

ビッグデータ医療とAI創薬

東京医科歯科大学

東北大学 東北メディカル・メガバンク機構

田中 博



本日のトピック

- **ビッグデータ医療（ゲノム医療）**
 - 米国型のゲノム医療と欧州型のゲノム医療
 - メタオミックスの時代
 - ビッグデータ医療とAIによる知識発見
- **AI創薬**
 - 分子プロファイル型計算創薬・DR
 - Deep Learning を用いたAI創薬・DR

ビッグデータ医療（ゲノム医療）

ゲノム医療の2つの流れ

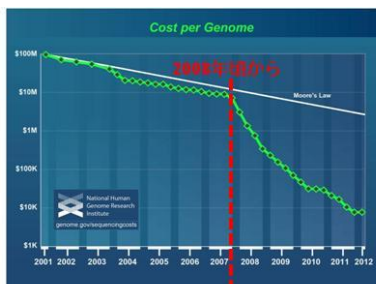
- 米国の流れ

- 次世代シーケンサの急激な発展による「シーケンス革命」からの怒濤の展開（2010から）
- 「治療医学」レベル質的向上のためにゲノム情報を取り入れた臨床実装の推進
 - 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

- 欧州の流れ

- 社会福祉国家の理念より国民医療の向上
- 「予防医学」レベル質的向上のためにゲノム情報を取り入れたバイオバンク推進
- 大規模前向きpopulation型バイオバンク/ゲノム・コホートの確立
 - 遺伝的素因だけでなく環境要因（生活習慣）との相互作用を解明し疾患発症を予測し、これに基づいて個別化予防する。
 - 疾患を発症前に対応して発症を防ぐ「先制医療(preemptive medicine)」や「予測医療(predictive medicine)の実現を目的

米国ゲノム医療の流れ



DNA Sequencing Cost: the National Human Genome Research Institute

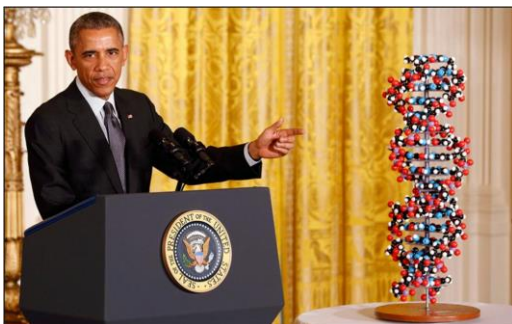
シーケンス革命 2007/8

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	ラビッドラン
解析時間	11日	27時間
リード長 (bp)	2 x 100	2 x 150
データ産出量 (Gb)	約600	約120
試薬コスト (ヒト1人全ゲノム)	数十万円	不可 エッセンスのみ

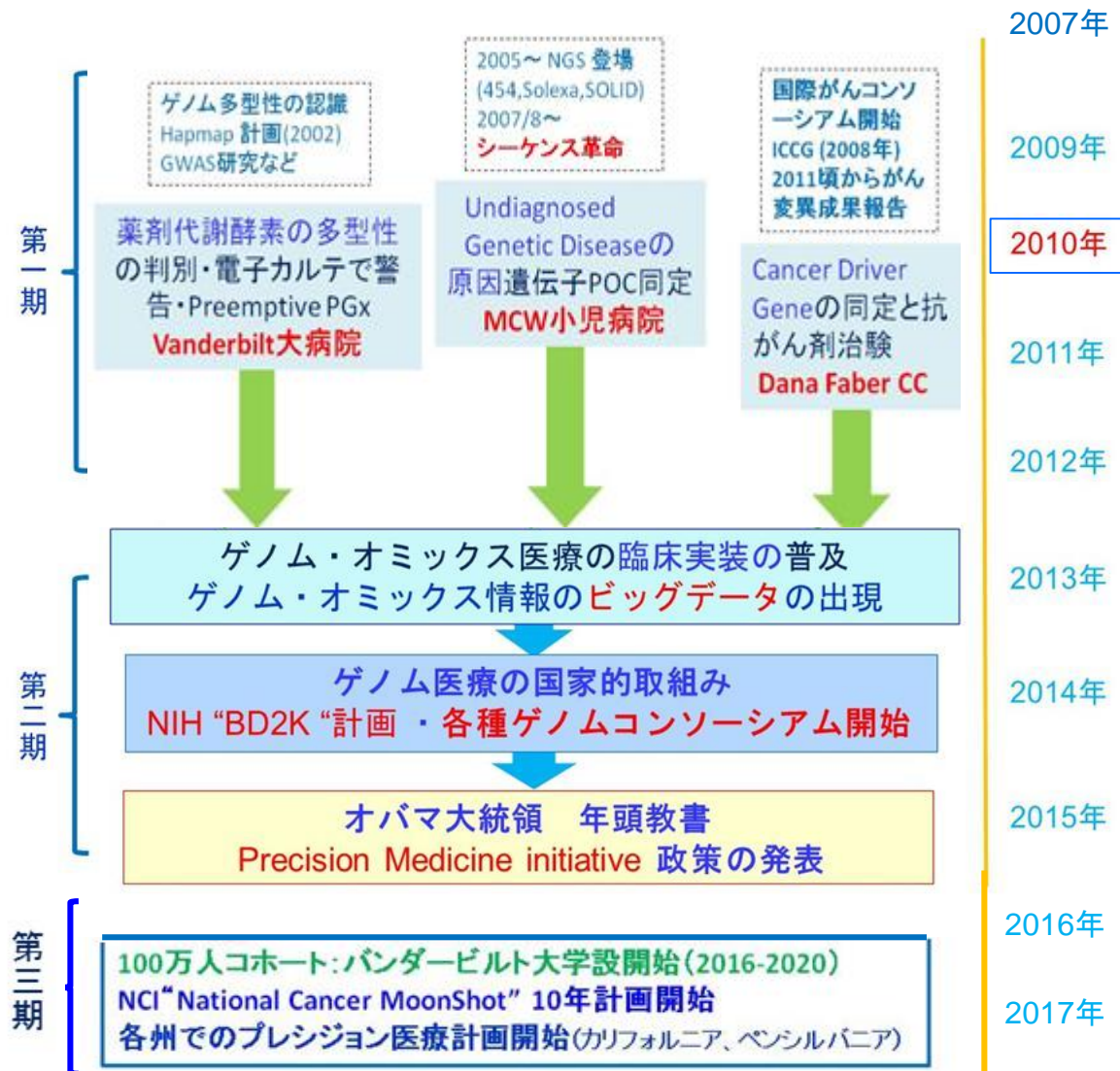
急速な高速化と廉価化
ヒトゲノム解読計画13年,3500億円
⇒1日,10万円



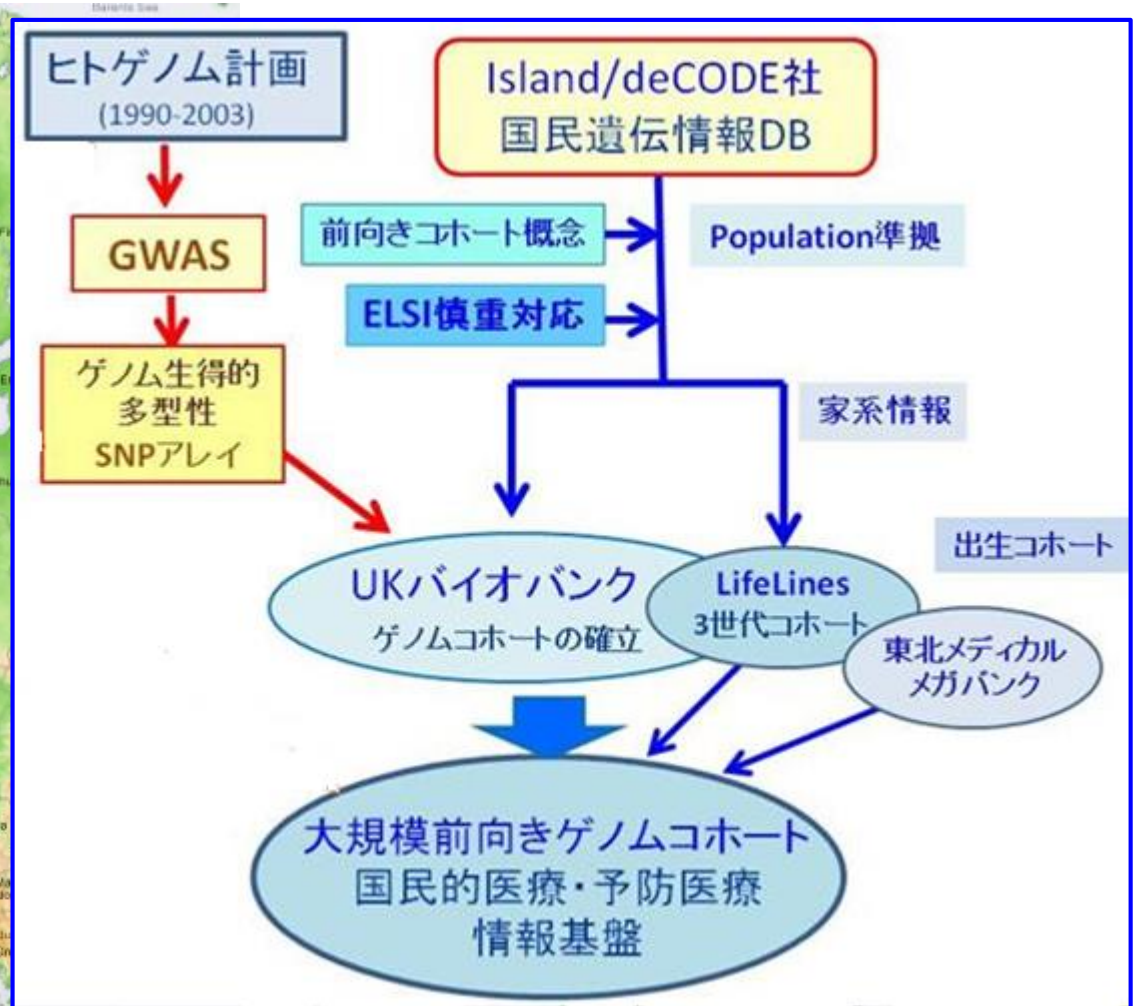
オバマ大統領 2015年1月 Precision Medicine Initiativeを開始
大統領一般年頭教書演説

先陣争いの時代

国家政策の時代

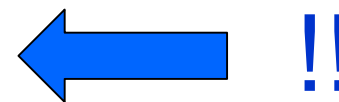


欧州のバイオバンクの流れ



医療ビッグデータ時代の到来

- (1) 次世代シーケンサなどによる「ゲノム/オミックス医療」
＜網羅的分子情報＞と＜臨床表現型情報＞の急速な蓄積
- (2) 「大規模 Biobank」による＜一般住民のゲノム情報＞と
＜環境・生活習慣情報＞の広範囲な蓄積

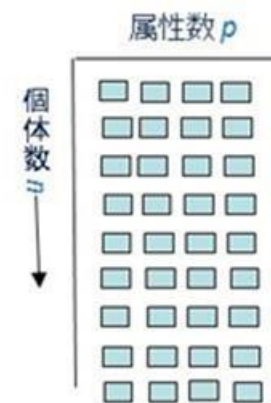


ゲノム：13年→1日(1/5000) 3500億→10万円(1/350万)

大量データの急激な
コストレス化かつ高精度化
医療ビッグデータの時代

ゲノム医療時代の ビッグデータ解析・人工知能

- ゲノム医療の2つの流れ
 - どちらにおいても超多次元相関ネットワークから「革新的知識」発見の必要性
- 治療医学：米国型
 - 〈網羅的分子情報と臨床表現型情報〉の相関ネットワークより知識発見
 - 分子画像やオミックス情報により複雑化
- 予防医学：欧州型
 - 〈遺伝的素因と環境・生活様式要因〉の相互作用と発症
- 医療ビッグデータ
 - 超多次元ネットワークから
 - 以下に「革新的知識」を獲得するか



(a)従来の医療情報のデータ形式



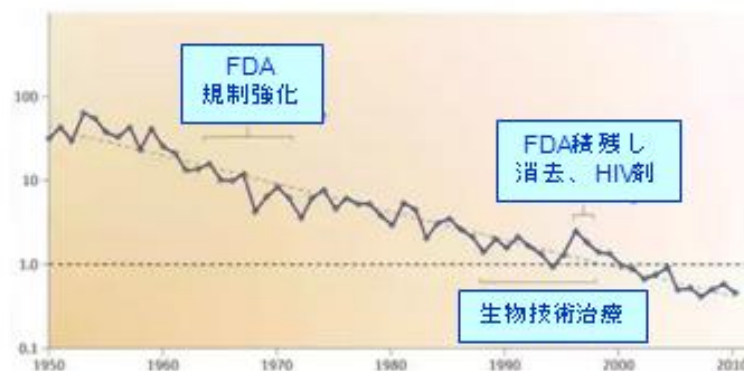
(b)新しいゲノム・オミックスのビッグデータ

ビッグデータ ・ AI創薬

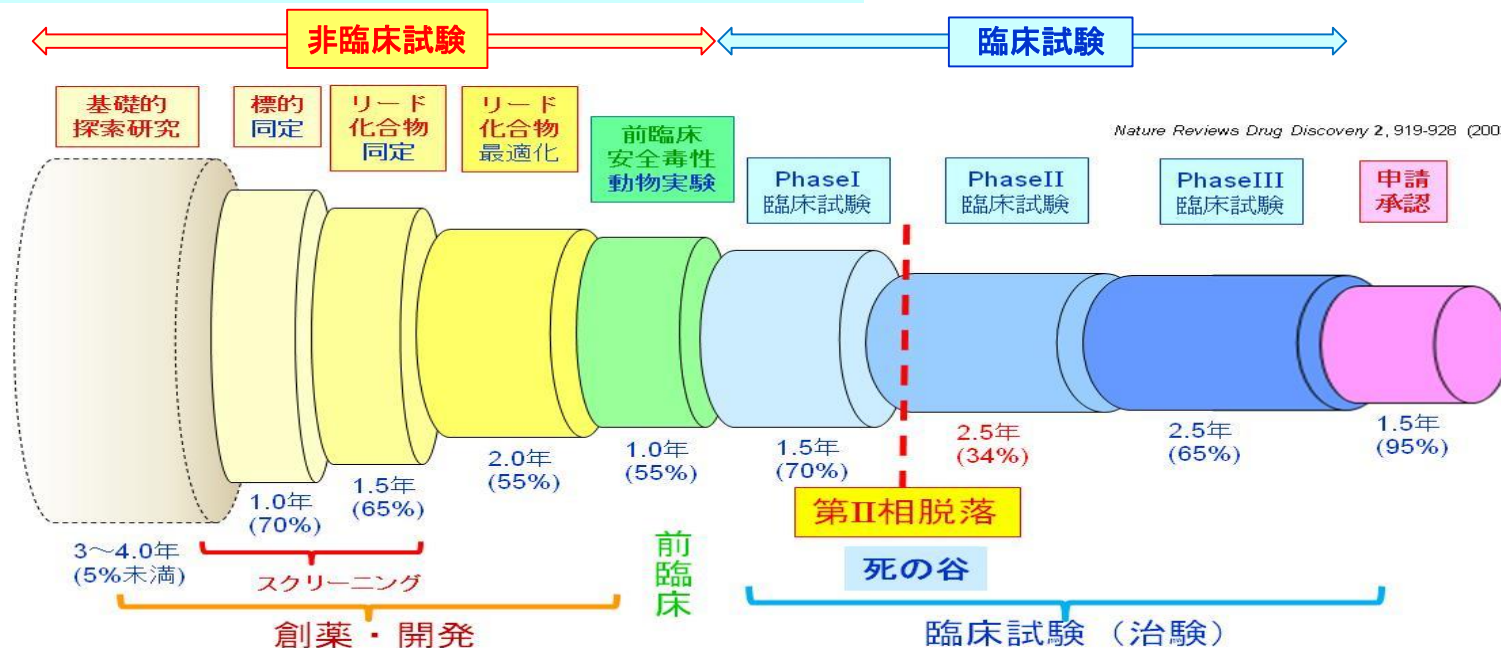
創薬をめぐる状況

- 医薬品の開発費の増大
 - 1 医薬品を上市するのに約1000億円以上
- 開発成功率の減少
 - 2万~3万分の1の成功率
 - とくに**非臨床試験**から**臨床試験**への間隙
 - **phase II attrition** (第2相脱落)
- 臨床的予測性
 - 医薬品開発過程の**できるだけ早い段階**での**有効性・毒性の予測**
- **臨床予測性の早期での実施**
 - 罹患者のiPS細胞を使う
 - ヒトの薬剤 - 生体関連のビッグデータを使う

10億ドル開発費で薬剤数



Nature Reviews Drug Discovery (2012)



Nature Reviews Drug Discovery 2, 919-928 (2003)

ドラッグ・リポジショニング

薬剤適応拡大

ヒトでの安全性と体内動態が十分に分かっている
既承認薬の標的分子や作用パスウェイなどを、体系的・論理的・網羅的に解析することにより**新しい薬理効果**を発見し、その薬を別の疾患治療薬として開発する創薬戦略

利 点

- (1) 既承認薬なので、ヒトでの安全性や体内動態などが既知で臨床試験で予想外の副作用や体内動態の問題により開発が失敗するリスクが少なく**開発の成功確率が高い**
- (2) 既にあるデータや技術（動物での安全性データや製剤のGMP製造技術など）を再利用することで、**開発にかかる時間とコストを大幅に削減できる**
- (3) **DR候補探索に疾患生命情報ビッグデータ知識DB**を使用できる。

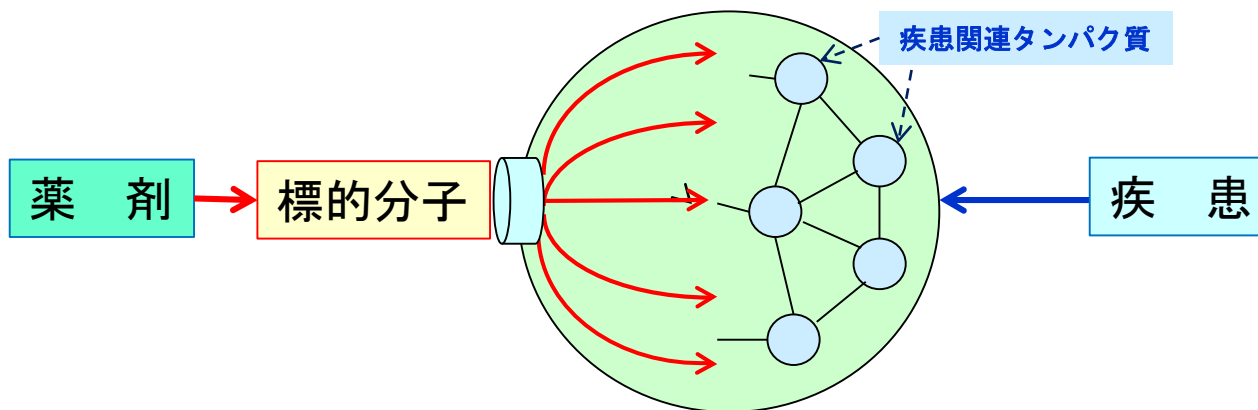
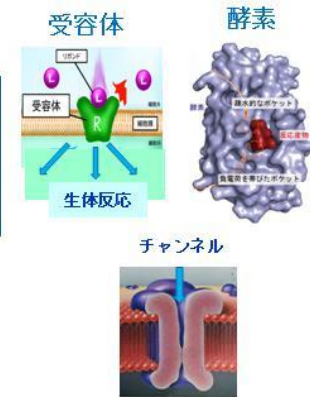
疾患・薬剤・標的の関係

病気の主要な要因

疾患要因タンパク質（複数）

薬：疾患要因タンパク質に影響を示す
標的タンパク質に作用し阻害する

薬剤の標的分子
受容体・酵素・チャンネルなど



生体システム/ネットワーク

ビッグデータ計算創薬 1

計算創薬(*in silico*創薬)の新しい方向

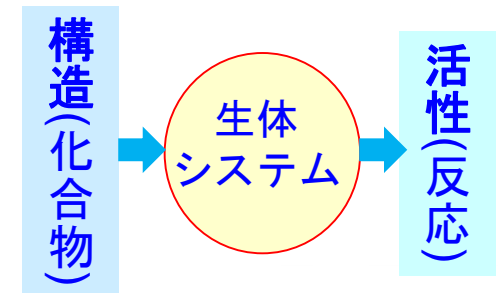
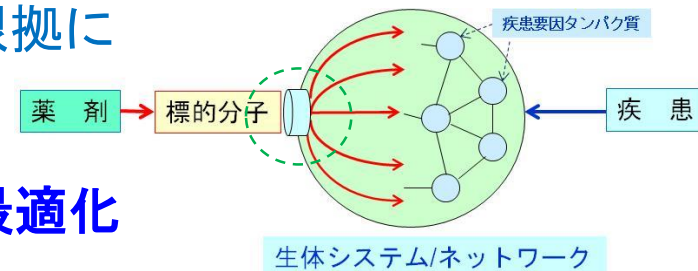
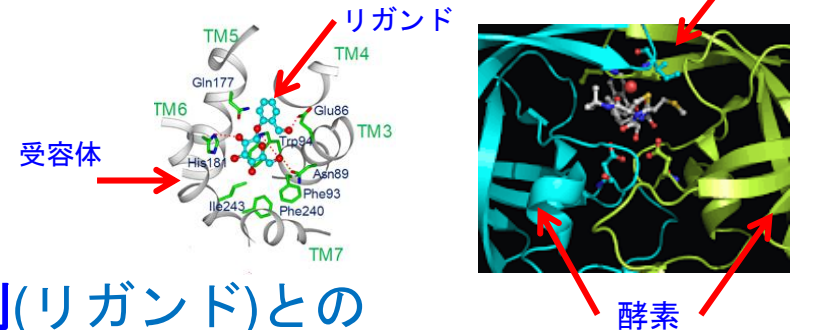
これまでの計算創薬

分子(結合構造)中心

- 分子構造解析・分子設計
- 標的分子(受容体・酵素)と薬剤(リガンド)との結合構造(ポケット)の分子構造を根拠に
- リガンドの分子設計(量子化学等)
 - 成功例: インフルエンザ薬 タミフル
- 標的に結合するリード化合物・構造最適化
- 結合後の生体システムの反応/振舞い
明確な取扱いがない

定量的構造活性相関(QSAR)

- 化合物の分子構造と生体活性の関係
- 両者の間には生体システムがある



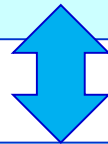
ビッグデータ計算創薬2

新しい計算論的創薬のアプローチ(網羅的分子プロファイル創薬)

疾患罹患状態における

疾患関連遺伝子(タンパク質)に起因し決定される
疾患時の生体のゲノムワイドな特異状態

疾患特異的な網羅的分子プロファイル変化

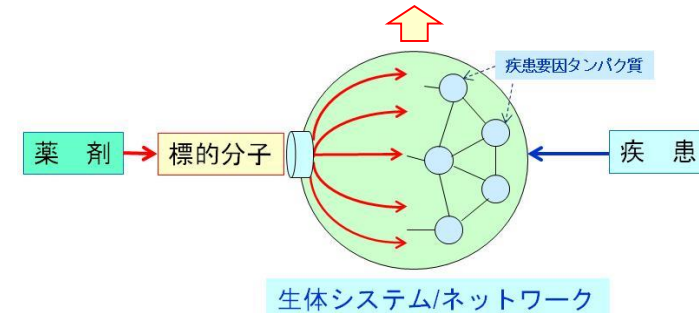
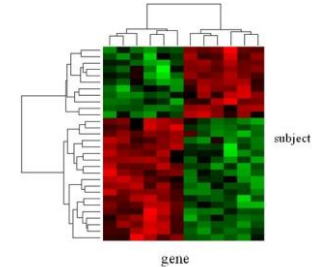


薬剤投与による

標的分子と薬剤分子の結合に起因し起こる
投与時の生体のゲノムワイドな反応/振舞い

薬剤特異的な網羅的分子プロファイル変化

遺伝子発現プロファイル変化
(疾患特異的/薬剤特異的)



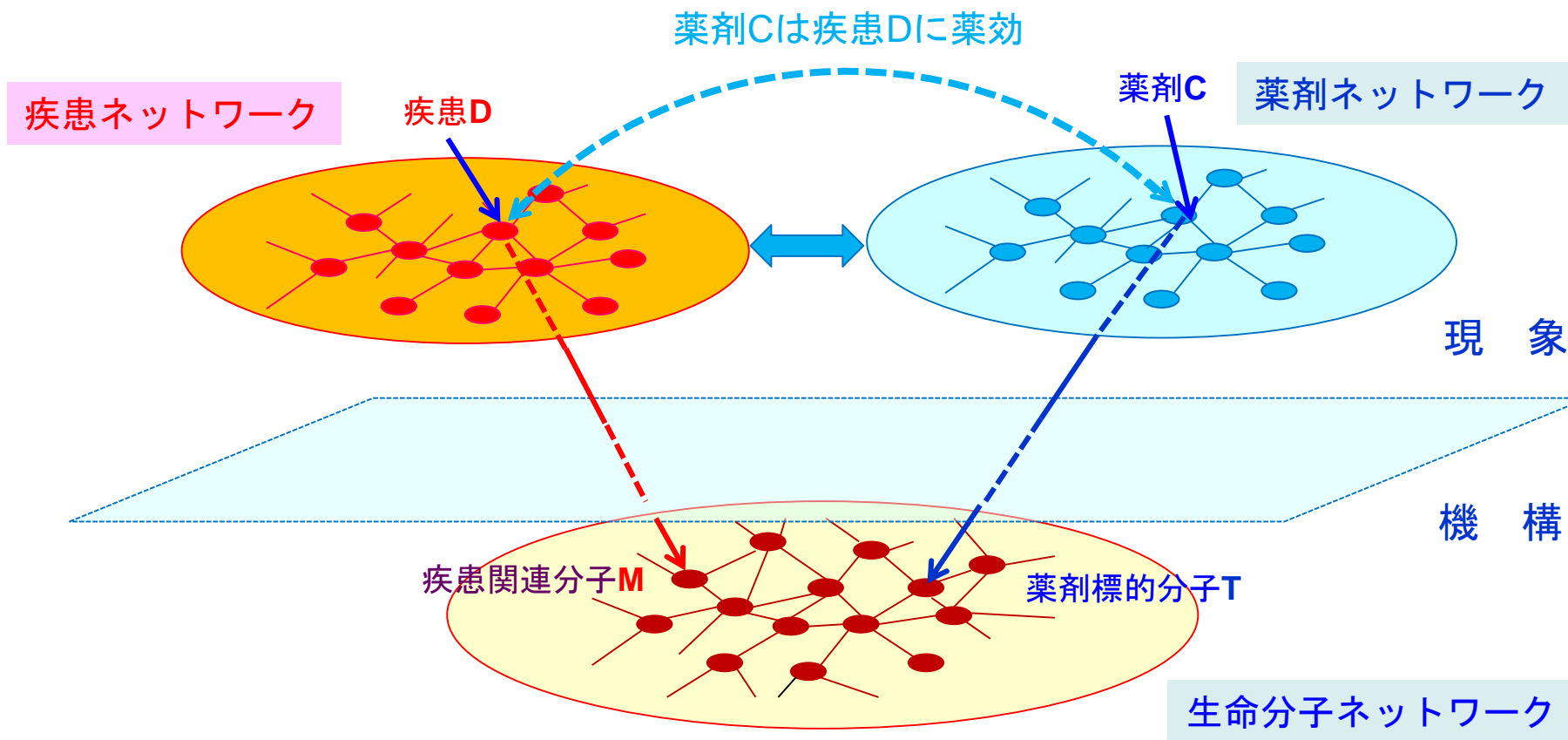
網羅的分子プロファイル⇒分子ネットワーク全体変化

＜疾患状態の生体＞に＜薬剤-標的分子の結合＞が引き起す作用によって
ゲノムワイドな生体分子環境がどう変化するか「生命システム観点からの理解」

化合物, 標的分子, 疾患間の関係の「ビッグデータ」DBを利用

Network型計算創薬・DRの基本的枠組み

3層の生体・薬剤のネットワーク間の関係図式



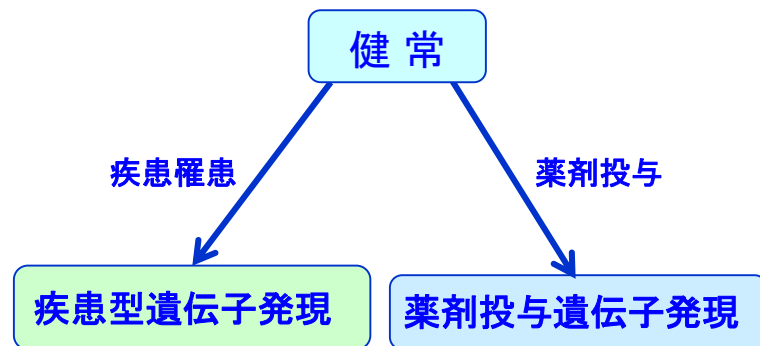
1. 遺伝子発現プロファイル比較型 創薬・DR

ビッグデータ計算創薬 発現プロファイル比較型創薬・DR

● 薬剤特異的遺伝子発現

– CMAP(Connectivity Map)

- 薬剤投与による遺伝子発現プロファイル変化
- 米国 ブロード研究所,1309化合物,
5種類のがんの培養細胞
約7000 遺伝子発現プロファイル
- シグネチャ (署名) 差別的発現遺伝子代表群
- DB利用：シグネチャを「問合せ」：
類似性の高い順に化合物を提示
- 最近はLINCSデータベース：100万種の薬剤特異的発現DBが存在



● 疾病特異的遺伝子発現

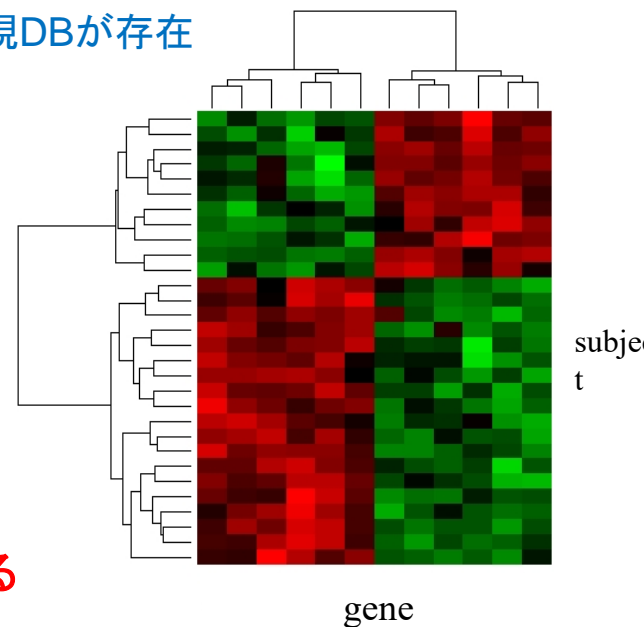
– GEO (Gene Expression Omnibus),

- 疾病罹患時の遺伝子発現プロファイルの変化
- 米国NCBI作成・運用 2万5千実験,
70万プロファイル (欧州 ArrayExpress)
- もEBIが作成、サンプル数同程度

基礎には分子ネットワークの疾病/薬剤特異的変化

遺伝子発現プロファイル変化

≈ 分子ネットワーク活性構造変化を反映する



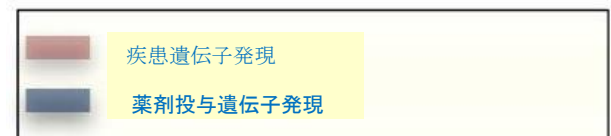
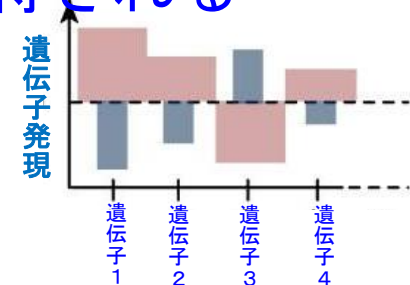
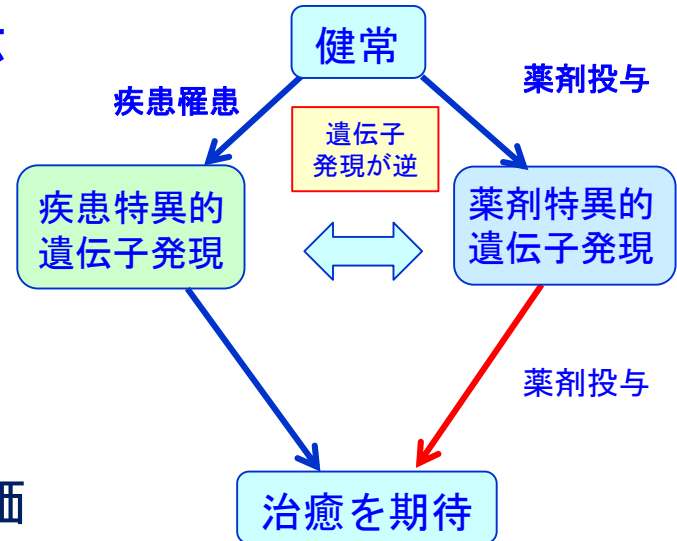
遺伝子発現プロファイルによる有効性予測

● 遺伝子発現シグネチャ逆位法

- 疾患によって**健常状態から変異**
「疾患特異的遺伝子発現プロファイル」
- これに**薬剤投与の変化を起こす**
「薬剤特異的遺伝子発現プロファイル」
- **両者のパターンが負に相関する**
- **ノンパラメトリックな相関尺度で評価**

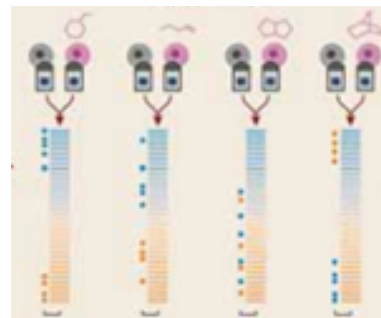
● 効果がお互いに打ち消すなら**有効性が期待される**

- 例：炎症性腸疾患に抗痙攣剤(topiramate),
骨格筋委縮にウルソール酸



ベースは疾患遺伝子発現
横の点は薬剤遺伝子発現

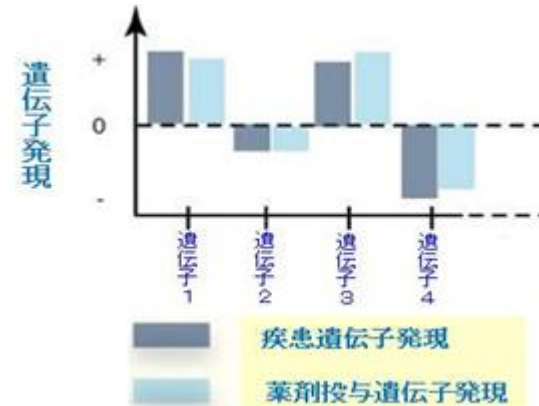
青は発現が**上昇**した遺伝子
赤は発現が**下降**した遺伝子



強正 弱正 弱負 強負

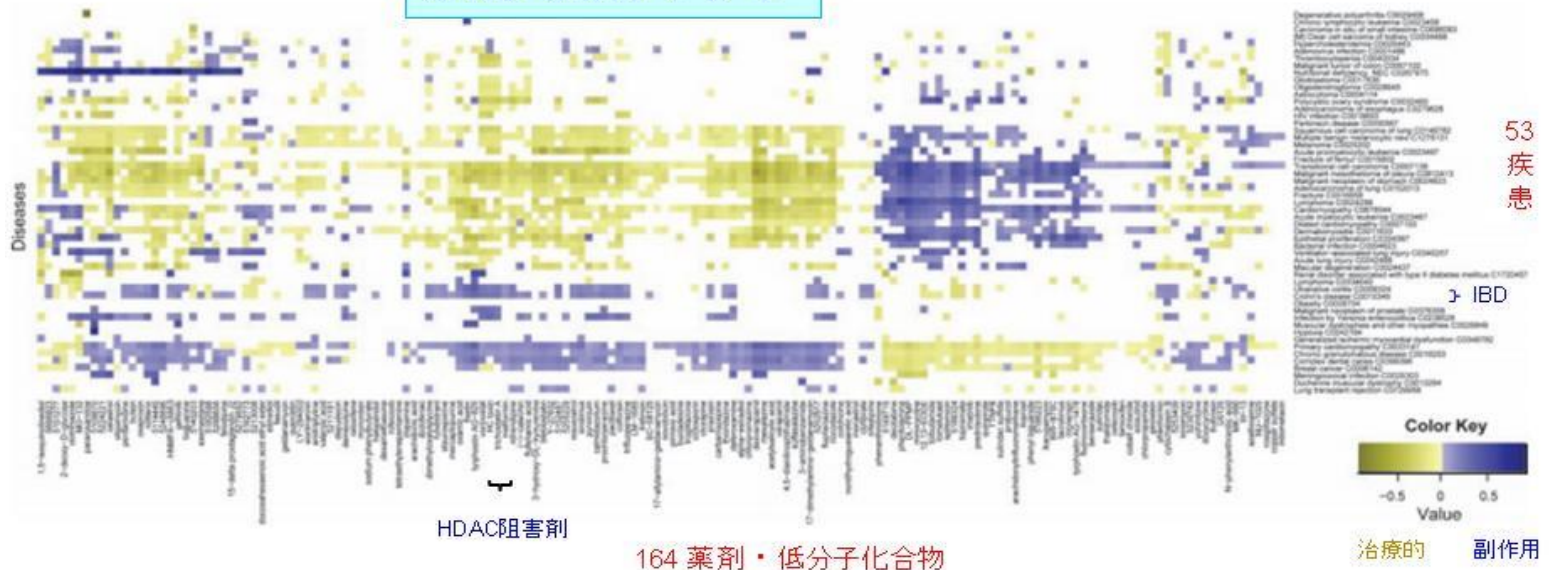
遺伝子発現プロファイルによる毒性予測

- 連座法 *guilt-by-association* :
- 薬剤-疾患間 副作用予測
 - 薬剤特異遺伝子発現プロファイルと
 - 疾患特異的遺伝子発現プロファイルが
 - ノンパラメトリック正に相関
 - 毒性・副作用の予測



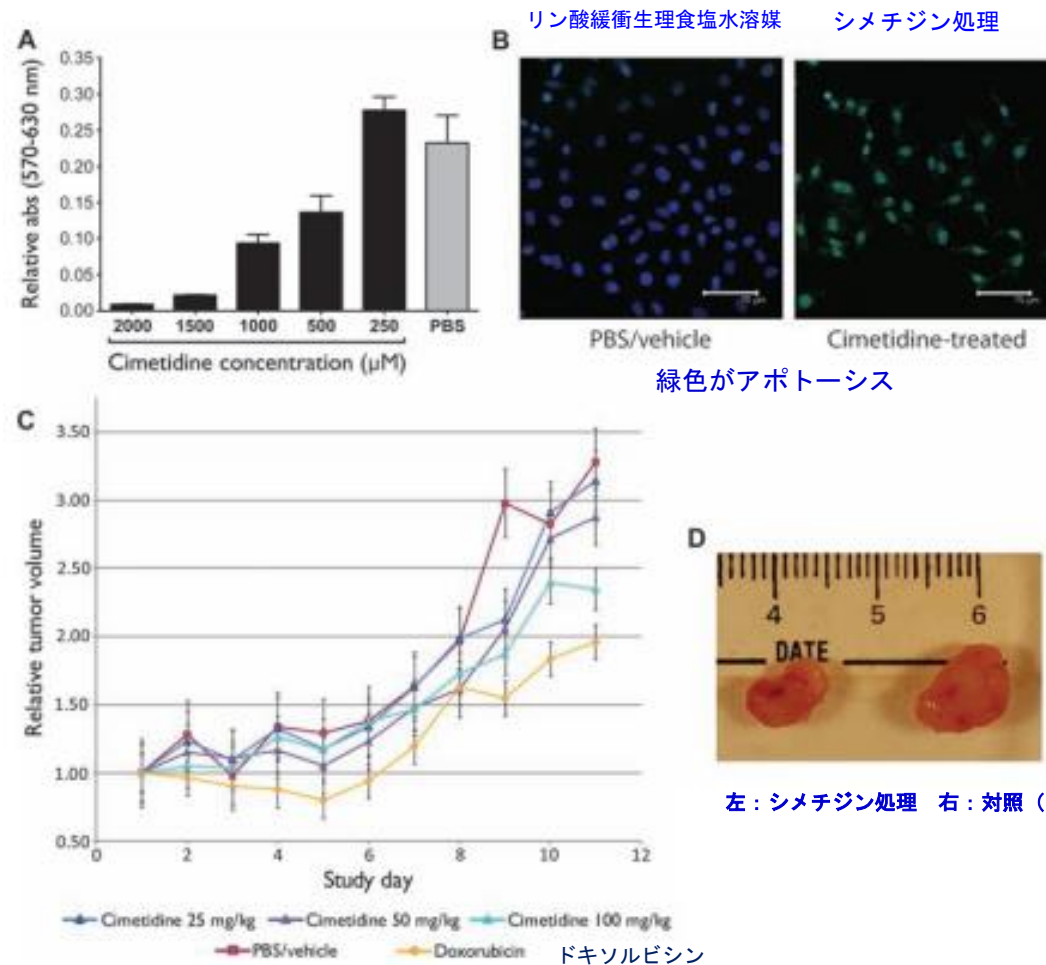
疾患-薬剤マップ

(Sirota, Butte 2011)



動物実験での実証

シメチジン(cimetidine:ヒスタミンH2受容体拮抗薬) →肺腺癌(LA)に有効か
 予測スコア -0.088 であったが gefinitib の-0.075より高い



遺伝子発現Profiling による疾患－薬剤ネットワーク (Hu, Agarwal)

遺伝子発現プロファイル(c-Map)での相関係数、ES指標によりネットワーク表示

疾患－疾患、薬剤－薬剤、疾患－薬剤のネットワークを発現プロファイルより構成

- 疾患 - 疾患 (disease-disease) 645 組
- 疾患-薬 (disease-drug) 5008 組
- 薬 - 薬 (drug-drug) 164,374 組

結果

①疾患関連の60%はMeSH (既知体系)
 その他は分子レベル疾患分類学
 Transcriptomeの類似性による疾患体系

②主な発見

<疾患 - 疾患>

HSP (Hereditary Spastic Paraplegia
 (遺伝性痙攣性対麻痺)

⇒bipolar 双極性障害 --精神障害も

Solar keratosis 日光性角化症

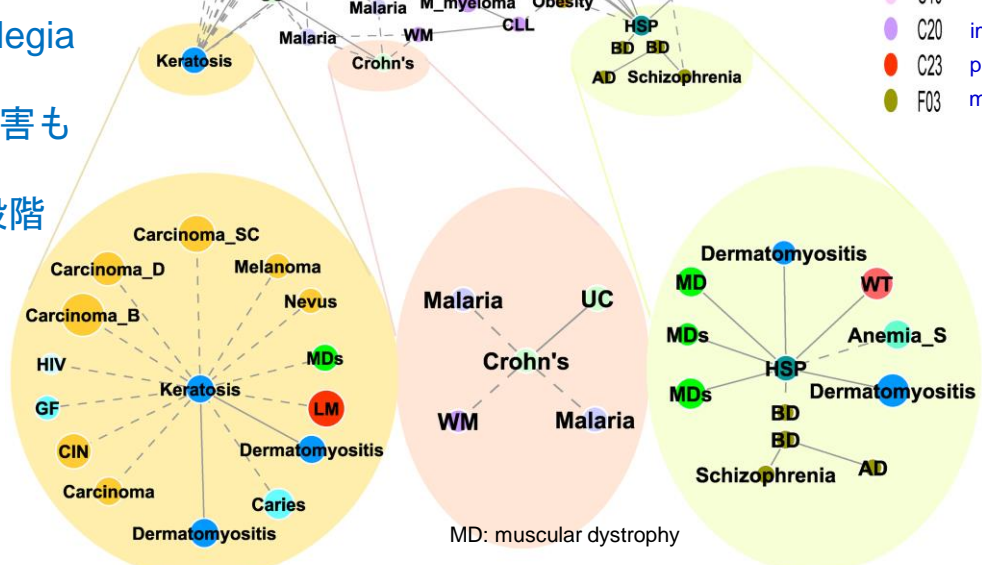
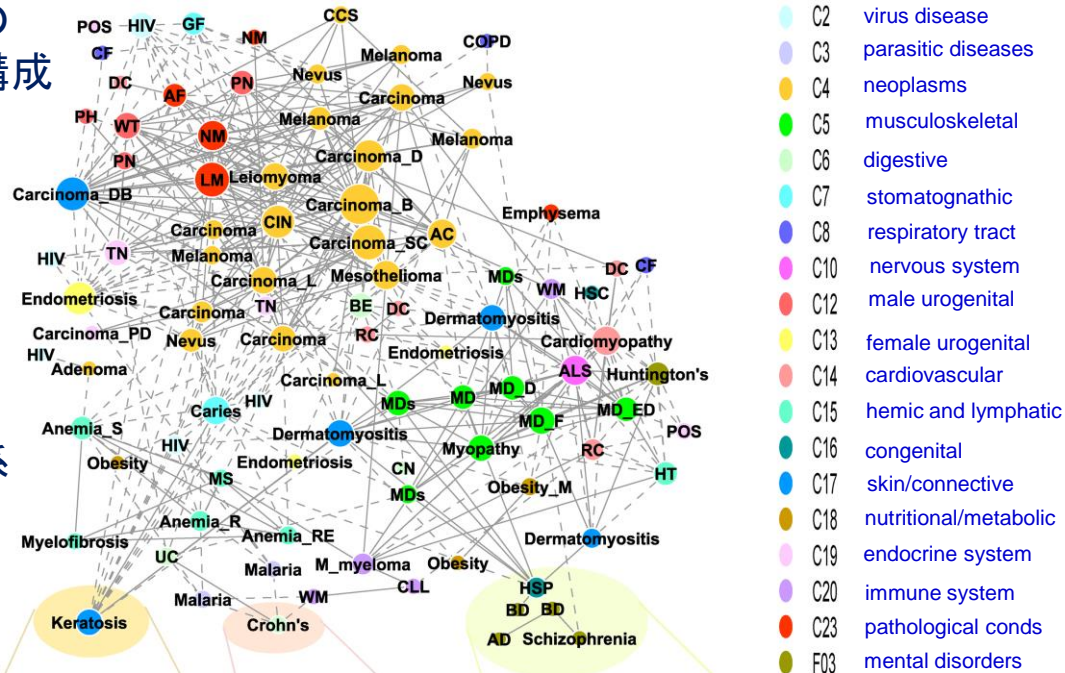
⇒ cancer(squamous) --前癌段階

<疾患 - 薬>

有効性：マラリア治療薬

⇒ Crohn's disease

ハンチントン病に種々の薬剤



カラーはMeSH
 同一カテゴリー
 実線はMeSH内
 破線はMeSH外

MD: muscular dystrophy

2. 疾患ネットワーク創薬/DR

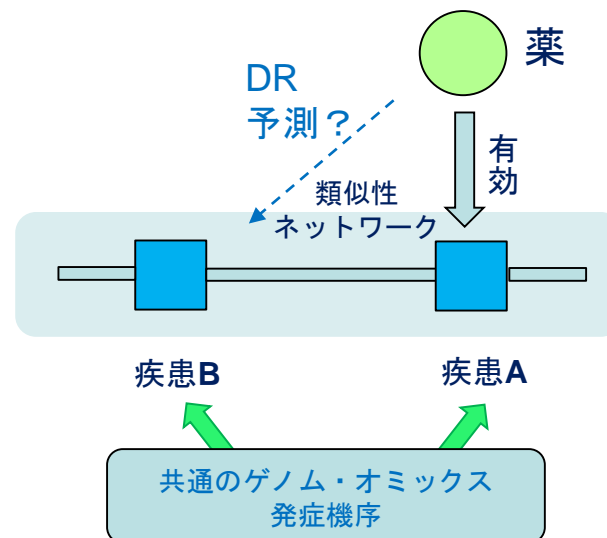
疾患ネットワーク空間を基礎にした
ビッグデータ創薬/DR

＜疾患ネットワークでの近接性＞

ビッグデータ創薬/DRの基本原理2

疾患ネットワーク準拠創薬/DR

- 従来の疾患体系 nosology
 - Linne以降300年に亘って表現型による疾病分類
 - 臓器別・病理形態学別の疾患分類学
- ゲノム・オミックスレベルでの発症機構での疾患分類
 - 発症の**内在的 (intrinsic)機構の類似性**を**基準に**疾患ネットワーク（疾患マップ）をつくる
 - ゲノム・オミックスによる内在的疾患機序の概念が基礎

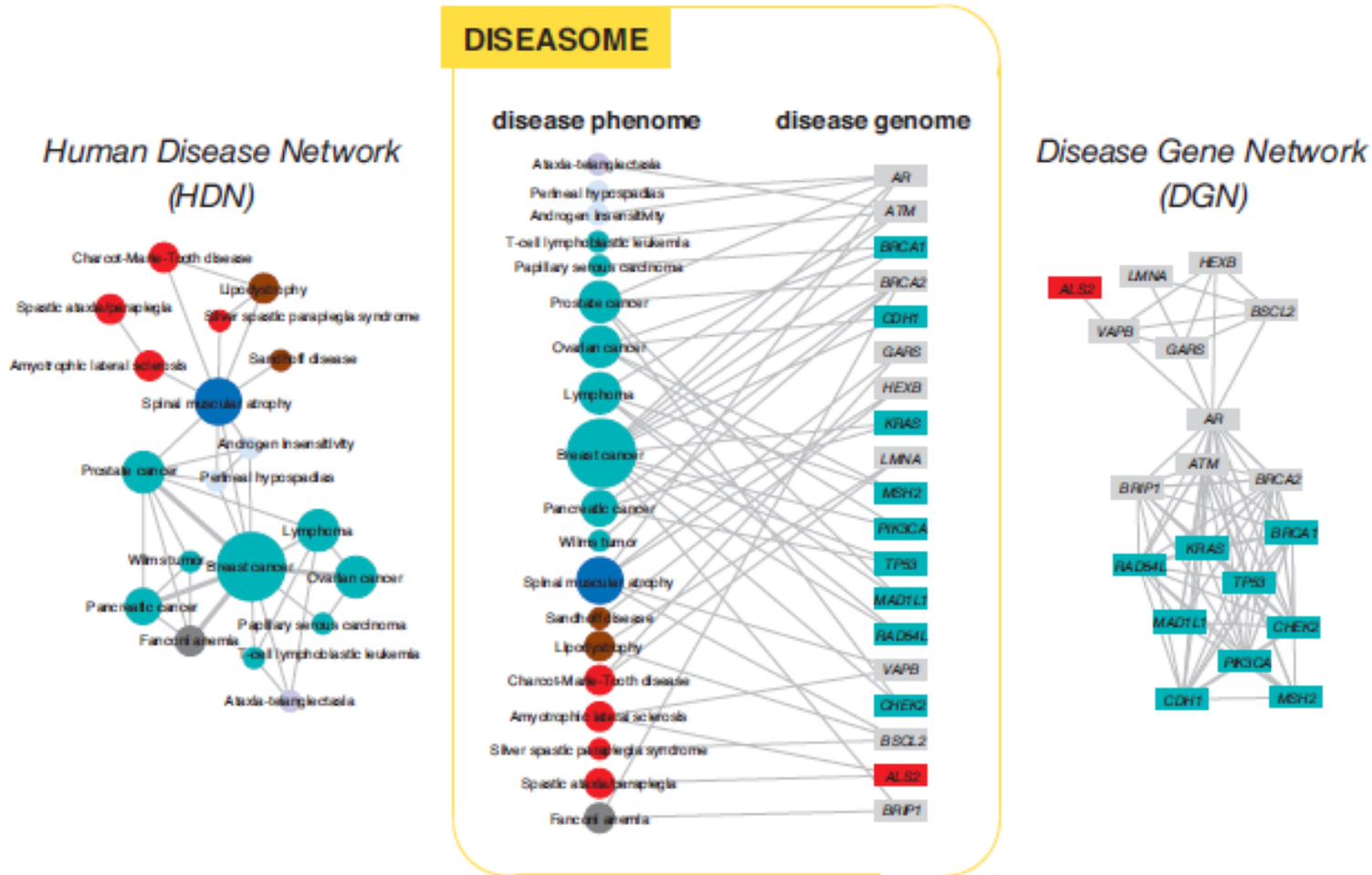


第1世代型

Diseasomeと疾患遺伝子

- **OMIM**から 1,284 疾患と 1,777 疾患遺伝子を抽出
- **ヒト疾患ネットワーク (HDN)**
 - 867疾患は他疾患へリンクを持つ 細胞型や器官に非依存
 - 516疾患が巨大クラスターを形成
 - 大腸がん、乳がんがハブ形成
 - がんはP53 やPTENなどにより最結合疾患 がんなどは後天的変異
 - 疾患を網羅的に見る見方：臓器や病理形態学に非依存
 - リンネ（12疾患群分類）以来300年続いた分類学を越える
- **疾患遺伝子ネットワーク (DGN)**
 - 1377遺伝子は他の遺伝子へ結合
 - 903遺伝子が巨大クラスター
 - P53がハブ
- ランダム化した疾患/遺伝子ネットワークに比べ
 - 巨大クラスターのサイズが有意に小さい
- **疾患遺伝子は機能的なモジュール構造**
 - 同じモジュールに属する遺伝子は相互作用し
 - 同一の組織で共発現し、同じ**GO**（遺伝子オントロジー）を持つ

疾患ネットワーク Diseasome (Goh, Barabasi et al.)



1つ以上の疾患関連遺伝子を共有する疾患

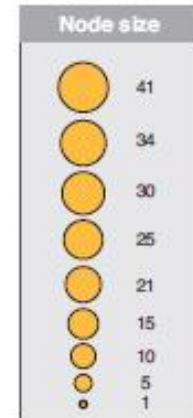
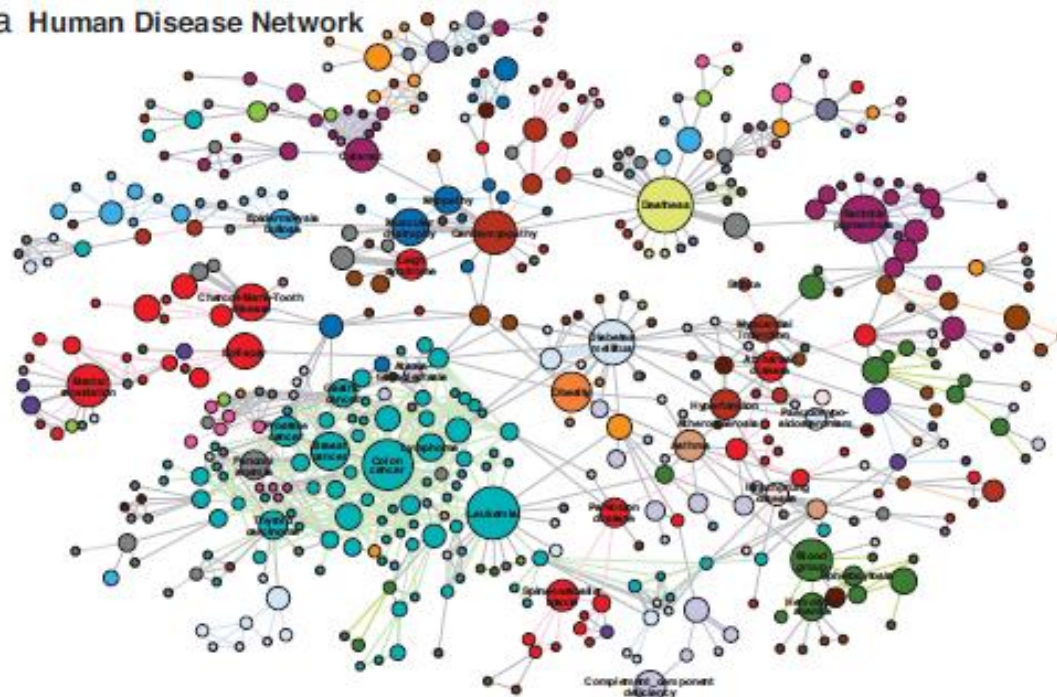
1つ以上の疾患を共有する疾患関連遺伝子

Kwang-Il Goh*, Michael E. Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-Laszlo Barabasi The human disease network PNAS2007

疾患ネットワーク (HDN)

Nodeの直径
 疾患に関与している原因遺伝子の数に比例
リンクの太さ
 疾患間で共有している原因遺伝子の数

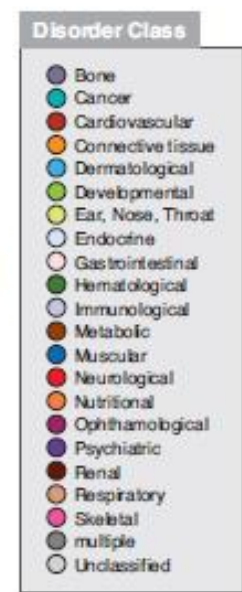
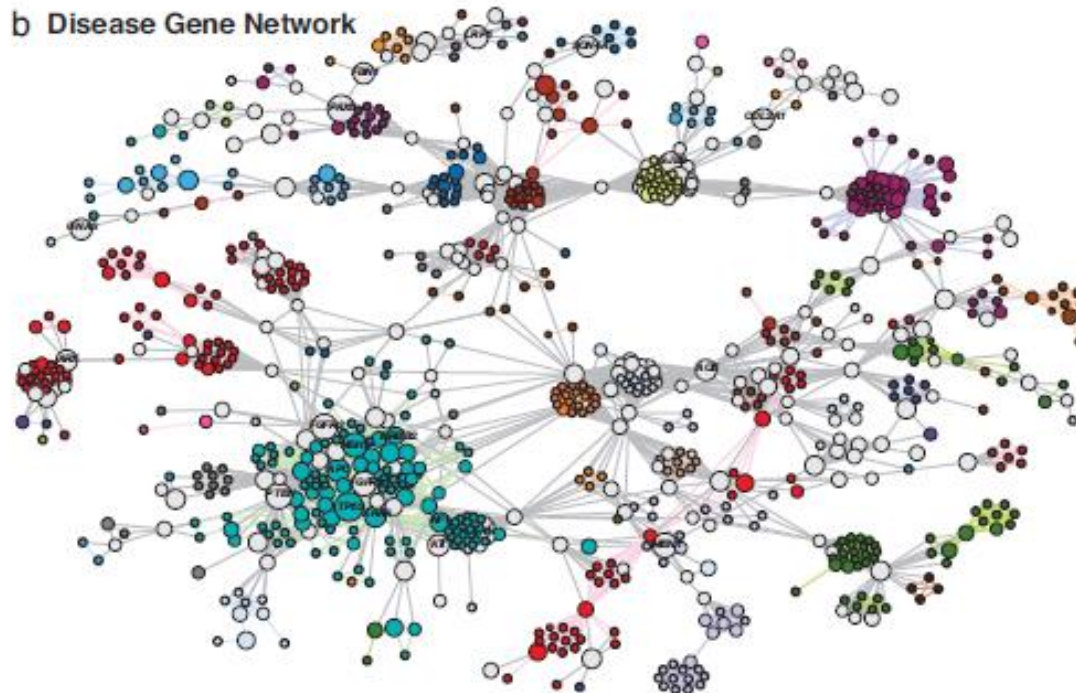
a Human Disease Network



疾患遺伝子ネットワーク (DGN)

Nodeの径
 その遺伝子を原因にしている疾患の数に比例
 2つ以上の疾患に関与していると明灰色の遺伝子ノード

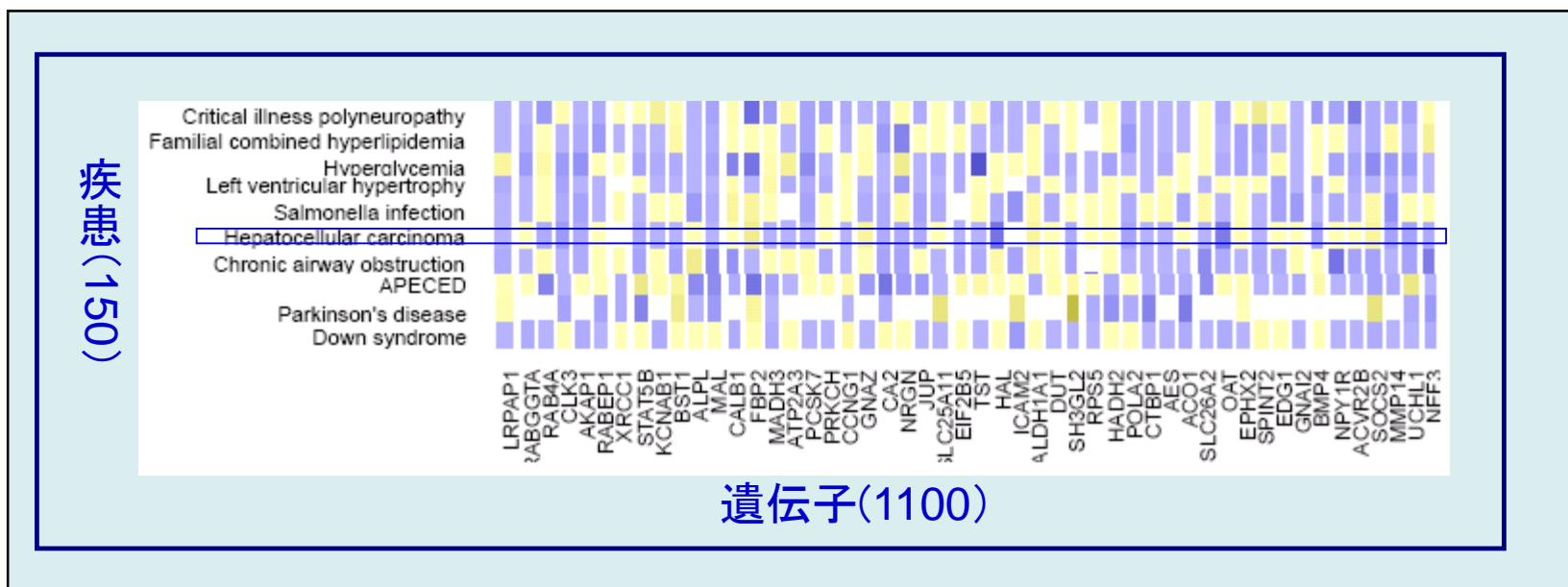
b Disease Gene Network



第2世代型

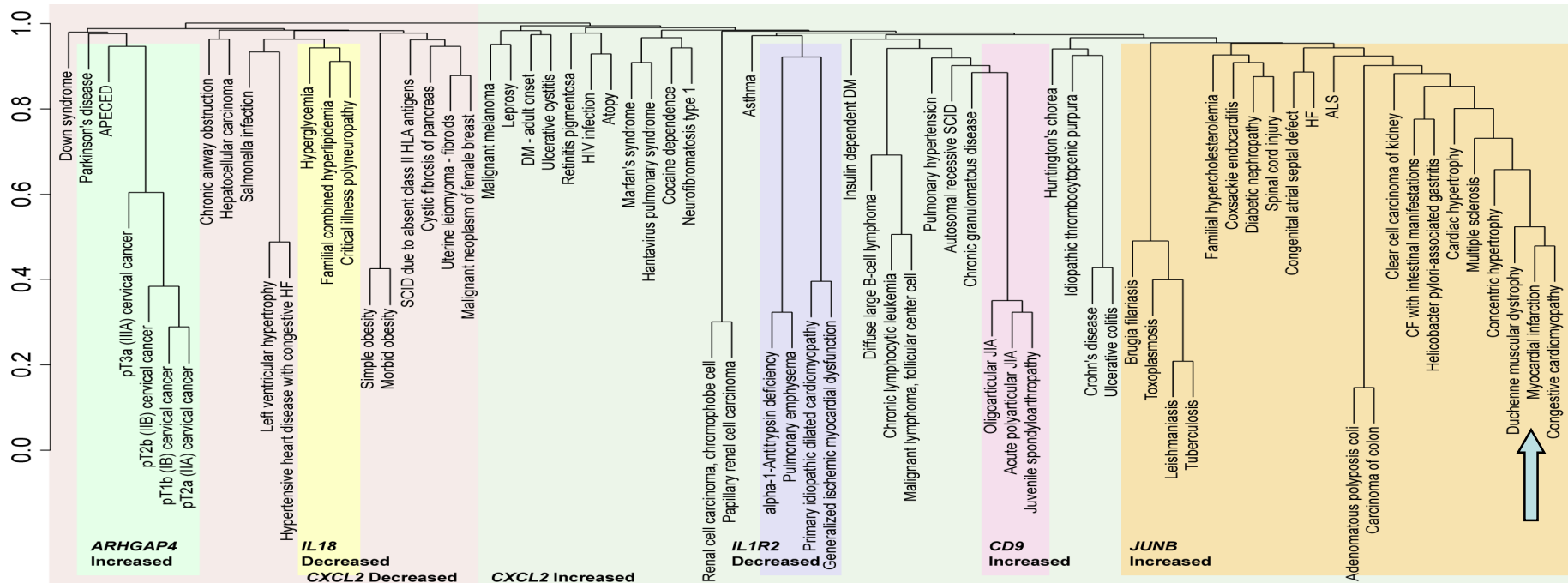
GENOMED (A. Butte et al)

- 遺伝子発現DBのGEO (Gene Expression Omnibus) 利用
 - 約20万のサンプル
- 疾患名は注釈文より用語集UMLSを用いて抽出
- 疾患ごとに多数の遺伝子発現パターンを平均化



Gene-Expression Nosology of Medicine

- 疾患を平均遺伝子発現パターンよりクラスター分類
 - 臓器別疾患分類では予想できない疾患間の親近性
 - 分類項目はサイトカインの遺伝子発現と相関
 - 疾患の再体系化に基づいた医薬の repositioning
- さらに656種類の臨床検査を結合した分析
- 心筋梗塞・デュシャンヌ型筋ジストロフィーに近い



＜疾患-薬剤-標的分子＞の
多階層ネットワークによる
ビッグデータ創薬/DRの枠組み

3層の生体・薬剤のネットワーク間の関係図式

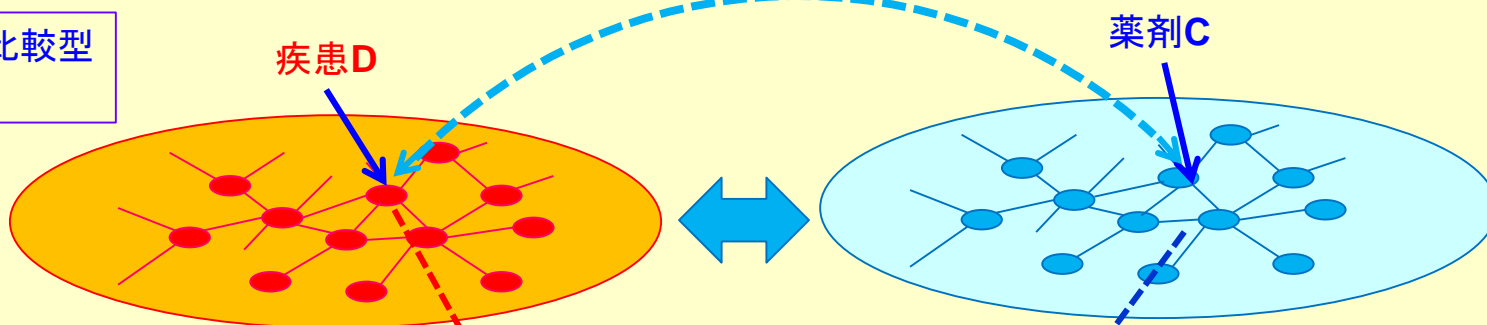
現象的マクロ的対応

薬剤Cは疾患Dに薬効

薬剤ネットワーク

疾患ネットワーク

プロファイル比較型
創薬/DR



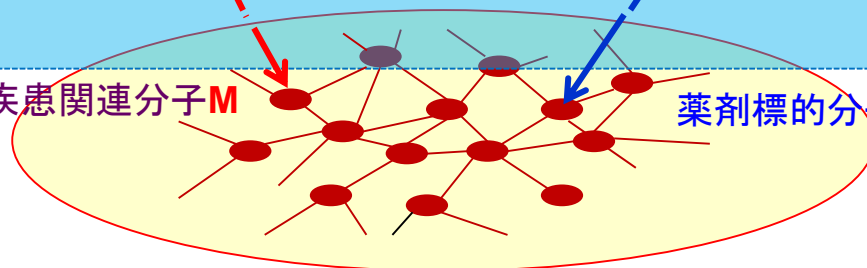
分子ネットワーク型
創薬/DR

疾患関連分子M

薬剤標的分子T

機構

生命システム



3層の生体・薬剤のネットワーク間の関係図式

薬剤ネットワーク

薬剤Cは疾患Dに薬効

現象

機構

生命システム



疾患ネットワーク

プロファイル比較型
創薬/DR

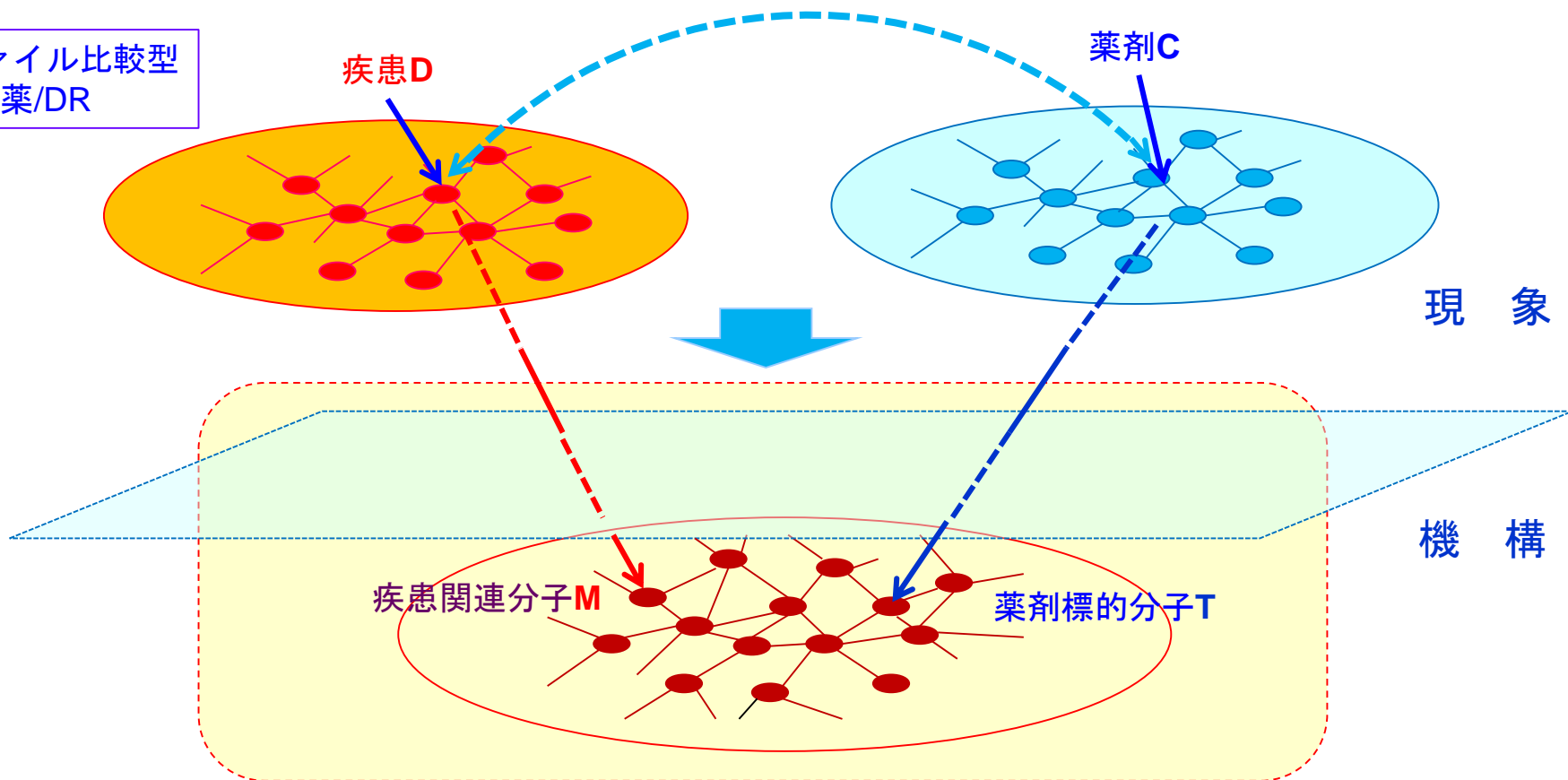
疾患D

薬剤C

疾患関連分子M

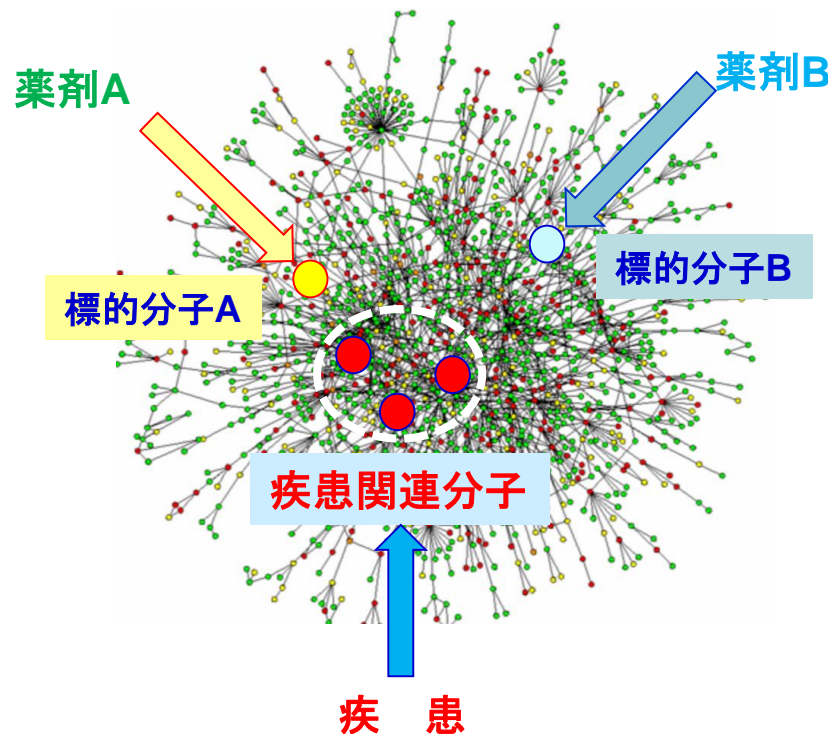
薬剤標的分子T

分子ネットワーク型
創薬/DR



標的分子や疾患要因分子の タンパク質相互作用ネットワーク (PPIN)

- 薬剤ネットワークと疾患ネットワークを媒介する第3の生体ネットワーク
- タンパク質相互作用ネットワーク (PPIN) での創薬/DR戦略
- PPIネットワーク場を基礎にして距離 (類似性) を検討
- **薬 剤** : 薬剤の**標的分子** (タンパク質) によって PPI場と繋がる
- **疾 患** : 疾患特異的発現遺伝子を**疾患要因分子** (タンパク質) へ翻訳、
- PPIN場内での**薬剤標的分子**と**疾患**の「**代理人(疾患遺伝子)**」の**距離・親近性**を基準に、**薬理作用のインパクト**力を評価



タンパク質相互作用
ネットワーク (PPIN)

PPIの基づくDR（肺腺癌の例）

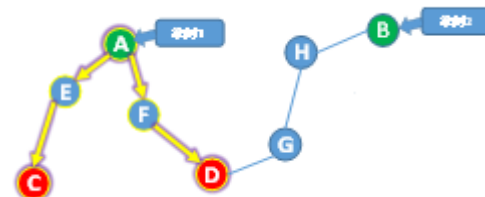
- **Interactome**(タンパク質相互作用)ネットワーク (Sun, 2016)

- **HPRD** (Human Protein Reference Database)

- 37,070 PPI, 9465 タンパク質

- **STRING** (Search Tool for the Retrieval of INteracting Genes/proteins)

- 184 M PPI, 9,643,763タンパク質 --- 個々に計算



- **薬剤⇒標的分子** : **DrugBank**

- 7,759 薬剤、4300タンパク質

- 12,604 の薬剤-標的分子組 (4,452薬剤, 1,617タンパク質)

- **疾患遺伝子の差別的遺伝子発現データ (DEG)**

- **TCGA** (The Cancer Genome Atlas)より差別的発現遺伝子を同定

- 445 肺腺癌例, 19 正常例, 疾患遺伝子 FC >2.0 or <0.5, FDR<0.01, **927** 差別的発現遺伝子

- **薬剤の疾患遺伝子への影響力 評価IPS** (Impact power score)

- **薬剤の標的分子と疾患遺伝子の間のネットワーク距離の総合評価**

- 「再出発ありランダム歩行RWR」でネットワーク距離を評価

- 標的分子からランダム歩行を繰り返す (出発点から再出発あり)

- s時点後, 疾患遺伝子のノードにどれだけの確率で滞在しているかを**IPS**とする

- 一定の時間が過ぎると、定常状態になり、歩行で滞在確率分布は変化しない。

- 定常状態での疾患遺伝子ノードに滞在している確率の総和が薬剤の評価になる

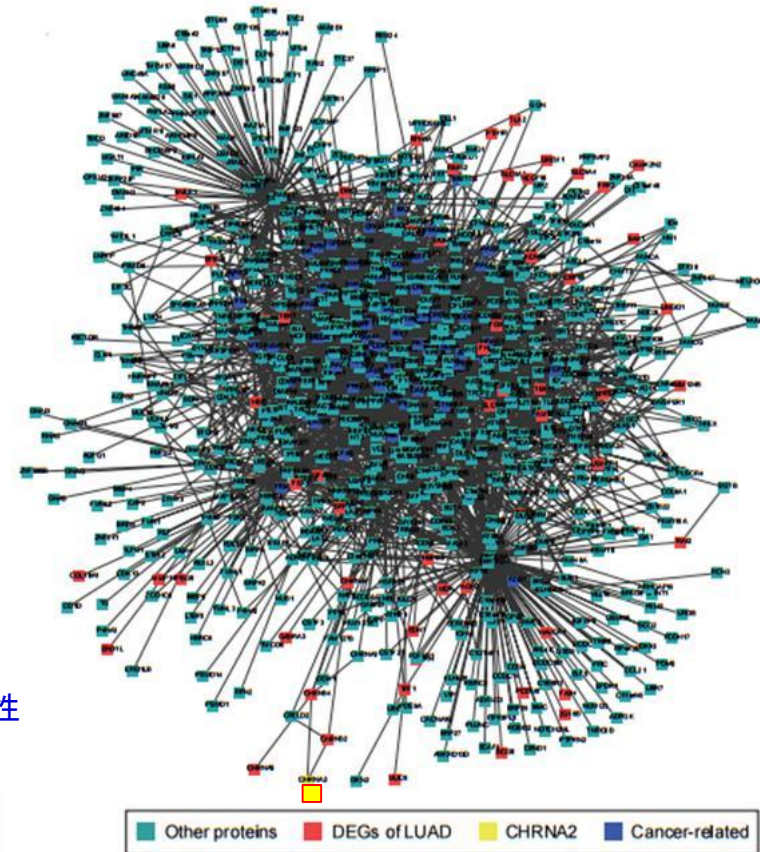
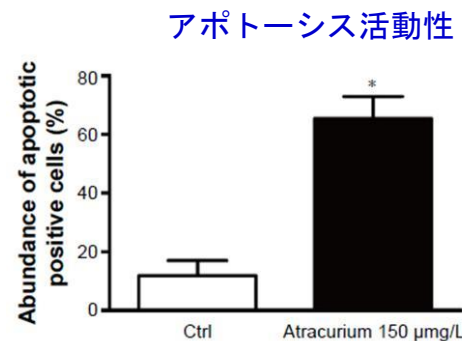
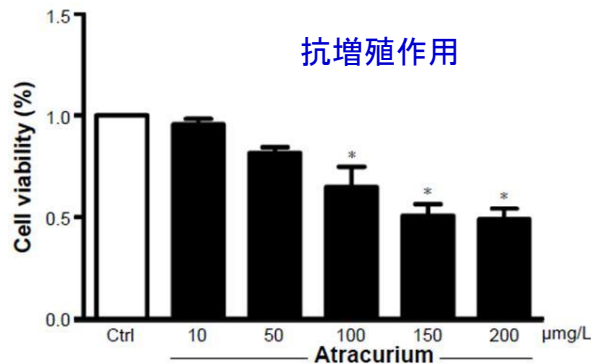
$$\mathbf{P}^{s+1} = (1-\gamma)\mathbf{M}\mathbf{P}^s + \gamma\mathbf{P}^0$$

\mathbf{P}^s : 時点sでの各ノードでの滞在確率 \mathbf{M} : 各ノードへの遷移確率 γ : 再出発確率

Interactome DR 結果の検証

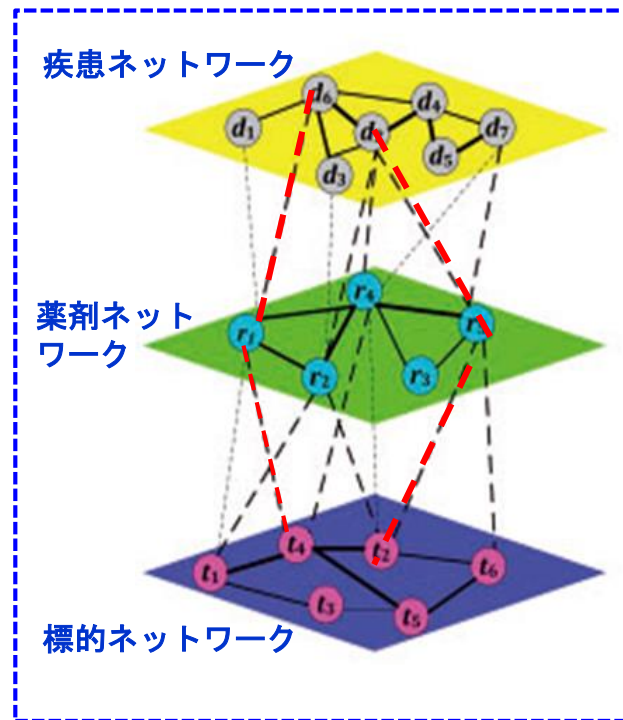
Drug ID	Drug name	Target	Score	Rank
DB00416	Metocurine Iodide	CHRNA2	0.966581	1
DB00565	Cisatracurium besylate	CHRNA2	0.966581	1
DB00732	Atracurium	CHRNA2	0.966581	1
DB00657	Mecamylamine	CHRNA2	0.966581	1
DB02457	Undecyl-phosphinic acid butyl ester	LIPF	0.953846	5

- HPRDとSTRINGSの両方のPPINのランダム歩行でtop5%で共通な145薬剤を同定
- 最高スコアを挙げたAtractiumを選択
- 薬剤標的はCHRNA2(Cholinergic Receptor Nicotinic Alpha 2)でアポトーシス経路である
- 培養細胞A549 (ヒト肺胞基底上皮腺癌細胞) の抗増殖作用を確認



3階層生命ネットワークでの創薬/DR

- 3階層の生体ネットワーク
 - 疾患ネットワーク：網羅的分子による内在的機序
 - 薬剤ネットワーク：化学構造によってネットワーク
 - 標的ネットワーク：薬剤と標的（DrugBank参照）
- 各層のネットワーク内結合
 - 稠密に自己完結的に構築可能
- 各層ネットワーク間のリンク
 - 成功した<疾患-薬剤>の事実の根拠のみ
 - 階層間はスパースな結合である



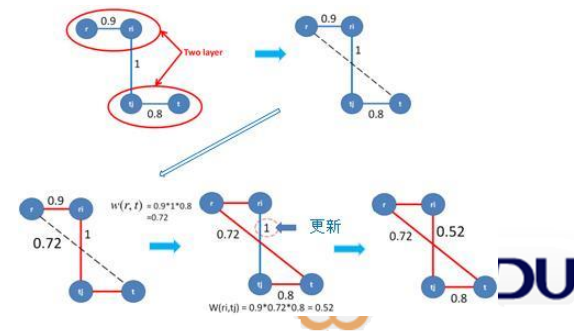
(Wang et al. 2014)

創薬/DRとは

未発見の階層間リンクを
既存の階層間リンクの事実と
各層のネットワークから推測

Wang et al. 2014は

- 階層間リンク（事実）と各階層内のリンクより階層間のリンクの強さを計算する方法を提案している



人工知能（AI）と医療・創薬

医療分野の人工知能の歴史

記号（シンボル）的知識処理

ニューロネットワーク処理

1970

問題解決の一般探索手法 GPS
解決木の高速探索（ゲーム）

ニューロネットワーク
3層の学習機械 Perceptron
入力層、隠れ層、出力層

1980

推論システム（if-thenルールシステム）
知識の表現と利用（専門家システム）
医療診断システム（Mycin, Internist-I）
大ブーム 医療から産業応用の期待波及

多層型ニューロネット
後方伝播 Back Propagation
結合係数修正アルゴリズム

1990

期待消滅！

知識発見 機械学習
Machine Learning, KDD
診断知識のDBからの学習

しばらく停滞！

2000

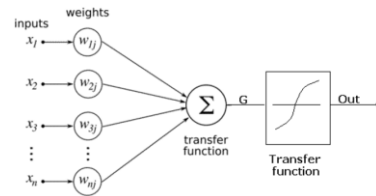
知識準拠診療支援（DSS）
医療ターミノロジー
医療オントロジー

ニューロネットワーク型
多層型ニューロネット
深層学習 Deep Learning
結合係数修正アルゴリズム
画像処理から創薬まで

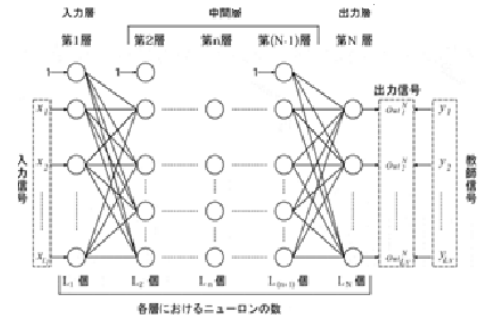
Deep Learning 型人工知能の 革命性

Deep Learning による 人工知能革命

- 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



神経情報素子



多層ニューロネットワーク

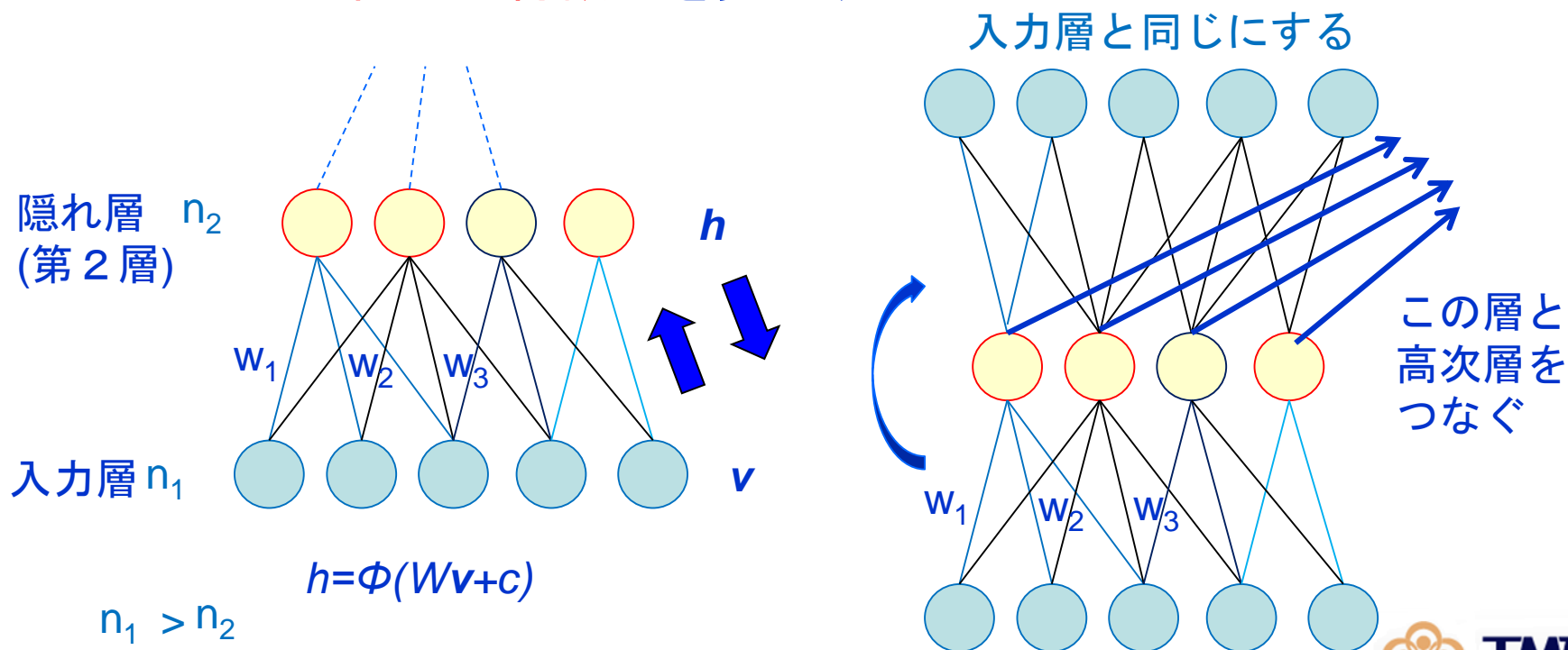
- Deep Learningの革命性
 - 「教師なし学習」

- 対象の特徴表現や対象の高次特徴量を自ら学ぶ



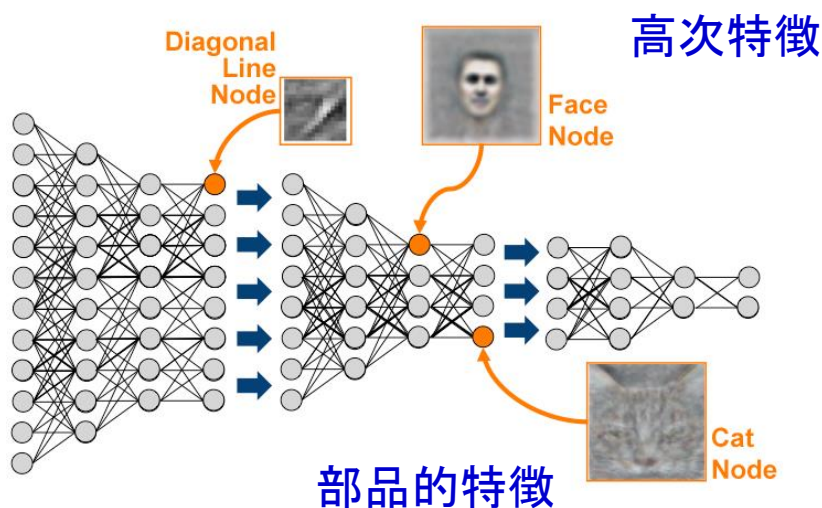
DLの革命点 Autoencode 1

- 対象に固有な**内在的特徴**を学ぶ自己符号化の原理
- 格段ごとに入力を少ない中間層を介して復元できるかを行なう
- 次元を圧縮されて可及的に復元する
 - できるだけ復元に**効果的な**特徴量を探索する
 - 内在的な特徴量**を見出す

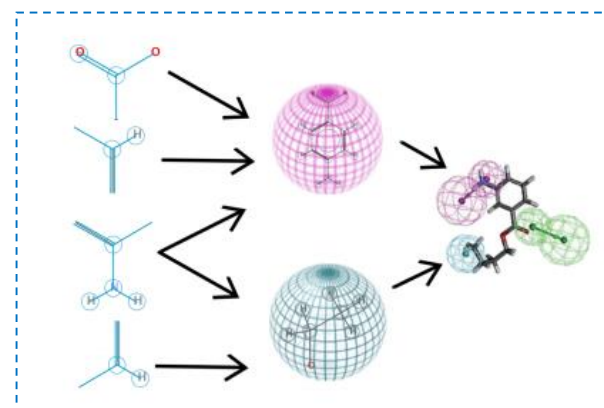


DLの革命点 Autoencoder 2

- 各層ごとに自己符号化を行うので**何層でも組める**
 - 各層間で「自己符号化」の積上げ (autoencoder stack)
- 第一層で学習した特徴量を使って次の階層を作るので**高次の特徴量**が作られる
- 特徴的表現と概念を結びつけるため「**教師あり学習**」が最後に必要。
- 自動特徴抽出によってこれまでの学習手法の限界を克服した
 - 内在的な特徴量による構造的な理解
- 人間の「思考の枠組み」を超えた正解の低次
 - 「アルファGo」が定石にない手で碁の名人に勝つ



Pharmacophoreの抽出

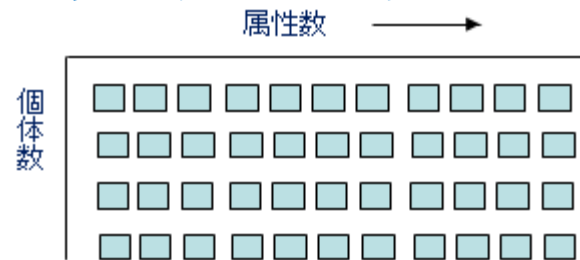


Deep Learningの医療・創薬へ応用

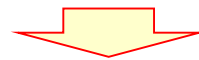
「ビッグデータ」のData 縮約原理

問題点 属性値数(p) ≫ サンプル数(n)

p: 数億になる場合あり n: 多くても数万、通常数千



これら膨大な属性変数がすべて独立ならばビッグデータの構造解析は不可能。単変量解析の羅列 (GWASのManhattan Plot) しか可能でない



ビッグデータ・スパース仮説

ビッグデータは、多数であるが属性値数より少ない独立成分が基底となって、相互にModificationして構成されている。
(独立成分の推定は、サンプル数とともに増加する)

データ次元縮約の原理 (**principle of compositionality**)

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–臨床表現型情報>
 - ゲノムコホートにおける<遺伝子情報–環境（生活様式）情報>
- Deep Learning-based Network Contraction
「DLネットワーク縮約法」

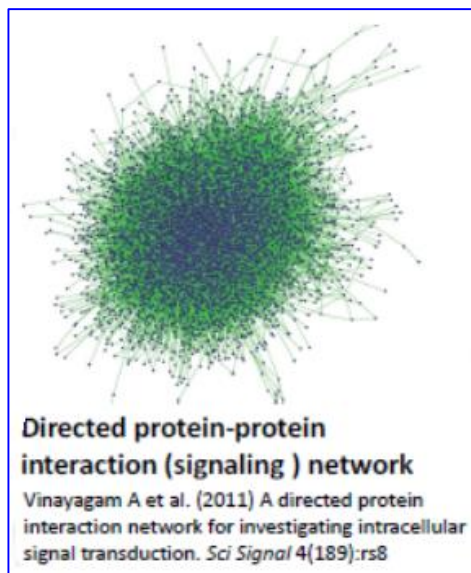
超多次元ネットワーク情報構造⇒
少数の特徴的ネットワーク基底に分解
- 線形分解ではない。非線形分解で基底への射影

特徴的ネットワーク基底への分解

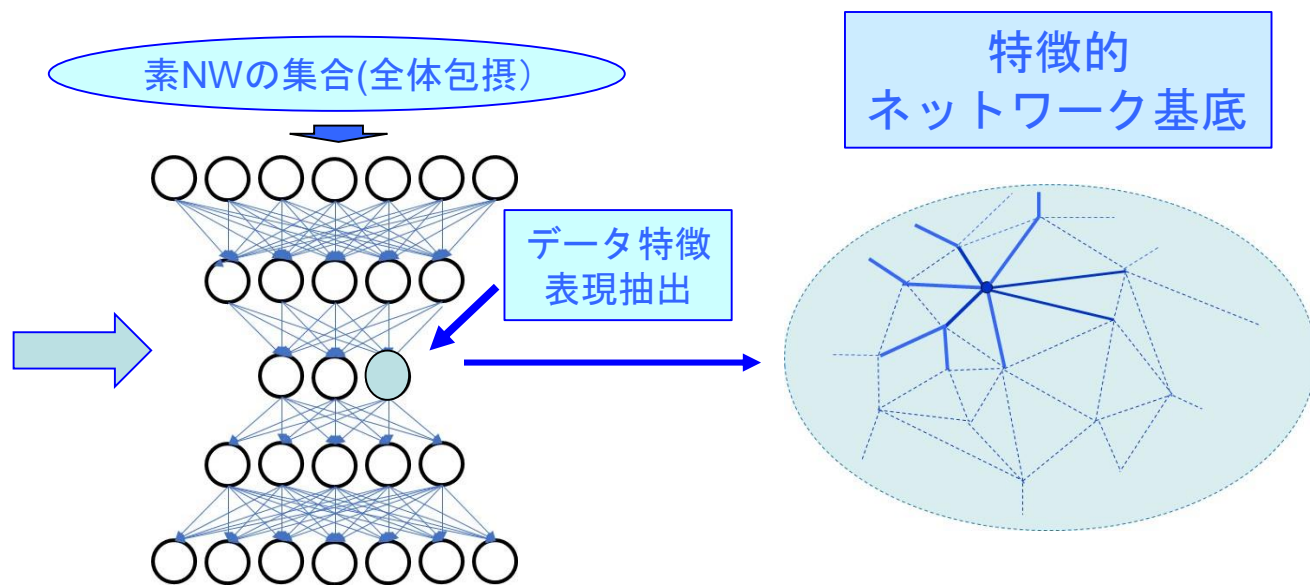
特徴的ネットワーク基底の和に縮約

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

1) 生体ネットワーク (PPIN) 特徴量の抽出

- タンパク質相互作用ネットワーク(PPIN)のNW結合を学習し**特徴表現** (特徴NW基底) を出力。
- 学習集合を部分ネットワークの集合から決める
- ノードを起点とした素NWでPPIN全体を覆う集合

2) 多層Stacked Auto-encoderのDLで学習

- 特徴的NW基底の「教師無し」学習
- 次元縮約による特徴的NW基底の抽出

3) DL特徴NW基底空間における正例補完

- DrugBankからの正例とその増加 (SMOTE法)

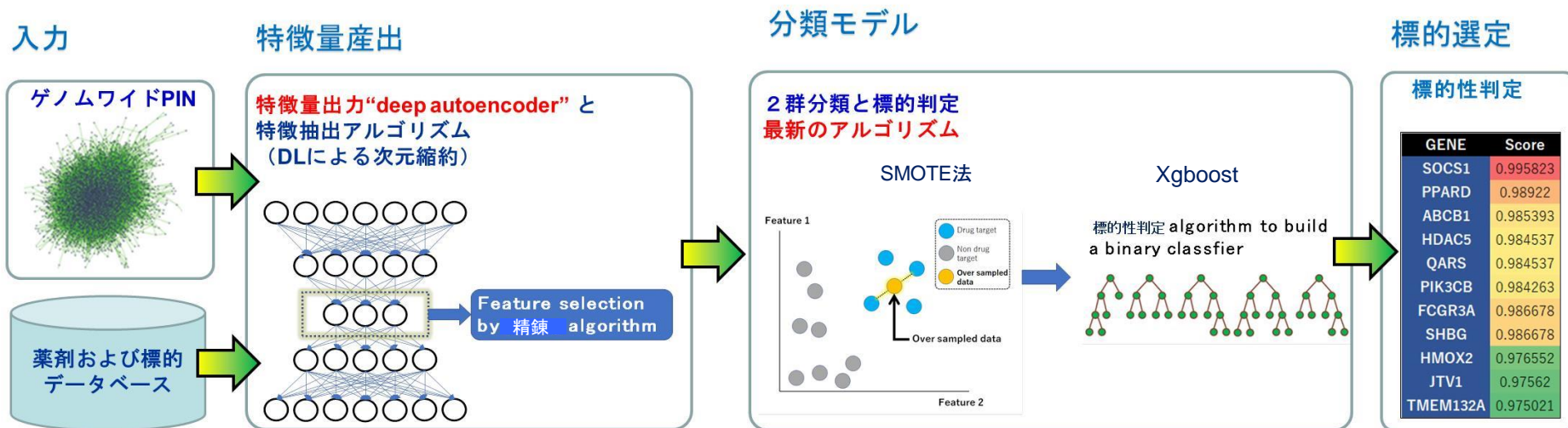
4) DL特徴NW基底量を用いた機械学習分類

- Xgboot法などを用いたDL特徴量からの判別ネットワーク・タンパク質の標的性の判定

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



従来の機械学習 (Random Forrest)と同じ成果は得られている

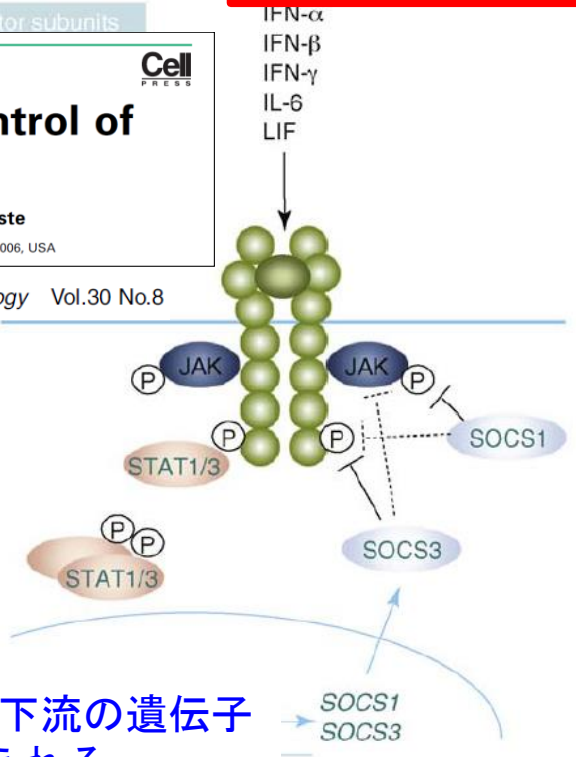
<疾患- 標的分子> 予測結果

Target disease	Top 25 genes for potential novel drug target for a target disease
Anti-Alzheimer's disease	SOCS1 ; <i>EPHA2; PRKCS; YES; UBB; HSP110A; HSPG2; PARD6A; BLNK; PRKCI; YES1; PPP3CA; MMP11; SPTAN1; PTPRC; FBLN1; RPLP1; FAM107A; TRADD; NR4A1; LAT; NF2; PRKCE; KIT; NID1</i>
Anti-Anxiety	<i>S1PR1; CNR1; MTNR1A; CCL4; F2; TEC; IL8; CRHR1; AGTR2; OPRD1; IL8RA; RNF43; RHO; SP6; RAB13; DRD4; IL8RB; MMP9; MMP2; OPRM1; IL1B; GNAS; S1PR3; KIT; GRM2</i>
Anti-Rheumatoid	<i>SLC22A5; GRASP; KIT; SLC22A4; CFH; COX2; HSP90AA1; UBB; DHRS3; SCTR; ADORA1; MUC1; C6orf47; NR4A1</i>
Anti-Breast Cancer	<i>SHC1; NFKB1; RELA; ID2; RAC1; SRC; JUN; HDAC2; I</i>
Anti-Colorectal	<i>SHC1; GHR; HRAS; PLCG1; NFKB1; LYN</i>
Anti-Pancreatic	<i>SH2; DSCC1; TAF9; AIF1; RNF223; NSMCE4A; ASPH;</i>
Anti-Melanoma	<i>RAP; HMGB2; FEN1; NFKB1; MUC1; LMNB1; GDE1;</i>

SOCS1はJAK/STAT pathwayを介してサイトカインの応答を変動させ、中枢神経系の炎症を制御

Review
SOCS1 and SOCS3 in the control of CNS immunity
 Brandi J. Baker, Lisa Nowoslawski Akhtar and Etty N. Benveniste
 Department of Cell Biology, The University of Alabama at Birmingham, Birmingham, AL 35294-0006, USA

Trends in Immunology Vol.30 No.8



しかし、SOCS1は上流の遺伝子なので、この下流の遺伝子を標的にした方が、長期投与には良いとも考えれる。

疾患-標的分子リンクの同定よりDRへ

機械学習で予測された、新規標的の情報(disease A と target の情報, 標的が disease A の新規標的分子、青いリンク)を、既知の drug-target-disease interaction network をマップし薬剤の新しい適用疾患 (赤リンク) を予測

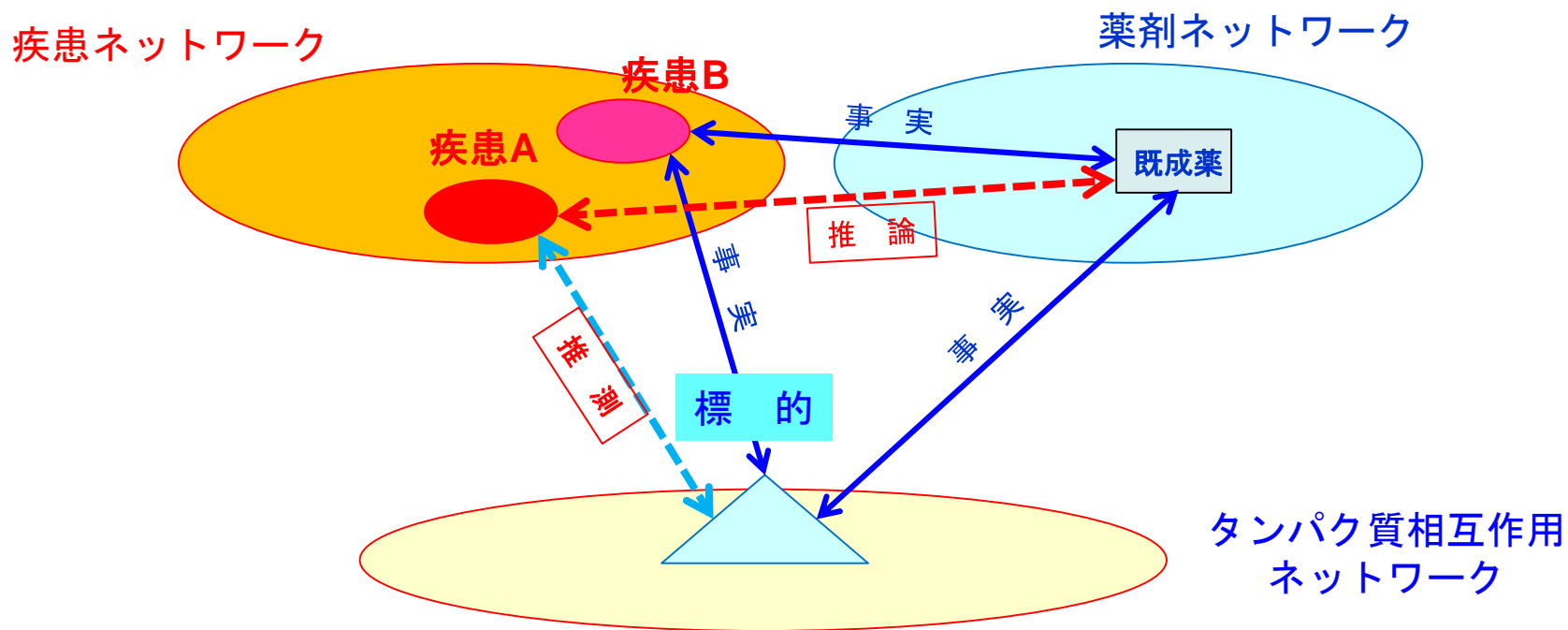


Table 3. Predicted repositionable drug candidates

Target disease	Candidate repositionable drug
Anti-Alzheimer's disease	Imatinib; Marimastat; Nilotinib ; Regorafenib; Sorafenib; Tamoxifen; Urokinase
Anti-Anxiety	3-Methylfentanyl; Agomelatine; Amitriptyline; Amoxapine; Antihemophilic Factor; Apomorphine; Aripiprazole; Bromocriptine; Buprenorphine; Butorphanol; Cabergoline; Canakinumab; Captopril; Chlorpromazine; Coagulation Factor IX; Codeine; Dextromethorphan; Dextropropoxyphene; Dopamine; Drotrecogin alfa; Ethylmorphine; Etorphine; Fentanyl; Halothane; Hirulog; Hydrocodone; Hydromorphone; Imatinib; Ketamine; Ketobemidone; L-DOPA; Lepirudin; Levallorphan; Levorphanol; Lisuride; Loperamide; Loperidine; Marimastat; Melatonin; Menadione; Methadone; Methadyl Acetate; Methotrimeprazine; Minocycline; Morphine; Naloxone; Naltrexone; Nilotinib; Olanzapine; Ondansetron; Oxycodone; Oxymorphone; Paliperidone; Pergolide; Pethidine; Pramipexole; Promazine; Propiomazine; Quetiapine; Regorafenib; Remifentanyl; Remoxipride; Risperidone; Ropinirole; Rotigotine; Sorafenib; Sufentanil; Suramin; Thiothylperazine; Ziprasidone
Anti-Rheumatoid	Acetylcholine; Adenosine; Amiloride; Aminohippurate; Aminophylline; Amphetamine; Ampicillin; Azidocillin; Benzylpenicillin; Cefalotin; Cefdinir; Cefixime; Cephalexin; Choline; Cimetidine; Clonidine; Cyclacillin; Desipramine; Diphenhydramine; Dopamine; Dyphylline; Enprofylline; Epinephrine; Furosemide; Grepafloxacin; Histamine Phosphate; Imatinib; Imipramine; Insulin, isophane; Ipratropium bromide; L-Arginine; L-Carnitine; Levofloxacin; Lidocaine; Liothyronine; Lomefloxacin; Mepyramine; Methamphetamine; Nicotin; Nicotine; Nilotinib; Norepinephrine; Norfloxacin; Ofloxacin; Oxtriphylline; Pentoxifylline; Probenecid; Procainamide; Quinidine; Quinine; Regorafenib; Rifabutin; Secretin; Sorafenib; Spermine; Testadenoson; Testosterone; Tetraethylammonium; Theophylline; Thiamine;

Neuroscience 304 (2015) 316–327

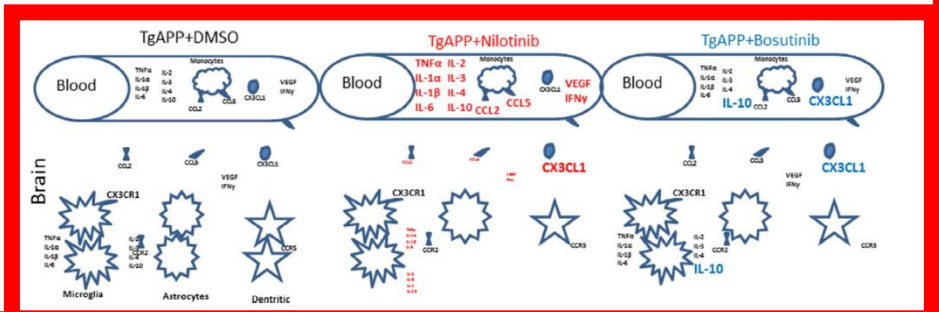
慢性白血病の抗がん剤であるニロチニブがアルツハイマーのDR薬剤として選定

NILOTINIB AND BOSUTINIB MODULATE PRE-PLAQUE ALTERATIONS OF BLOOD IMMUNE MARKERS AND NEURO-INFLAMMATION IN ALZHEIMER'S DISEASE MODELS

I. LONSKAYA,^a M. L. HEBRON,^a S. T. SELBY,^a
R. S. TURNER^b AND C. E.-H. MOUSSA^{a*}

^a Department of Neurology, Laboratory for Dementia and Parkinsonism, Georgetown University Medical Center, Washington D.C. 20007, USA

^b Department of Neurology, Memory Disorders Program, Georgetown University Medical Center, Washington D.C. 20007, USA



DL型NNへの期待と困難点

- 医療・創薬の応用は開始段階で応用成功例は少ない
 - 本質的に「教師なし学習」:人間が思いつかない解を提示
 - 画像分類・解釈と文章理解が優れているので、遺伝子発現プロファイル解析や病態推移の理解への応用が期待される
 - 例：ヒトmicrobiomeの分類・階層的表現を得た
 - 6つのがんで遺伝子発現をmiRNAとともに分類した。
 - 異なったMicroarrayを含むがん発現を分類の特徴表現を導き分類した。
 - Convolution ネットワークを使用して画像としての遺伝子発現を分類した。
 - 遺伝子発現プロファイルの自動アノテーション
 - 医療における超多次元ネットワークから革新的知識を発見できるであろう
- DL型ニューラルネットの困難点
 - 特徴表現を自己学習するが基本的にはBlack Box
 - 大量のデータを必要とする
 - DL型NNには、ハイパーパラメータが多種類があり、使用に関して選択問題が残る
 - 計算時間が長く、コストが大きい。

第2世代のゲノム医療に向けて

ゲノム医療の第2世代

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: $G \times G$
環境と遺伝子の相互作用が: $G \times E$

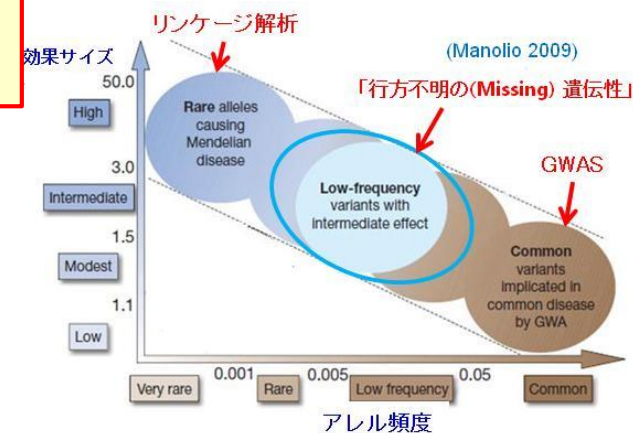
SNPの相対リスク
低い(1.1~1.3)理由
 $G \times E$ 組合せ特異的効果
を環境要因の平均



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



第2世代網羅的分子医療 メタオミックス

＜遺伝子要因と環境との相互作用の基底＞はどんな機序で行われているか

エピゲノム

オランダ
飢饉 (1944)



環境によるエピゲネティック修飾

DOHaD(Developed Origin of Health and Diseases) 学説

オランダ飢饉のとき、母親の胎内にいた人々
出生30年後、肥満、糖尿病、心疾患、高罹患率

過度な低栄養：肝臓のPPARα/γ (儉約遺伝子) 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

環境因子

Epigenome変化

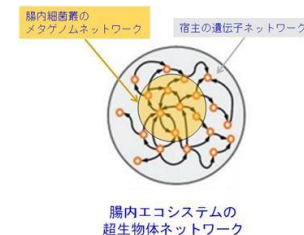
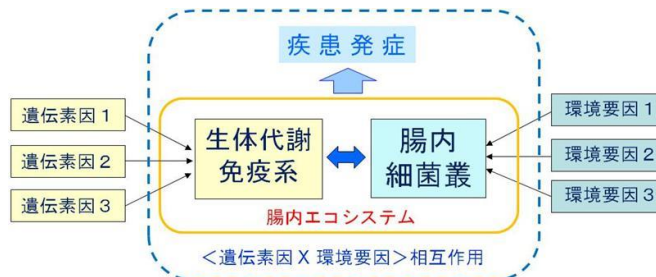
遺伝子発現調節

疾病発症

メタゲノム

Microbiomeにおける生体細菌叢相互作用

- ・ 食事などの栄養物質による環境要因は、**腸内細菌叢の代謝物**を介して、宿主の生体機構に相互作用
- ・ 心筋梗塞や糖尿病、**腸内細菌が産出する代謝物**（短鎖脂肪酸やTMAOなど）が**生体シグナル物質**や**生体活性物質**となって**受容体や転写因子の活性化**して生体側の**遺伝子ネットワーク**に働きかける。
- ・ 腸内細菌叢と生体の＜**超生物系; hologenome**＞において＜環境要因x遺伝素因＞の相互作用



ホロゲノム
hologenome

免疫ゲノム

TCRのゲノム配列の多様性解析（レパトア解析）病原環境によって変化

今後の戦略・方向

- 第2世代のゲノム医療・創薬
- Deep Learningによる〈多次元ネットワーク情報構造〉の縮約
 - 創薬だけでなく、ビッグデータ医療への適応可能
 - ゲノム医療の〈網羅的分子情報—臨床表現型〉の
相関ネットワーク構造
 - バイオバンクの〈遺伝素因—環境要因〉と発症
- AI創薬の「枠組み」実行方向は「見えてきた」
- 本年中に、いよいよAI創薬の実装に着手しなければならない。米国に持って行かれる。
 - 製薬企業、IT企業、医療機関を束ねた集中的プロジェクトを推進するために「ビッグデータ医療・AI創薬コンソーシアム」を設立する

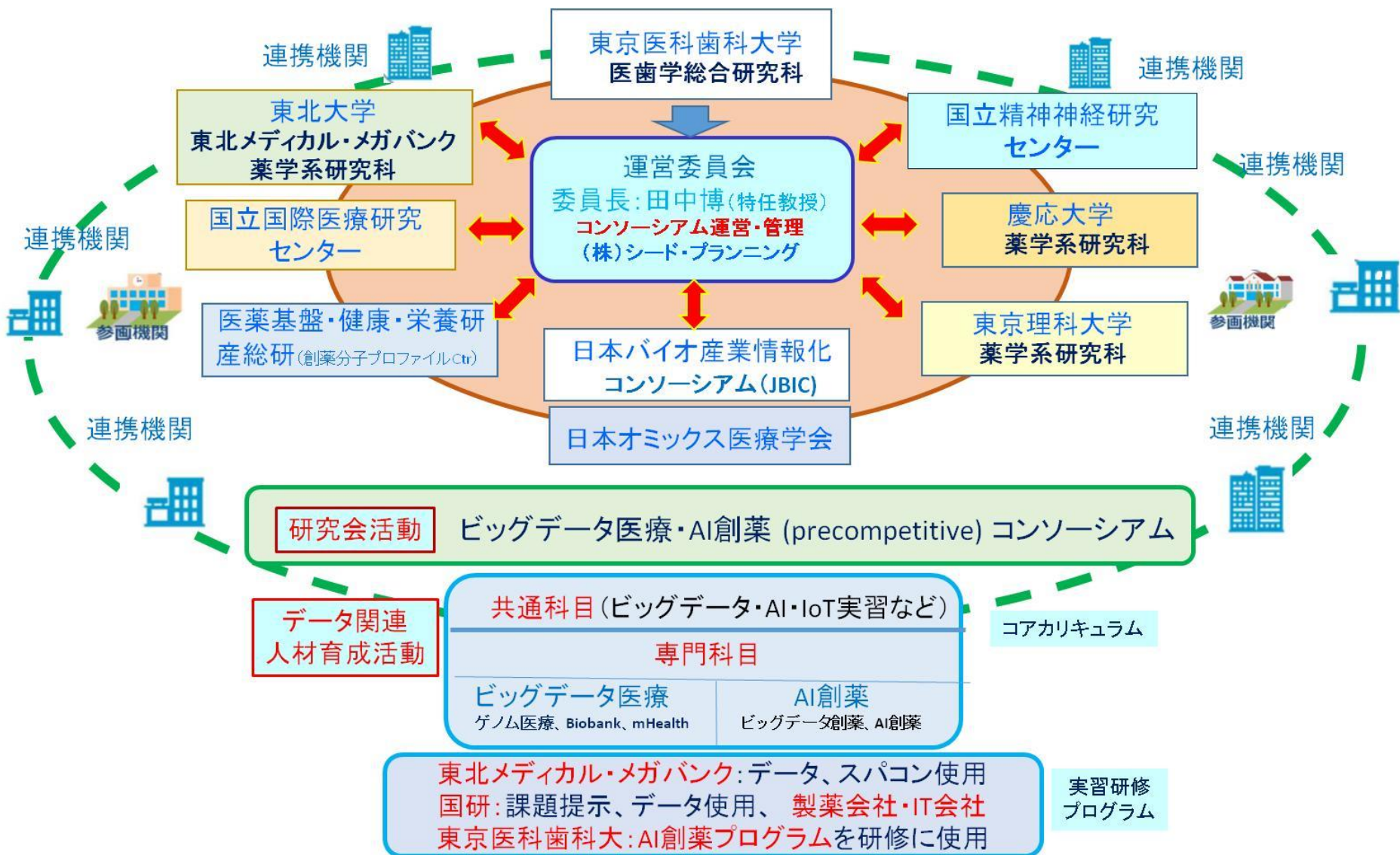
田中 博 著

「AI創薬・ビッグデータ創薬」

薬事日報社 6月19日刊行



ビッグデータ医療・AI創薬コンソーシアム



ご清聴有難う御座いました

