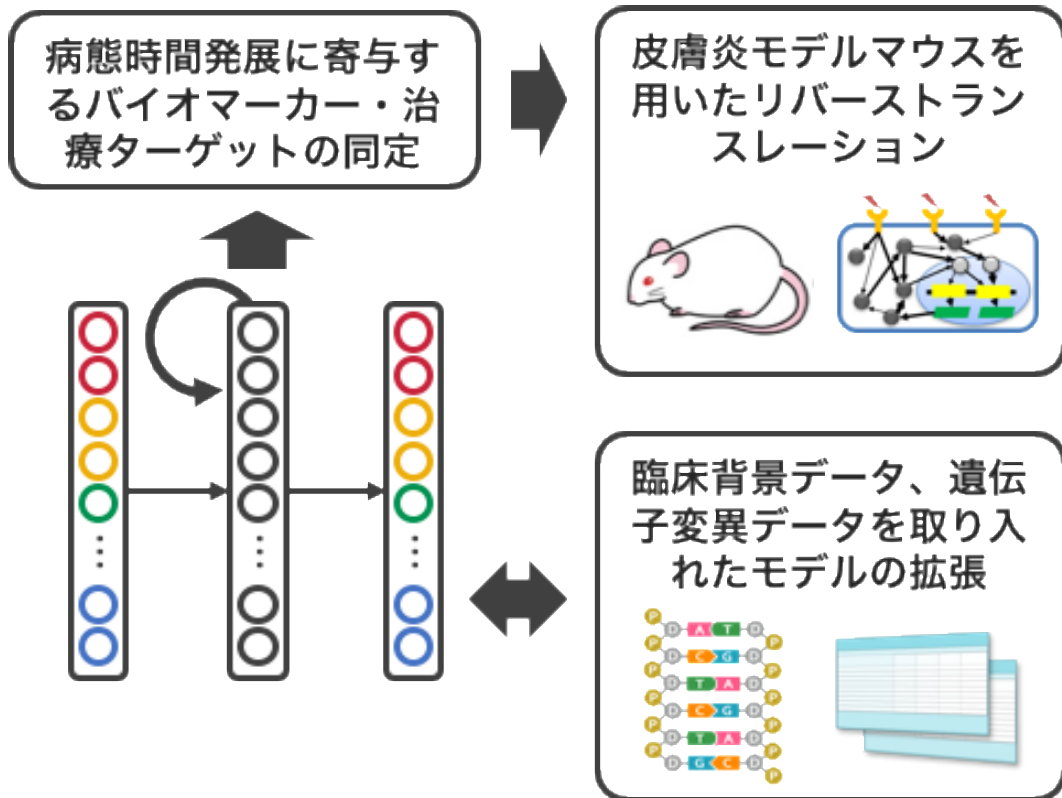
3'UTR-seqによるトランスクリプトーム測定

16S seqによるマイクロバイオーム測定

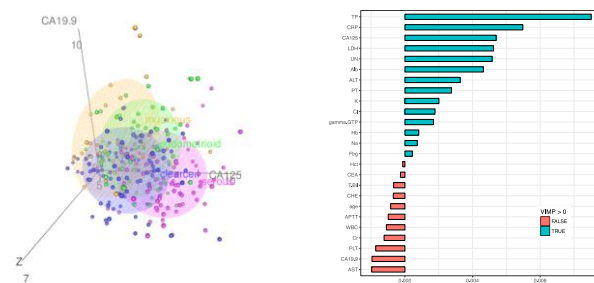
Bio-Plexによる48項目サイトカイン測定

臨床情報データ取得

状態遷移予測モデルに基づく疾患の予測、理解



様々な皮膚炎モデルマウスにおけるサイトカイン時間変動の違い



機械学習による特徴選択と臨床背景への関連付け

研究推進体制

次元削減・圧縮手法の開発

川上 英良 (医科学イノベーションハブ)

中岡 慎治 (さきがけ数学協働領域)

古関 恵太 (医科学イノベーションハブ)

状態遷移予測アルゴリズム開発

川上 英良 (医科学イノベーションハブ)

中岡 慎治 (さきがけ数学協働領域)

石川 哲朗 (医科学イノベーションハブ)

五島 祐樹 (医科学イノベーションハブ)

臨床データ収集・リバーストランスレーション

川崎 洋 (慶応大学・皮膚科)

福島 彩乃 (慶応大学・皮膚科)

海老原 全 (慶応大学・皮膚科)

梶下 紘貴 (統合生命医科学研究センター)

システム生物学、数理科学、計算機科学、生物学、臨床医学を高いレベルで融合することで、データ駆動型モデル化のリーディングケースに