医療から見た期待(必要性）
ゲノム・オミックス医療の将来

東京医科歯科大学　データ科学推進室
東北大学　東北メディカル・メガバンク機構
田中　博

ゲノム・オミックス医療

バイオテクノロジーの
急速な進展による
「ビッグデータ医療」時代の到来

医療・創薬への超大なインパクト
ビッグデータ時代の到来

(1) 次世代シーケンサ（Clinical Sequencing)を始めとする
「ゲノム/オミックス医療」における網羅的分子情報収集/蓄積
(2) **Biobank**/ゲノムコホート普及による分子・環境情報の蓄積
(3) モバイルヘルス(mHealth）によるWearable センサの連続計測による生理データの蓄積（unobstructed monitoring)

↓

コストレスで良質なデータが大量に収集可能

↓

治療医学の的確性の飛躍的進展:「精密医療」
医療の国民レベル・生涯ヘルスケアの進展

ゲノム・オミックス医療の２つの流れ

米国でのゲノム医療

Precision Medicine (精密医療）

- 「シーケンス革命」(2007）からの**怒涛の展開**（2010から）
- 個々の患者の「治療医学」レベル質的向上:臨床実装の推進
  - 稀少疾患の原因遺伝子変異の同定
  - がんのドライバー遺伝子変異の同定と分子標的薬の選択
  - 薬剤代謝酵素の多型性の同定と個別化投与

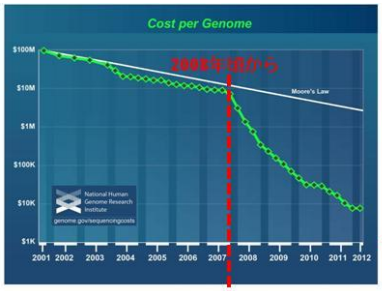
欧州でのゲノム医療

Genomic Biobank (バイオバンク)

- 「集合的遺伝情報」の価値⇒**ゲノム・バイオバンクへ流れ**
- 国民医療（医療の国民レベル）の向上: 社会福祉国家の理念
- 「予防医学」レベル質的向上のためにゲノム情報導入
  - 大規模前向きpopulation型バイオバンク/ゲノム・コホートの確立
  - 遺伝的素因と環境要因（生活習慣）との相互作用に基づいた「多因子疾患」の発症予測を通した「国民医療の向上」
  - 生涯的健康/疾病管理へ

TMDU

米国ゲノム・オミックス医療の流れ



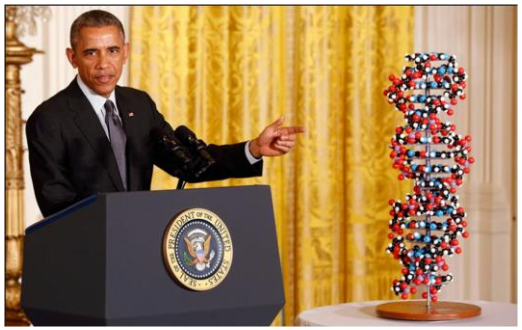
DNA Sequencing Cost: the National Human Genome Research Institute

シーケンス革命　2007/8

2005～ NGS 454 (LS,Roche)
2007/8～454, Solexa (Illumina), SOLiD (LT,TF)
シーケンス革命

	HiSeq2500		Ion Proton
本体価格	約1億円		約3500万円
モード / チップ	ハイアウトプット	ラピッドラン	Ion Proton I
解析時間	11日	27時間	2時間
リード長（bp）	2 x 100	2 x 150	200
データ産出量(Gb)	約600	約120	10
試薬コスト（ヒト1人全ゲノム）	数十万円		不可 エクソームのみ

急速な高速化と廉価化
ヒトゲノム解読計画13年,3500億円
⇒1日,10万円



オバマ大前統領　Precision Medicine Initiativeを開始、2015年1月　大統領一般年頭教書演説

先陣争いの時代

第一期

ゲノム多型性の認識
Hapmap 計画(2002)
GWAS研究など

薬剤代謝酵素の多型性の判別・電子カルテで警告・Preemptive PGx
Vanderbilt大病院

2005～ NGS 登場
(454,Solexa,SOLID)
2007/8～
シーケンス革命

Undiagnosed Genetic Diseaseの原因遺伝子POC同定
MCW小児病院

国際がんコンソーシアム開始
ICCG (2008年)
2011頃からがん変異成果報告

Cancer Driver Geneの同定と抗がん剤治験
Dana Faber CC

ゲノム・オミックス医療の臨床実装の普及
ゲノム・オミックス情報のビッグデータの出現

国家政策の時代

第二期

ゲノム医療の国家的取組み
NIH “BD2K “計画 ・各種ゲノムコンソーシアム開始

オバマ大統領　年頭教書
Precision Medicine initiative 政策の発表

精密医療普及期

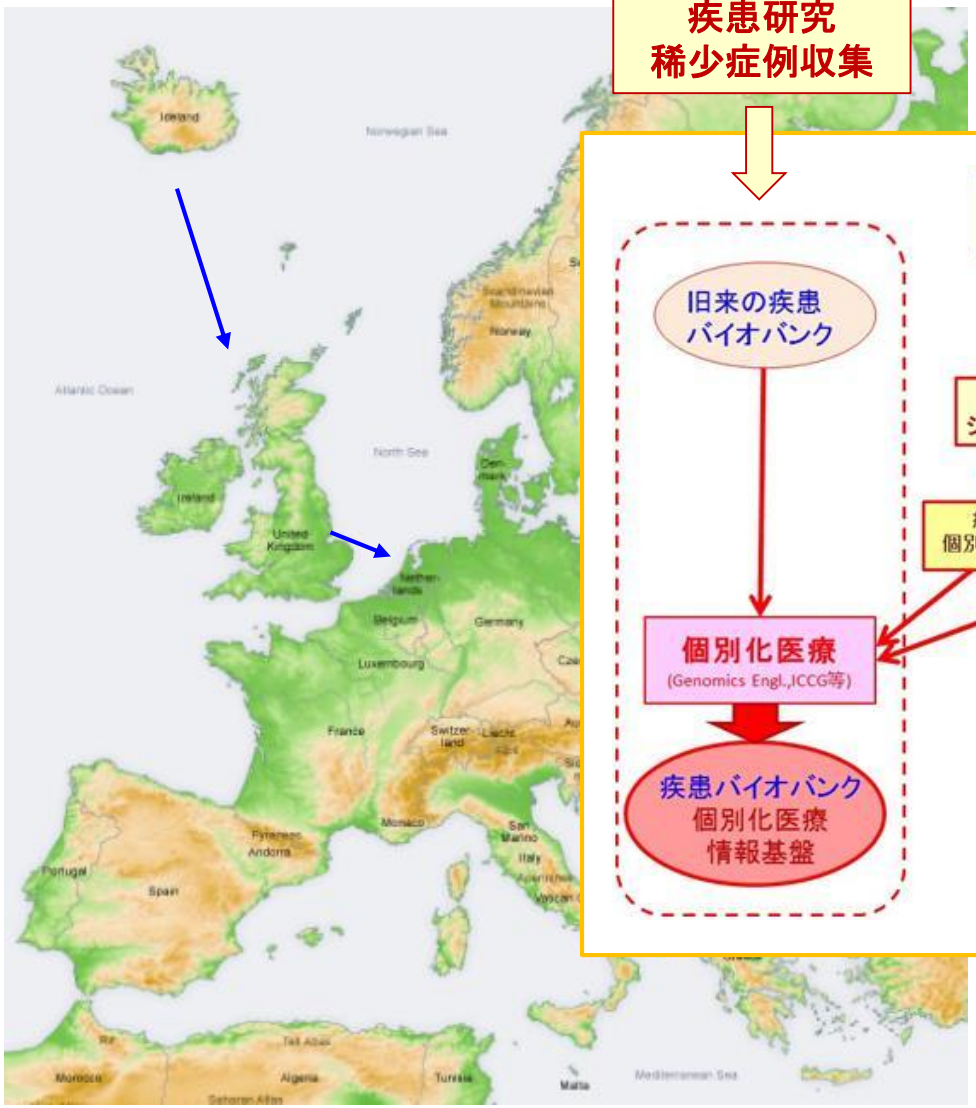
第三期

100万人コホート:バンダービルト大学設開始(2016-2020)
NCI“National Cancer MoonShot” 10年計画開始
各州でのプレシジョン医療計画開始(カリフォルニア、ペンシルバニア)

2007年
2009年
2010年
2011年
2012年
2013年
2014年
2015年
2016年
2017年

TMDU

第2の流れ　欧州のバイオバンクの普及



疾患研究
稀少症例収集

「集合的遺伝情報」による
国民レベルでの医療向上

旧来の疾患
バイオバンク

ヒトゲノム計画
(1990-2003)

次世代
シーケンス

GWAS

疾患
個別化機序

ゲノム
生得的
多型性
SNPアレイ

個別化医療
(Genomics Engl.,ICCG等)

疾患バイオバンク
個別化医療
情報基盤

Island/deCODE社
国民遺伝情報DB

前向きコホート概念

ELSI慎重対応

Population準拠

家系情報

UKバイオバンク
ゲノムコホートの確立

LifeLines
3世代コホート

東北メディカル
メガバンク

出生コホート

大規模前向きゲノムコホート
国民的医療・予防医療
情報基盤

ビッグデータ医学/医療の２つの流れに起因する大規模な生命情報DB/KBの出現と利用

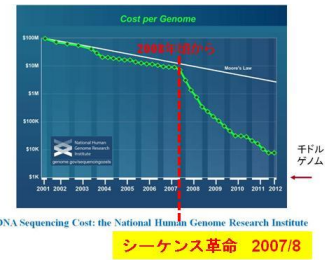
- ヒトゲノム解読計画以降急速に進展
  - Hapmapプロジェクト,1000 genome,がんICGC,TCGA、TopMED
  - ゲノム変異・多様体
    - dbSNP, HGMD, **Clinvar**, **Clingen**, OMIM,GWAS catalog
    - 表現型との対応: dbGaP, EGA
  - 遺伝子発現プロファイル
    - 疾患特異的transcriptome: **GEO, ArrayExpress**,
    - 薬剤特異的trancriptome: **c-Map, LINCS**
  - タンパク質
    - ３次元構造：PDB, Swiss-Prot,
    - タンパク質間相互作用：**HPRD, STRING,** BIND
  - 分子ネットワーク、パスウェイ
    - KEGG, TRANSFAC, BioCyc, Reactome
- 各種バイオバンク症例ベース（制限アクセス）
  - UK biobank, BMBRI, 東北メディカル・メガバンク
- これらの大規模DB/KBを組合せてゲノム医療/創薬を推進

TMDU

医学/医療へのビッグデータの衝撃

次世代シーケンサの登場

シーケンス革命（2007)



コストレスで高精度な網羅的分子情報の出現

1. ゲノム・オミックス医学/医療の進展
   - Clinical Sequencingによるゲノム・オミックス医療の臨床実装の急速な進展
2. Biobank/ゲノムコホートの世界的普及
   - 個別化医療/予防の情報基盤として普及
3. 大規模な生命情報DB/KBの出現
   - ゲノム・オミックスによるDB/KBの膨大化

わが国での現状「ゲノム医療元年 (2016)」

■「ゲノム医学実現推進協議会」(中間報告）2015.7

研究費を用いた試行的ゲノム医療であるが、いくつかの医療施設で
ゲノム・オミックス医療が試行されていた

- 例：がんの網羅的分子診断と個別化治療
  - 国立がん研究センター（Top-gear、SCRUM-Japan)
    - ドライバー遺伝子の診断。分子標的薬の治験グループに割当て
    - がんのゲノムパネル：来年先進医療（7施設）
  - 岡大,京大, 北大,千葉大 病院併設型BB
- 予定：2018年度より「がんゲノムパネル」先進医療（7か所）開始

■AMED（日本医療研究開発法人）がゲノム医療を推進に予算

- IRUD（Initiative on Rare and Undiagnosed Disease)
  未診断疾患の原因遺伝子をIRUD拠点病院が審査して解析センターがシーケンシング。その後、DB化する。
- ゲノム医療実現推進プラットフォーム事業
- 臨床ゲノム情報統合DB事業

ゲノム医療臨床実装では、米国と水を空けられている。しかし、
Biobank/Genomic Cohortでは我が国の状況は遅れていはいない。
米国とは異なったBiobank準拠のゲノム医療/創薬を推進するべき

TMDU

ビッグデータ医療への期待（必要性）

医療の「新しいビッグデータの革命性」

～ゲノム・オミックスデータの基軸的な特徴～

＜**目的**も**データ特性**も従来型と違う＞

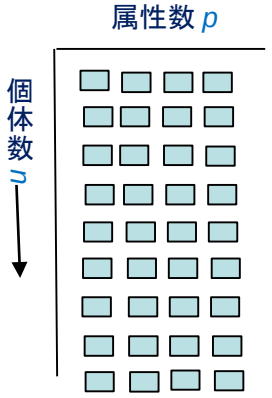
従来の医療情報の「ビッグデータ」(n≫p)

医療情報・疫学調査では **属性数**：**数十項目程度**

個体数：近年電子化の流れ⇒**個体数：膨大**

– 目的：Population （集合的）医学のBig Data

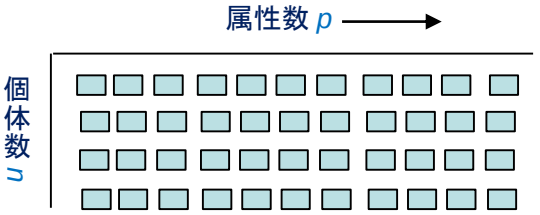
⇒個別を集めて「集合的法則」を見る



網羅的分子情報などのビッグデータ(p≫n)

１個体のデータ属性数が膨大（SNP4000千万）

ただし個体数は大規模biobankでも数十万



属性(*p*)>> 個体数(*n*) :従来の多変量統計学が無効

「新 ***np*** 問題」: GWASは単変量解析の羅列

– 目的：医療の場合 個別化医療 Personalized Medicine

⇒大量データを集めて「個別化パターン」の多様性を摘出

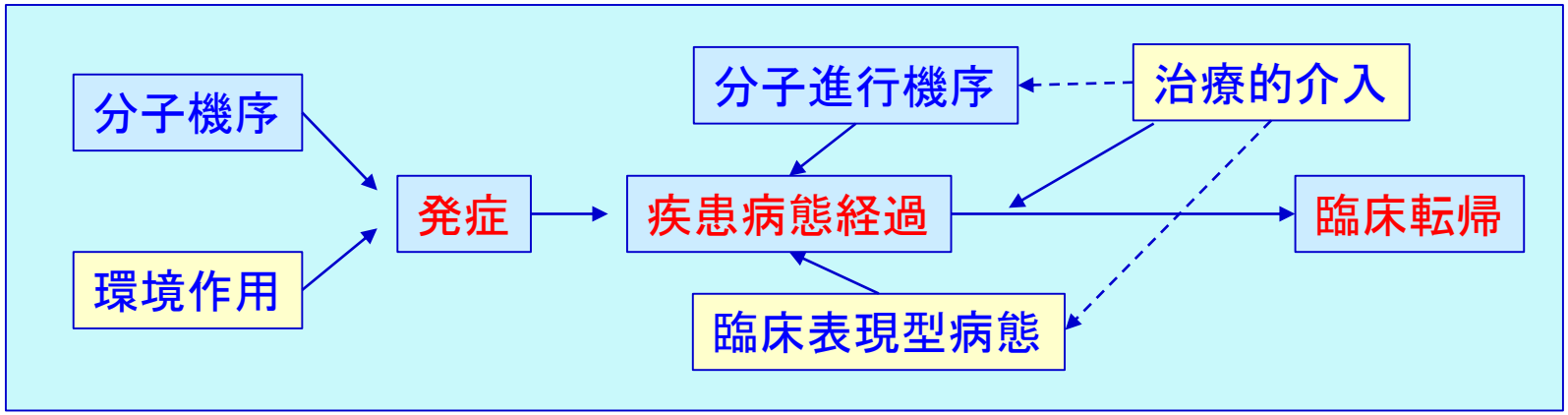
新しいデータ科学の必要性

TMDU

医療の「ビッグデータ」革命は どんな既存のパラダイムに変革しているか

- Population （集合的）医学からパラダイム転換
  - <One size fits for all>の集合的医療はもはや成り立たない
  - 個別化医療 “Personalized medicine”の概念
    - 個別化医療実現のために＜個別化・層別化パターン＞がどれだけ有るか網羅的に調べる：どこまでの粒度で個別化・層別化すればよいか
- Clinical research（臨床研究）のパラダイム転換
  - 臨床研究の基礎：従来の範型RCTは,個別化概念を取扱えない
  - <EBM: (statistical) evidence based>の呪縛からの解放
  - 「標本」統計・「推測」統計学に制約されない臨床研究
  - Real World Data・ビッグリアルワールドデータからの知識生成
  - Learning Health System: 学習的医療実践

期待１：対応する非ゲノム病態データの検証的「情報化」



疾患経過のオントロジー

疾患分子発症進行機序（生体分子ネットワーク）

対応する非分子機序の明確化

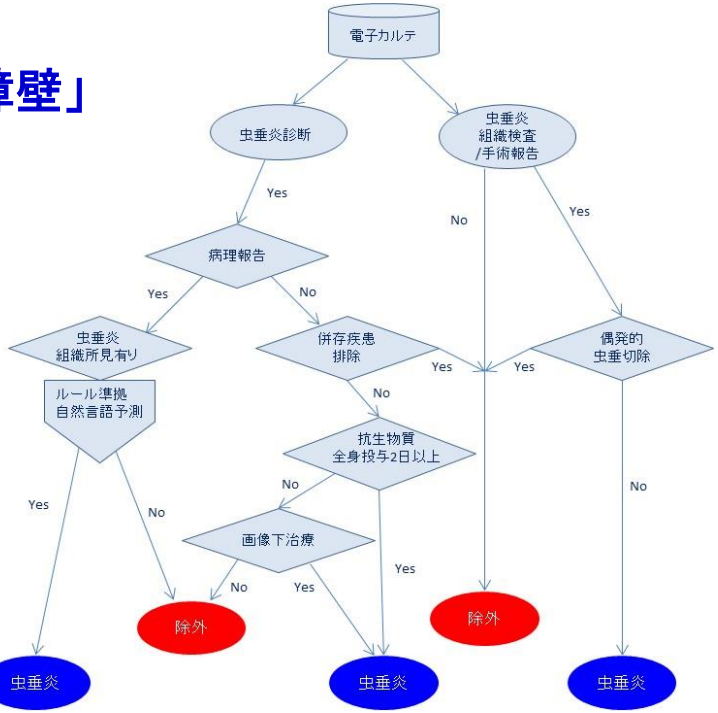
環境発症要因　臨床表現型情報　治療介入効果

臨床表現型との統合(phenotyping)

臨床表現型データ検証的抽出、「非構造化データ障壁」

electronic MEdical Records + GEnomics (NHRI-funded

- **phase I** (2007-2011) EMR-bsedゲノム研究の探求
  - EMR(臨床phenotyping)とbiorepositoryに基づくGWAS等（EMR-based GWAS）が可能か（LHS）。
    - 開始時はGWAS全盛時代。ゲノム医療の臨床実装未着手
  - 電子カルテより臨床表現型情報抽出 phenotypingルール
  - 計画開始時参加施設：Mayo, Vanderbilt Univ, Marshfield, Univ. Washington, Northwestern Univ.など５施設,
- **phase II** (2011-2015) 臨床実装へ舵を切る
  - MCWの臨床実装のインパクト,Vanderbiltの先制PG x
  - 電子カルテと遺伝情報の統合
    - 電子カルテへのゲノム情報の統合
    - **PheKB** (Phenotype Knowledge Base)
    - ゲノム医療の実装、PGxの臨床応用
    - 結果回付 **Return of Result,** ELSI等
  - 4つのサイトが新しく加わる
    - 小児病院グループとMount Sinai, Geisinger
- phase III （2015より始まる
  - NHGRIのコンソーシアムと連携
  - とくに**CSER** “Clinical Sequencing Exploratory Research”



PheKB: phenotyping ルール

期待２：生命医療ビッグデータから AIによる「革新的(innovative)知識」発見

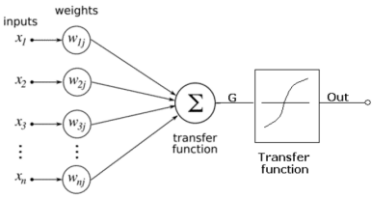
- 臨床ゲノム医学
  - 全ゲノム配列の普及、多層オミックス情報の収集、分子画像の発展
  - ビッグデータ化：超多次元相関ネットワーク
  - 〈網羅的分子情報と臨床表現型情報〉の相関
- 予防ゲノム医学
  - バイオバンクの大規模化、国際連携によるバーチャル連携
  - 〈遺伝的素因と環境/生活様式要因〉の相互作用と発症の相関ネットワーク

いずれも超多次元複雑ネットワークの縮約理論

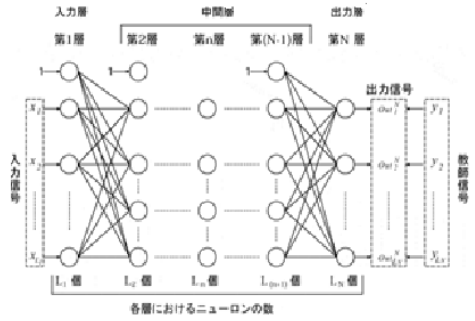
TMDU

人工知能 Deep Learning への期待

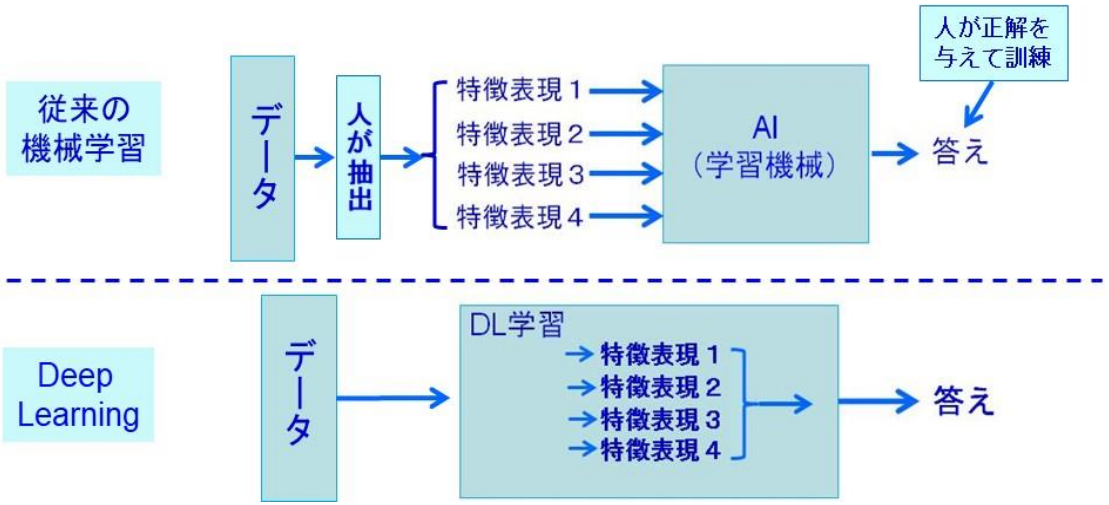
- 機械学習のこれまでの限界
  - 「教師あり学習」
    - 分類対象の特徴と正解を与え学習機械（AI）を構築
- Deep Learningの革命性
  - 「教師なし学習」
    - 対象の特徴表現や対象の高次特徴量を自ら学ぶ



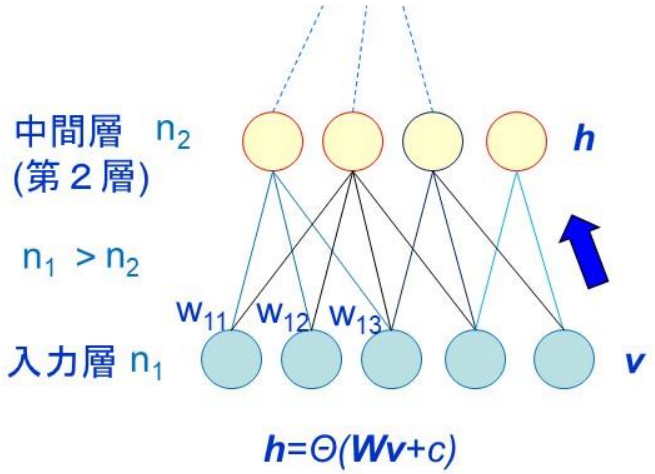
神経情報素子



多層ニューロネットワーク



Autoencoder方式



Deep Learningによる
多次元ネットワーク縮約法
(Hase, Tanaka 2017)

- 医療・創薬ビッグデータへの応用性は高い
- 超多次元ネットワーク情報構造の急増
  - ゲノム医療＜網羅的分子情報ー臨床表現型情報＞
  - ゲノムコホート＜遺伝素因ー環境要因(生活習慣)＞
- Deep Learning-based Network Contraction
  「DLネットワーク縮約法」

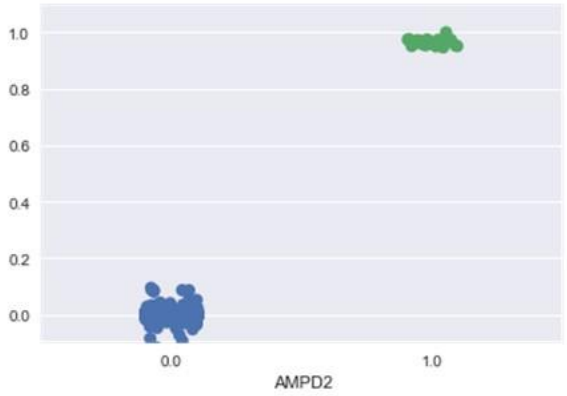
  超多次元ネットワーク情報構造⇒
  少数の特徴的ネットワーク基底に分解
- 線形分解ではない。非線形分解で基底への射影
  - 線形分解（特異値分解：SVD）との比較

TMDU

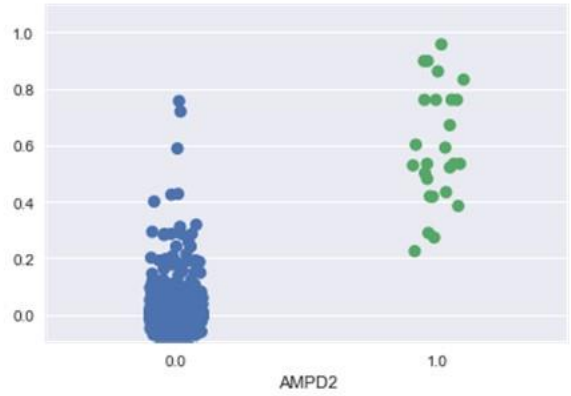
Deep Learningと特異値分解比較

それぞれの遺伝子（タンパク質）の「隣接ベクトル」の復元精度

AMPD2 (adenosine monophosphate deaminase 2)
degree=26



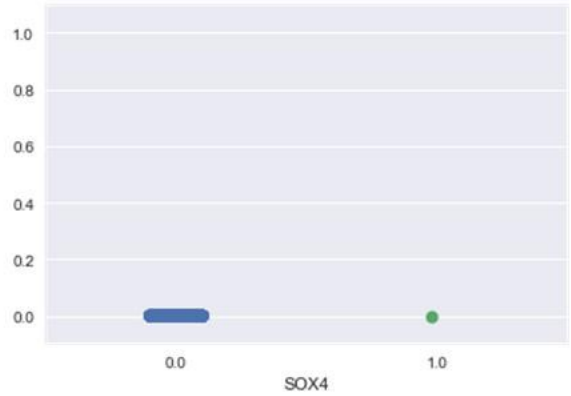
Autoencoder



SVD

SOX4 (SRY-box 4)
degree=1

Autoencoder



SVD

N=8,502

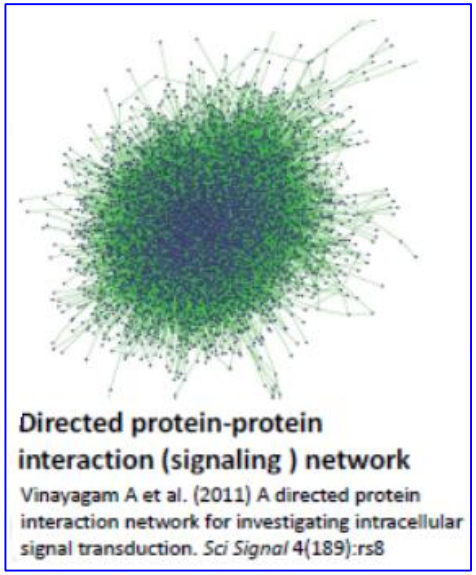
TMDU

特徴的ネットワーク基底への分解

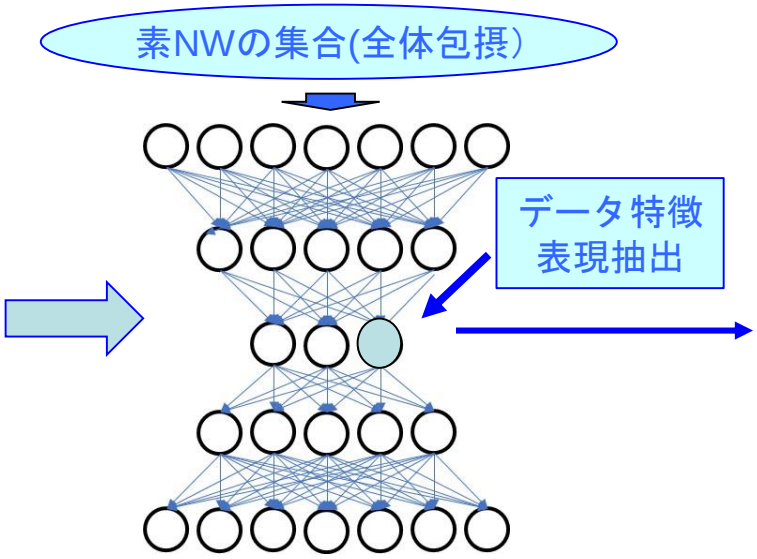
特徴的ネットワーク基底の和に縮約

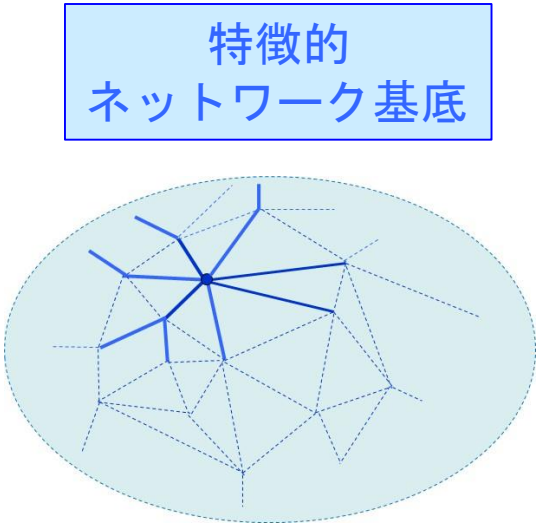
特定のノードを起点とした素NW（部分NW）の集合
全体NWを包摂する集合にDL反復自己学習

特徴的ネットワーク基底：トポロジーのみの構造/頻度構造



PPIネットワーク

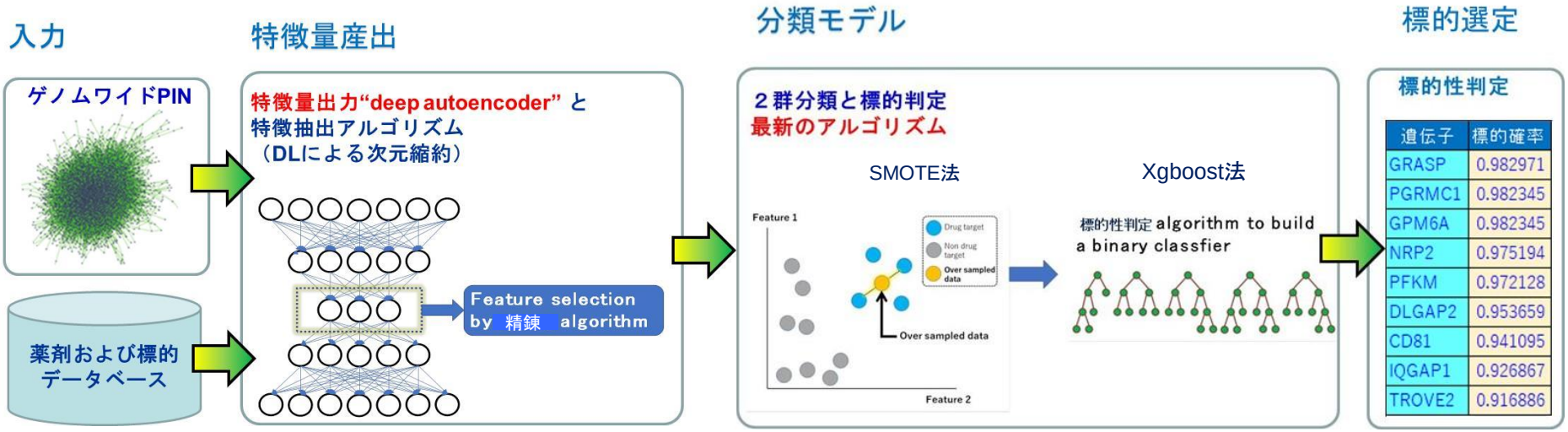




Deep Learningによる創薬・DR

分類部　**DrugBankを利用した**
当該分子を標的とする既製薬剤の探索

既製薬剤がない→新規薬剤探求（創薬）
既製薬剤がある→DRの検討



推定した標的分子は実験的研究でも検討されている

期待３：ゲノム医療の第2世代へ

成功した臨床実装

1. **希少先天遺伝疾患**の**原因遺伝子**を**病院の現場で**シーケンサにより同定
2. **がんのドライバー遺伝子変異**を同定、適切な**分子標的薬**を処方
3. 患者の**薬剤の代謝酵素の多型性**を**先制的に同定**し、副作用を防ぐ

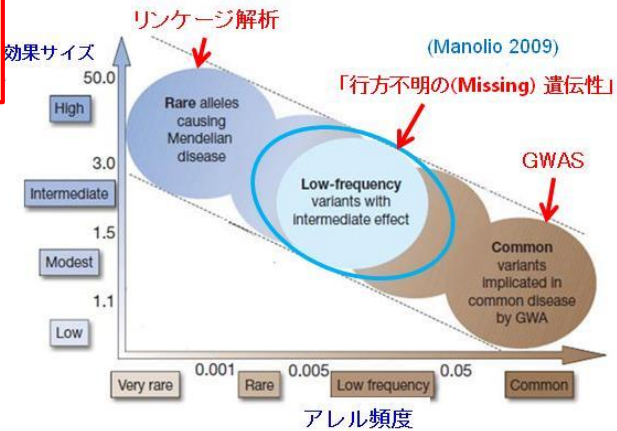
しかし

多因子疾患の機序/発症予測は無着手である

- 「単一遺伝的原因」帰着アプローチの限界
- 「行方不明の遺伝力」の主要な原因
  - 複数の疾患関連遺伝子間の相互作用: G x G
  - 環境と遺伝子の相互作用が：G x E

SNPの相対リスク
低い(1.1～1.3)理由
GxE組合せ特異的効果を環境要因の平均

多因子疾患は個人の＜遺伝的体質と環境要因＞の＜相互作用の結果。シーケンスだけでは解明不能



疾患発症の遺伝要因と環境要因の相互作用は
加算的 $(G \oplus E)$でもなく乗算的 $(G \otimes E)$ でもない
＜(G,E) 組合せ特異的な効果＞である
例　大腸がんの遺伝要因と環境（生活習慣）要因

発達プログラム説 DOHaD

Developmental Origin of Health and Disease)

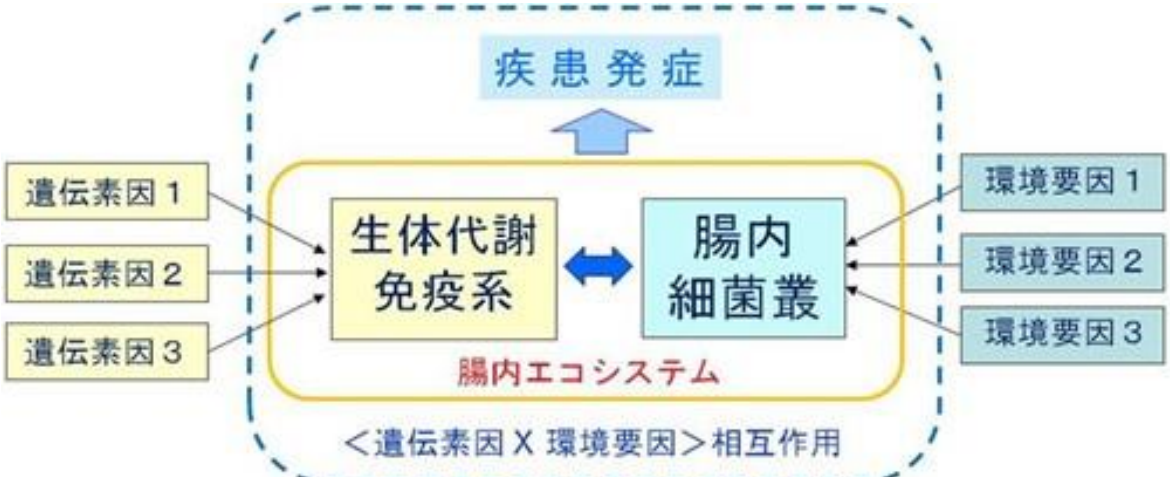
- オランダ飢饉
  - 第2次大戦末期、ナチスの封鎖、約半年間酷い飢饉
  - 飢饉の期間に胎児、戦後30年
  - 成人期:肥満,糖尿病,心筋梗塞,統合失調
- Baker仮説：英国心筋梗塞増加
- エピジェネティク機構
  - 過度な低栄養：肝臓のPPARα/γ（倹約遺伝子）メチル化低下・遺伝子発現がオン
  - エピジェネティック変化は**可変**：短期的変化、**長期的「記憶」**次の世代も

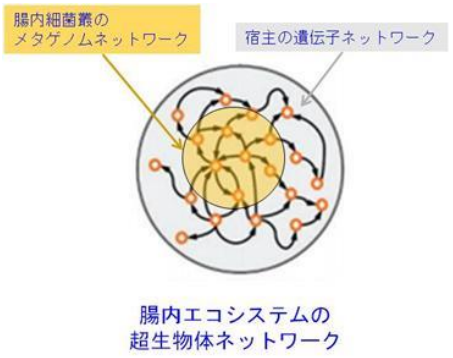
オランダ
飢饉 (1944)

環境因子 → Epigenome変化 → 遺伝子発現調節 → 疾病発症

腸内細菌叢microbiome：メタゲノム

- **疾患の環境発症要因**（exposome）
  - 腸内microbiome：環境要因の最大の１つ
- **腸管微生物叢**（gut microbiome）
  - 約1000種類、100兆個、総重量１～1.5kg,「**実質的な臓器**」
  - 遺伝子数個人あたり**約50万遺伝子**、総数：数100万遺伝子
- **免疫系、炎症系、粘膜免疫細胞群との相互作用**
  - **食物の難消化性の食物繊維**：腸内細菌によって嫌気的に代謝、酪酸などの「**短鎖脂肪酸**」がエネルギー源となる
  - 食事・栄養物質による環境要因は、腸内細菌叢の代謝物（短鎖脂肪酸やTMAOなど）から宿主の生体機構に相互作用

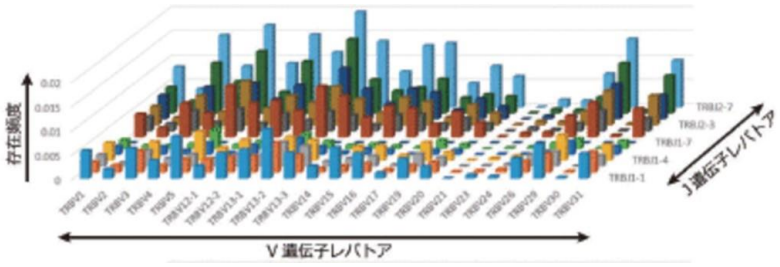




メタゲノム
超生物ネットワーク

免疫ゲノム

- 可変領域や相補性決定領域（特にCDR３）のDNAやRNAを次世代シーケンサ(HTS)で解析
- レパトア解析
  - 抗原受容体全体のプロファイルを俯瞰的に把握できる
  - V(D)Jなどの成分を基軸として3次元表示可能。
  - 疾患罹患ともに瞬時に全体像が変化する。
  - 網羅的病態全体像を提示する
    - VDJの使用頻度
    - 多様性(diversity)の変化
    - 疾病/加齢レパトア分布変化
- 臨床シーケンスに含まれる
- ３次元分布の特徴分析



（レパトア・ジェネシス社）

第2世代のゲノム・オミックス医療

- 生涯的全体性においてその個人の疾患可能性の全体性を把握し、個別化予防、個別化治療に取り組む
- ゲノム・オミックス情報と医療・健康
  - **Clinical Sequencingのインパクト**
- 第１世代ゲノム医療
  - ゲノムの変異・多型性の個別性に基づく
- 第２世代のゲノム医療
  - 多因子疾患が対象、環境情報との相互作用
  - エピゲノム、メタゲノム・免疫ゲノムなど

疾患メタ・オミックス修飾

TMDU

まとめ

- 「ビッグデータ医学・医療」時代の到来
  - 仮説駆動からデータ駆動型サイエンスへ
  - 人工知能（AI）の必要性、DLの次元縮約
  - 多次元相関ネットワークの基軸探索
    - 遺伝子要因X環境要因, 遺伝子要因X臨床表現型
- ゲノム・オミックス医療　第２世代へ
  - 多因子疾患のゲノム・オミックス医療へ
  - メタ・オミックス情報の収集
  - ３つのビッグデータ医療の統合デジタル医学へ
- データ科学準拠型医科学へ
  - 人材育成・大学医学教育体制の変革

ご清聴ありがとうございました