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

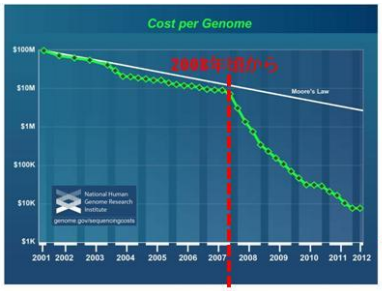
- ビッグデータ医療（ゲノム医療）
  - 米国型のゲノム医療と欧州型のゲノム医療
  - メタオミックスの時代
  - ビッグデータ医療とAIによる知識発見
- AI創薬
  - 分子プロファイル型計算創薬・DR
  - Deep Leaning を用いたAI創薬・DR

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」からの怒涛の展開（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



千ドル
ゲノム

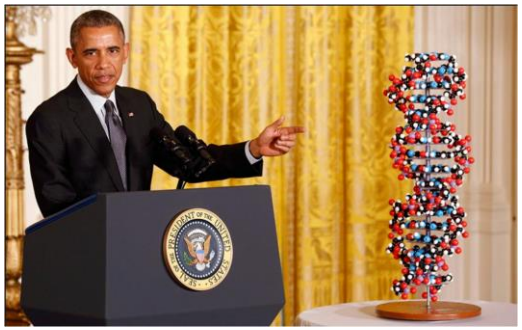
DNA Sequencing Cost: the National Human Genome Research Institute

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



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

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

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

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

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

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

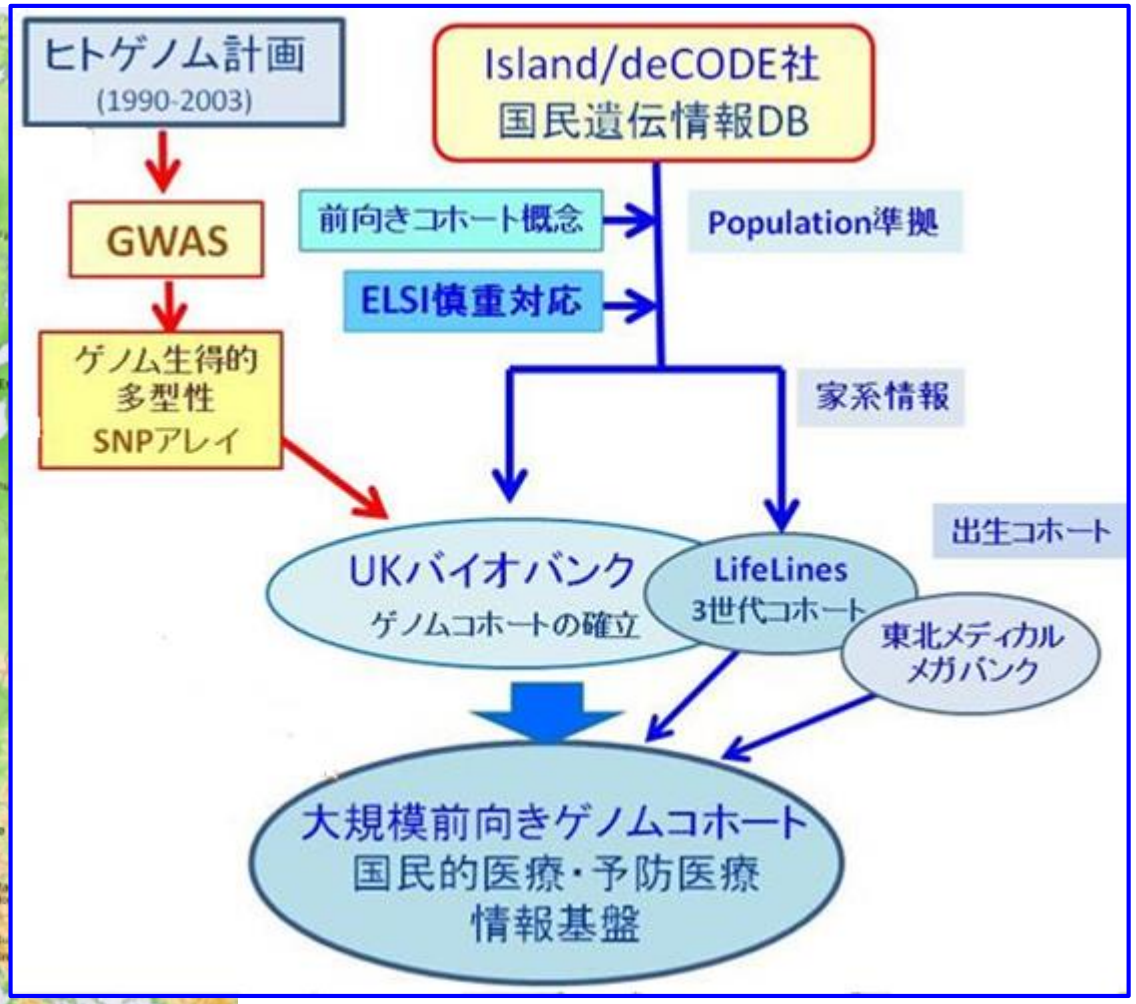
100万人コホート:バンダービルト大学設開始(2016-2020)
NCI“National Cancer MoonShot” 10年計画開始
各州でのプレシジョン医療計画開始(カリフォルニア、ペンシルバニア)

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

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

TMDU

欧州のバイオバンクの流れ



TMDU

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

TMDU

医療ビッグデータ

- 臨床ゲノム・オミックス医療の進展
  - Clinical Sequenceのインパクト
  - 網羅的分子情報、臨床表現型情報の統合
  - 個別化医療、Precision Medicine
- Biobank, 疾患レジストリの拡充
  - 疾患型：個別化医療の情報基盤
  - 住民型：慢性疾患発症予測・個別化予防
  - レジストリ準拠ランダム臨床治験
- 網羅的分子情報DBの大規模化と利用
  - 1000genome, GWAS, ICGC, Clinvar, Clingen, dbGaP, LINCS, HPRD,STRINGなど

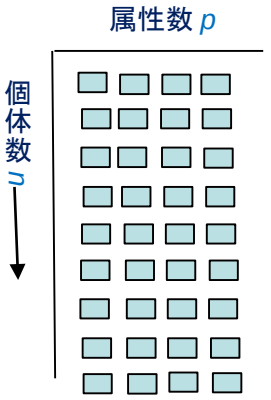
医療の「新しいビッグデータの革命性」

～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」

Big “Small Data”　(n≫p)



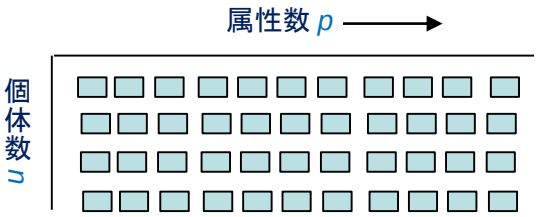
医療情報・疫学調査では 属性数：数十項目程度

– 目的：Population Medicine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

Small “Big Data”(p≫n)



1個体に関するデータ属性種類数が膨大

属性(*p*)に比べて個体数(*n*)少数:従来の統計学が無効

「新 ***np*** 問題」: GWASは単変量解析の羅列

– 目的：例えば医療の場合 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

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 (innovative insight)」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革新的知の発見
  - 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

    如何に「innovation knowledge」を獲得するか

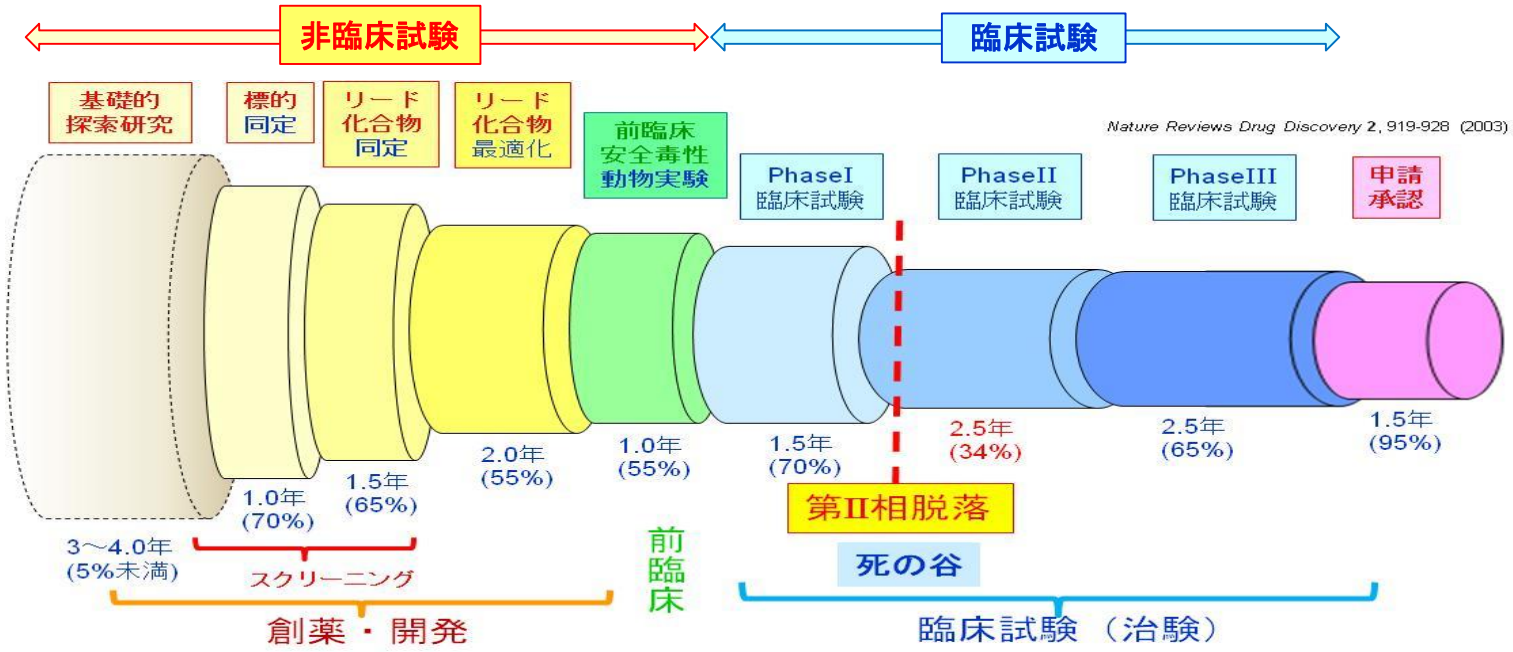
TMDU

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**

Nature Reviews Drug Discovery (2012)



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。

Tanaka K…,**Tanaka H**, ..,Mizushima T, *Nature Comm*, 4:2686, 2013

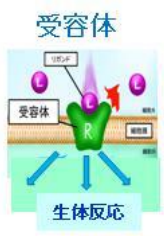
疾患・薬剤・標的の関係

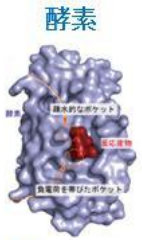
病気の主要な要因

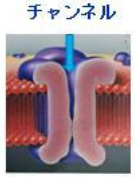
疾患関連タンパク質（複数）

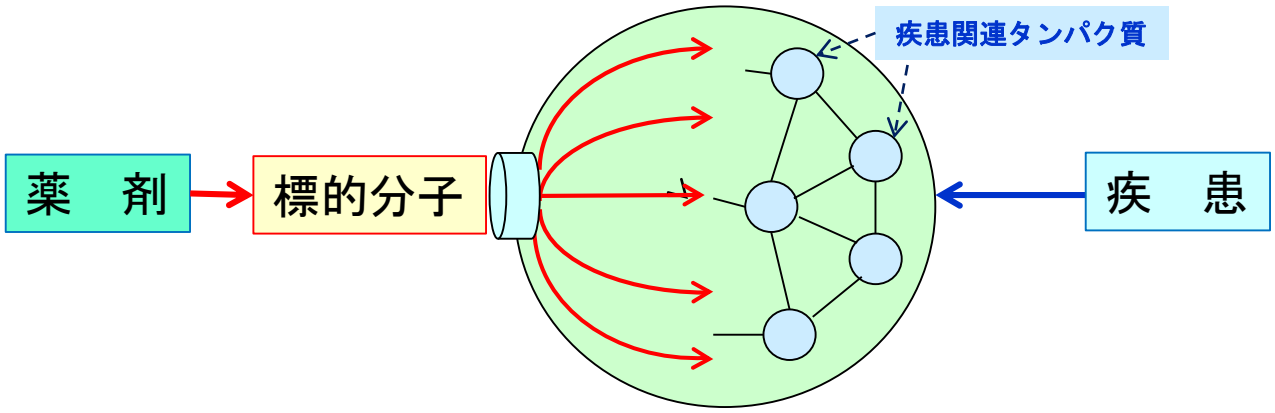
薬：疾患関連タンパク質に影響を示す
標的タンパク質に作用し阻害する

薬剤の標的分子
受容体・酵素・チャンネルなど







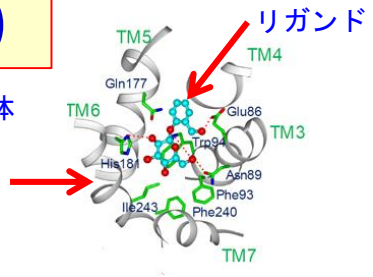


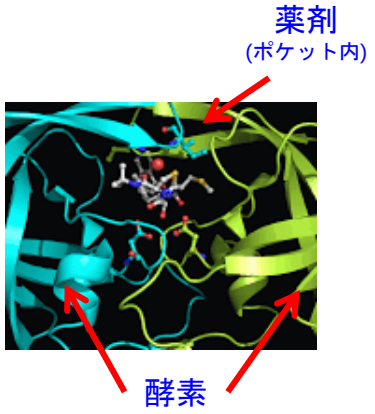
生体システム/ネットワーク

ビッグデータ計算創薬１

計算創薬(computational drug discovery) の新しい方向

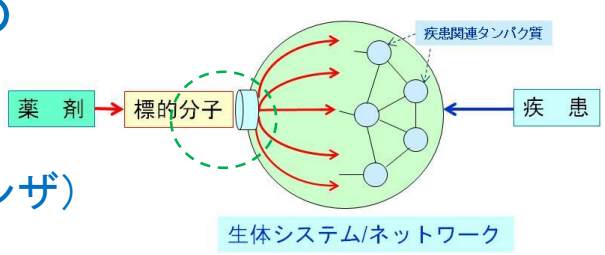
これまでの計算創薬 (*in silico* 創薬)





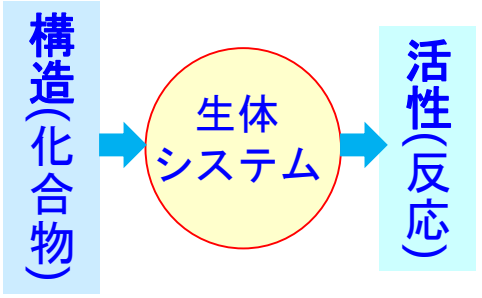
分子(結合構造)中心

- **分子構造解析・分子設計**
- **Structure-based rational drug design**
- **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(量子化学等を使う)**
  - 成功例：インフルエンザ薬　ザナミビル（リレンザ）
- 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

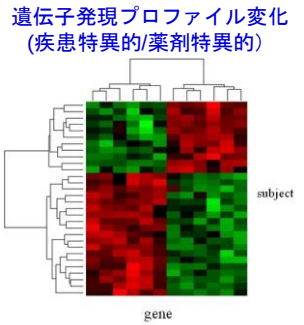


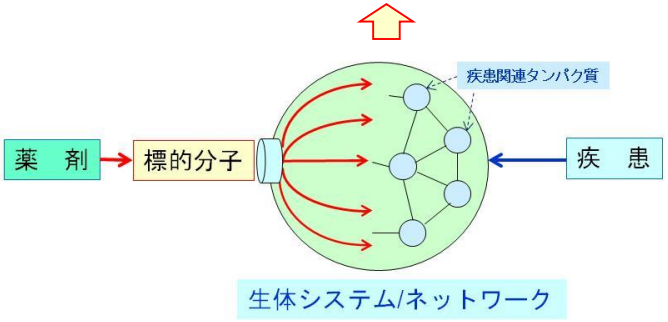
ビッグデータ計算創薬２

新しい計算論的創薬のアプローチ（生体分子プロファイル型創薬）

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





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＜薬剤ー標的分子の結合＞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

TMDU

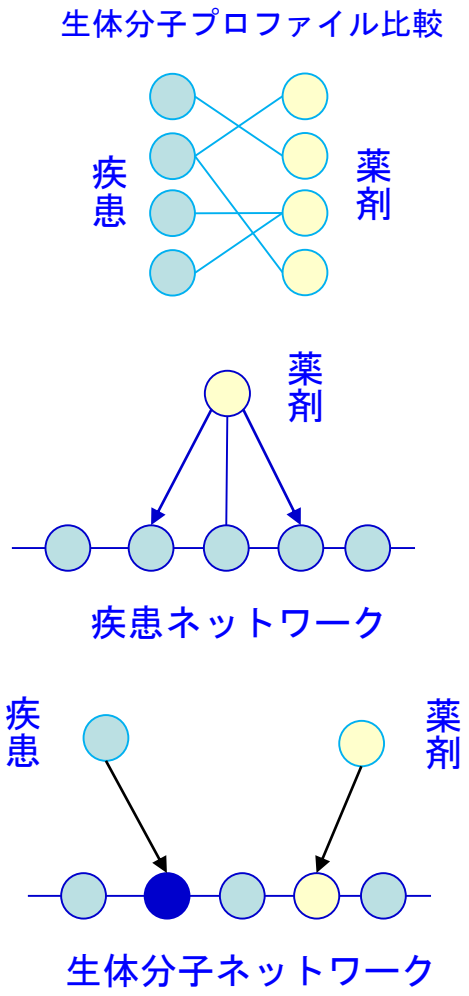
生体分子プロファイル型創薬/DR 方法論の深化

深化

第１段階：疾患・薬剤プロファイル直接比較
- 疾患罹患時と薬剤投与時の生体反応の遺伝子発現プロファイルを比較。
- パターン正負相関性に基づく有効性毒性予測

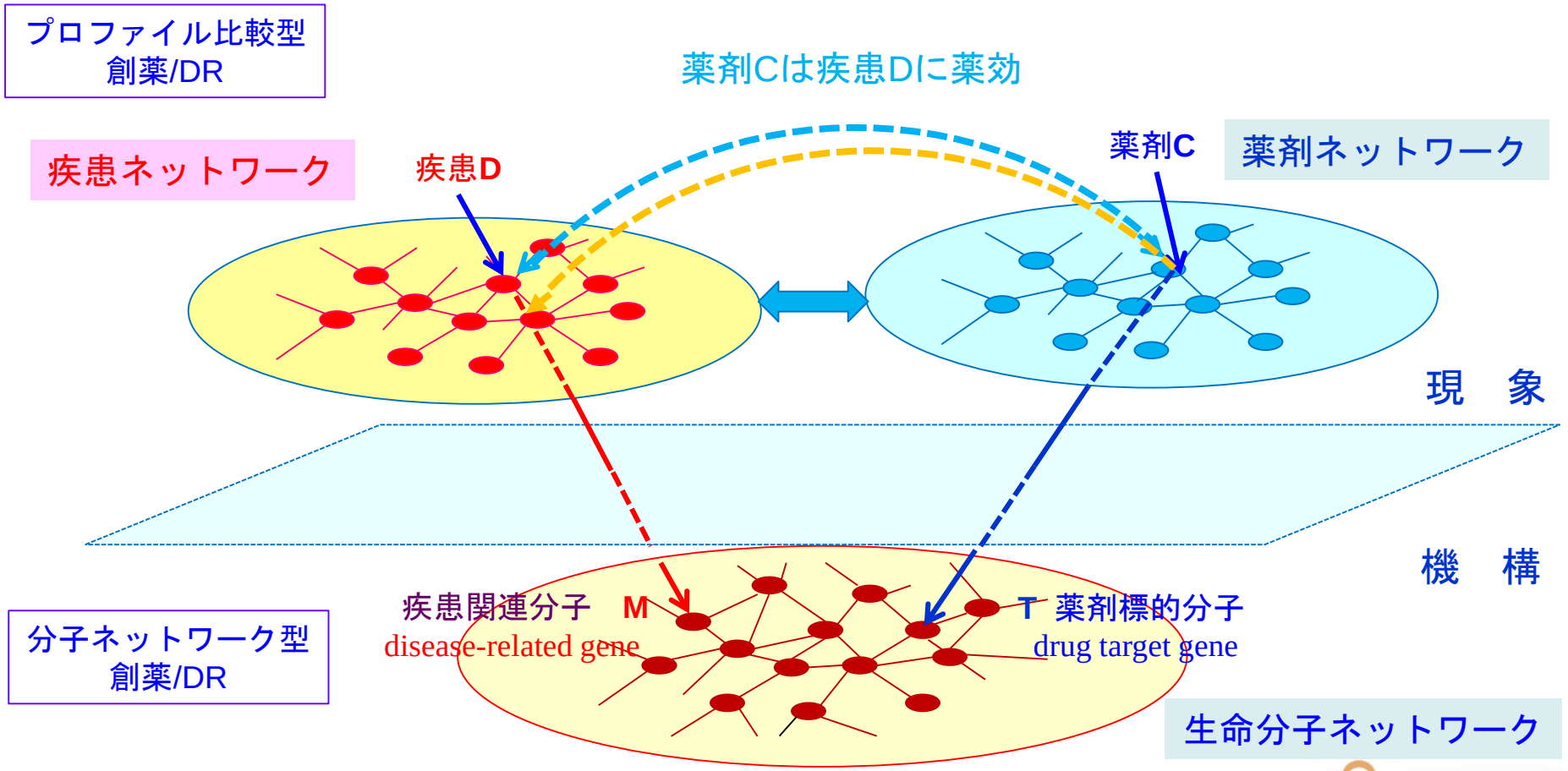
第２段階：疾患・薬剤ネットワーク近接解析
- 疾患あるいは薬剤の集合をネットワーク表現
- ネットワークに準拠して有効性・毒性予測

第３段階：生体ネットワーク媒介型比較
- 生体分子ネットワークを＜場＞として疾患・薬剤の足場分子を同定
- 足場分子間の相互作用としての有効性・毒性予測



生体分子プロファイル型計算創薬/DRの基本的枠組み

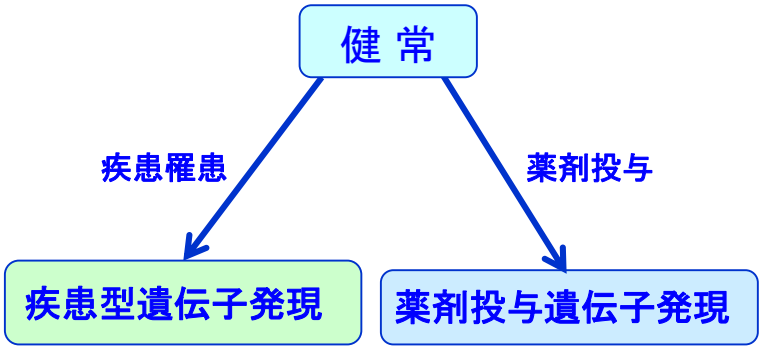
3層の生体・薬剤のネットワーク間の関係図式



１．遺伝子発現プロファイル比較型創薬・DR

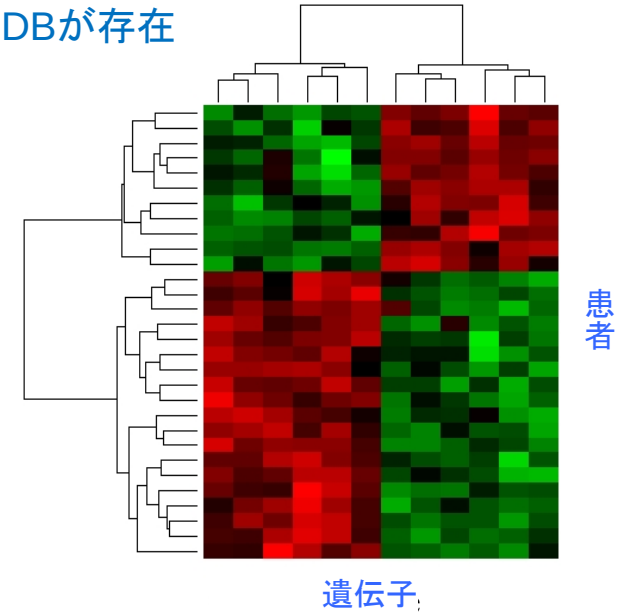
ビッグデータ計算創薬
発現プロファイル比較型創薬・DR

- 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
    - シグネチュア (署名)差異的発現遺伝子代表群
    - DB利用：シグネチャアを「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



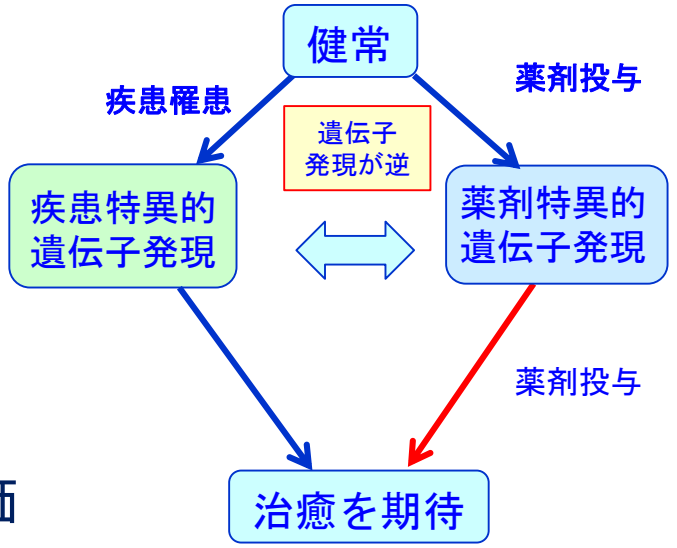
- 疾病特異的遺伝子発現
  - GEO（Gene Expression Omnibus),
    - 疾病罹患時の遺伝子発現プロファイルの変化
    - 米国NCBI作成・運用　2万5千実験,
      70万プロファイル（欧州　ArrayExpress）
    - もEBIが作成、サンプル数同程度

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

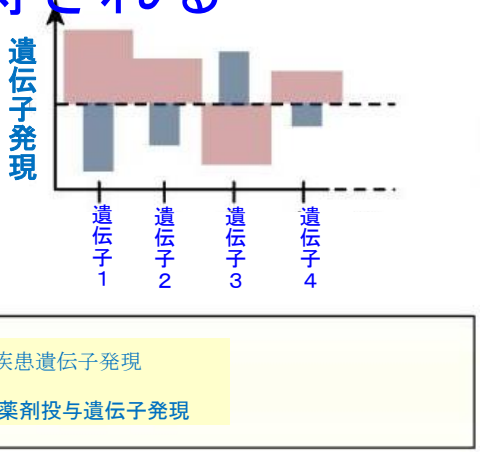


遺伝子発現プロファイルによる有効性予測

- 遺伝子発現シグネチャア逆位法
  - 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

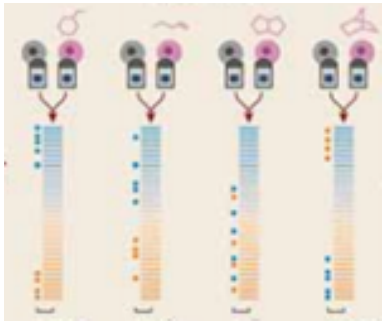


- 効果がお互いに打ち消すなら**有効性**が期待される
  - 例：炎症性腸疾患に 抗痙攣剤(topiramate),
    骨格筋委縮にウルソール酸



ベースは疾患遺伝子発現
横の点は薬剤遺伝子発現

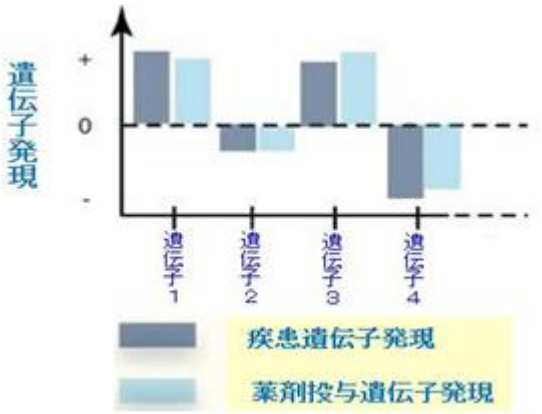
青は発現が**上昇**した遺伝子
赤は発現が**下降**した遺伝子



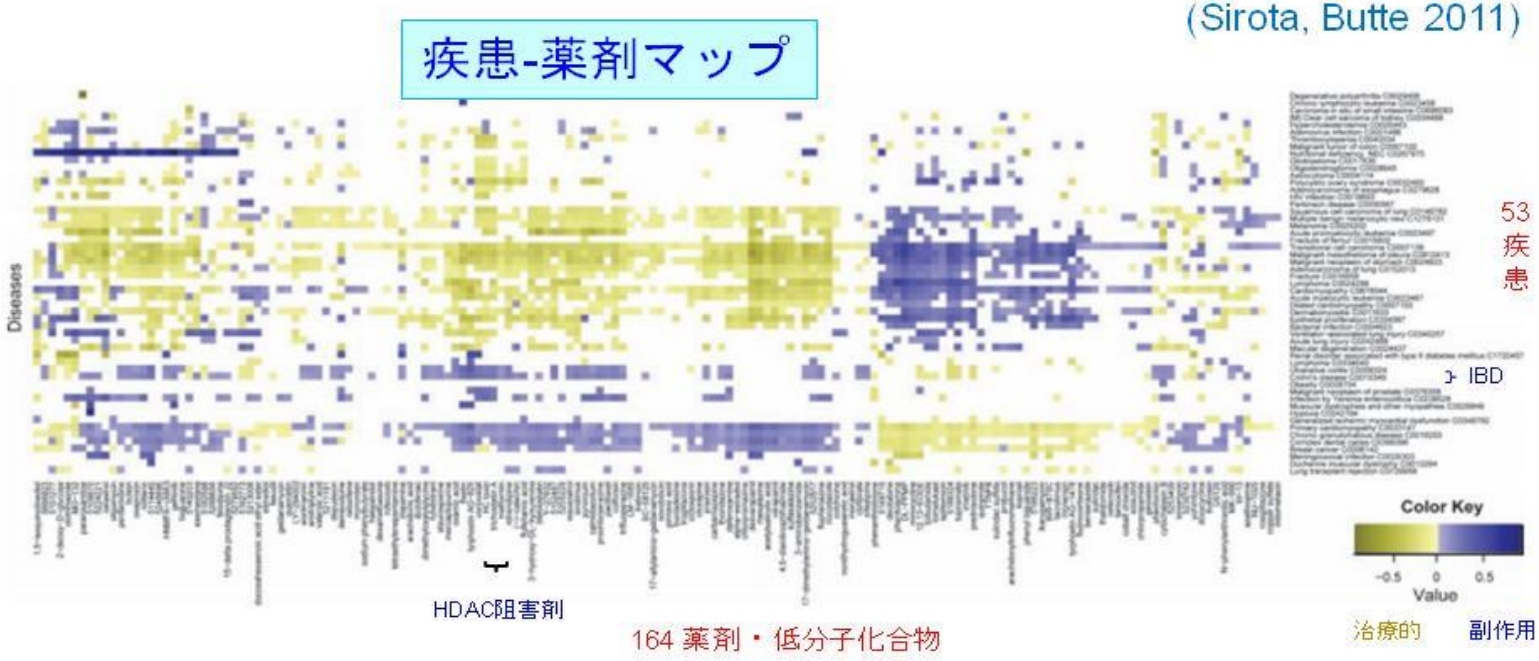
TMDU

遺伝子発現プロファイルによる毒性予測

- 連座法 guilt-by-association ：
- **薬剤ー疾患間　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

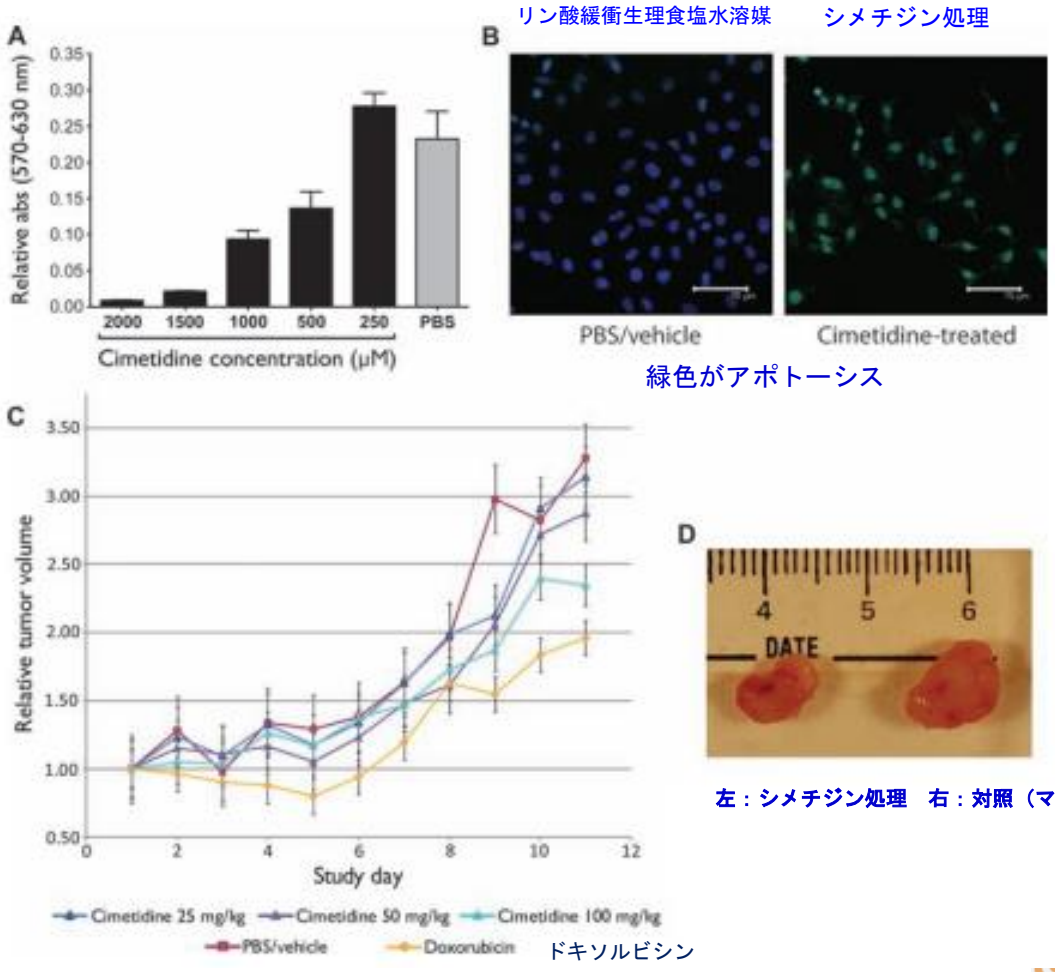


(Sirota, Butte 2011)



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



遺伝子発現プロファイル による疾患ー薬剤ネットワーク

遺伝子発現プロファイルの類似性を相関係数、ESによってリンク （Hu, Agarwal, 2009）

疾患ー疾患、薬剤ー薬剤、疾患ー薬剤の
ネットワークを発現プロファイルより構成

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-疾患NWの60%はMeSH** (既知体系)
その他は**分子レベル疾患分類学**
遺伝子発現の類似性による疾患体系

②**主な発見**

＜疾患 - 疾患＞

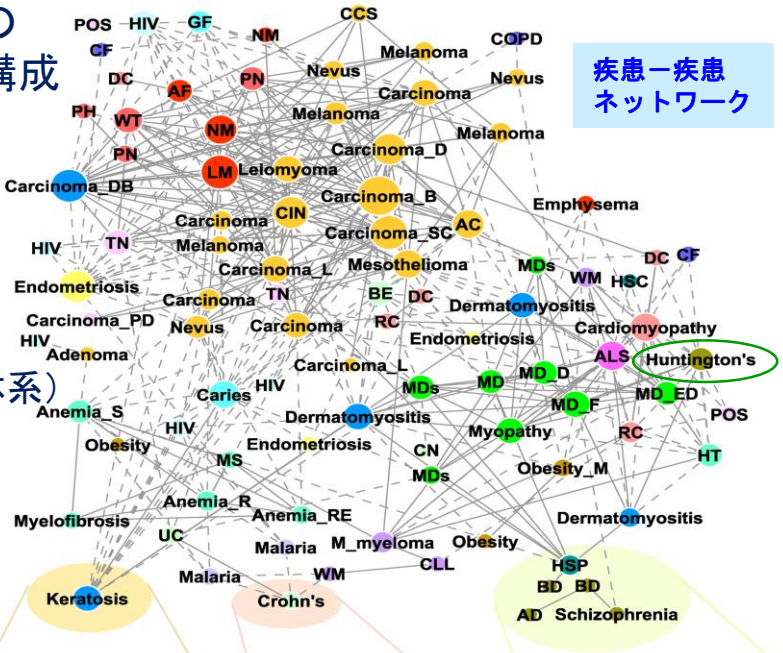
HSP（Hereditary Spastic Paraplegia
（遺伝性痙攣性対麻痺）
⇒bipolar 双極性障害

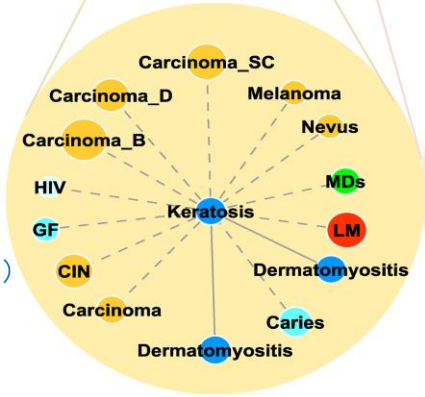
Solar keratosis 日光性角化症
⇒ cancer(squamous)

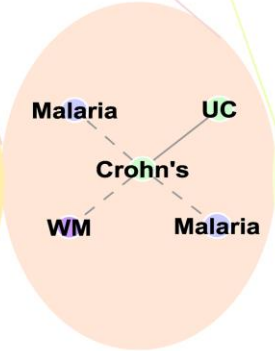
＜疾患 - 薬＞

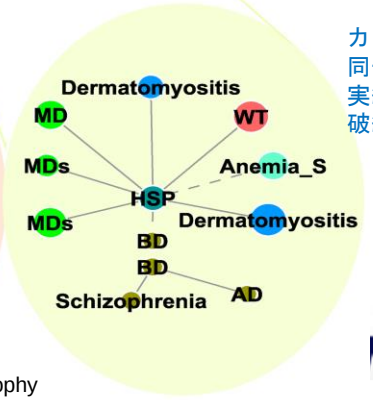
有効性：マラリア治療薬
⇒ Crohn's disease

（ベトナム経験：クローン病罹患保護効果）
ハンチントン病に種々の薬剤









C2 virus disease
C3 parasitic diseases
C4 neoplasms
C5 musculoskeletal
C6 digestive
C7 stomatognathic
C8 respiratory tract
C10 nervous system
C12 male urogenital
C13 female urogenital
C14 cardiovascular
C15 hemic and lymphatic
C16 congenital
C17 skin/connective
C18 nutritional/metabolic
C19 endocrine system
C20 immune system
C23 pathological conds
F03 mental disorders

カラーはMeSH
同一カテゴリー
実線はMeSH内
破線はMeSH外

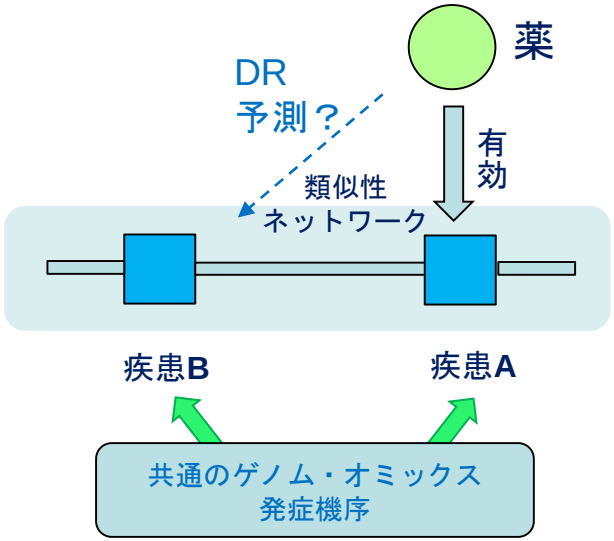
MD: muscular dystrophy

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年に亙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病機序の概念が基礎



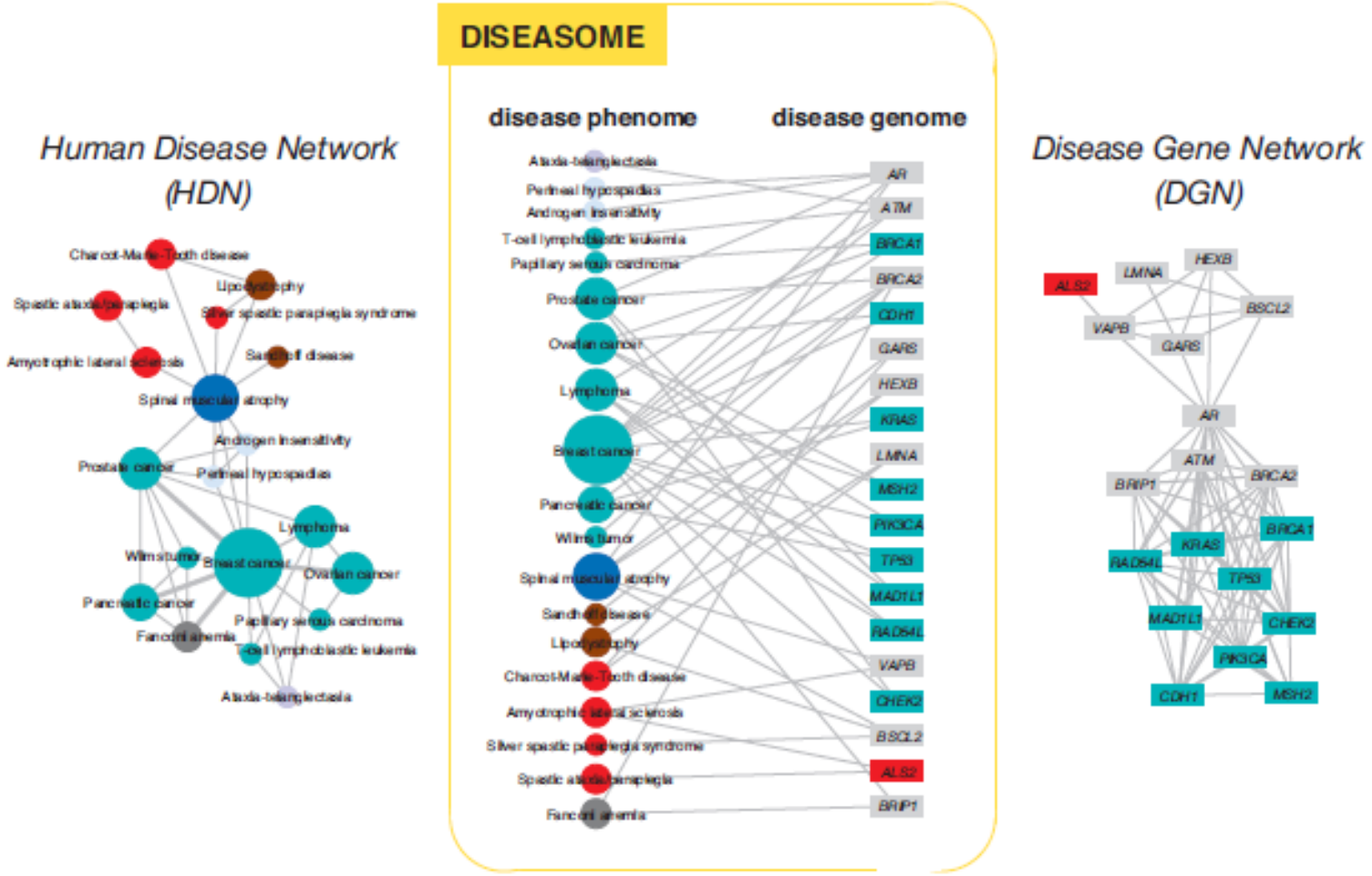
TMDU

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

TMDU

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

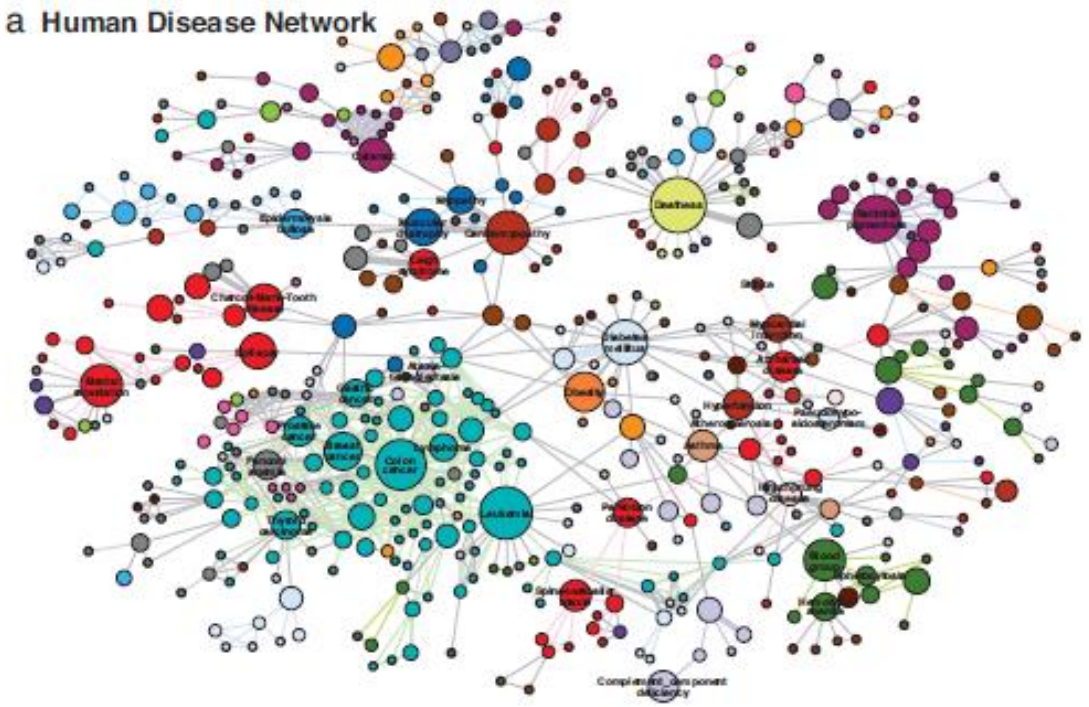
１つ以上の疾患を共有する疾患関連遺伝子

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

TMDU

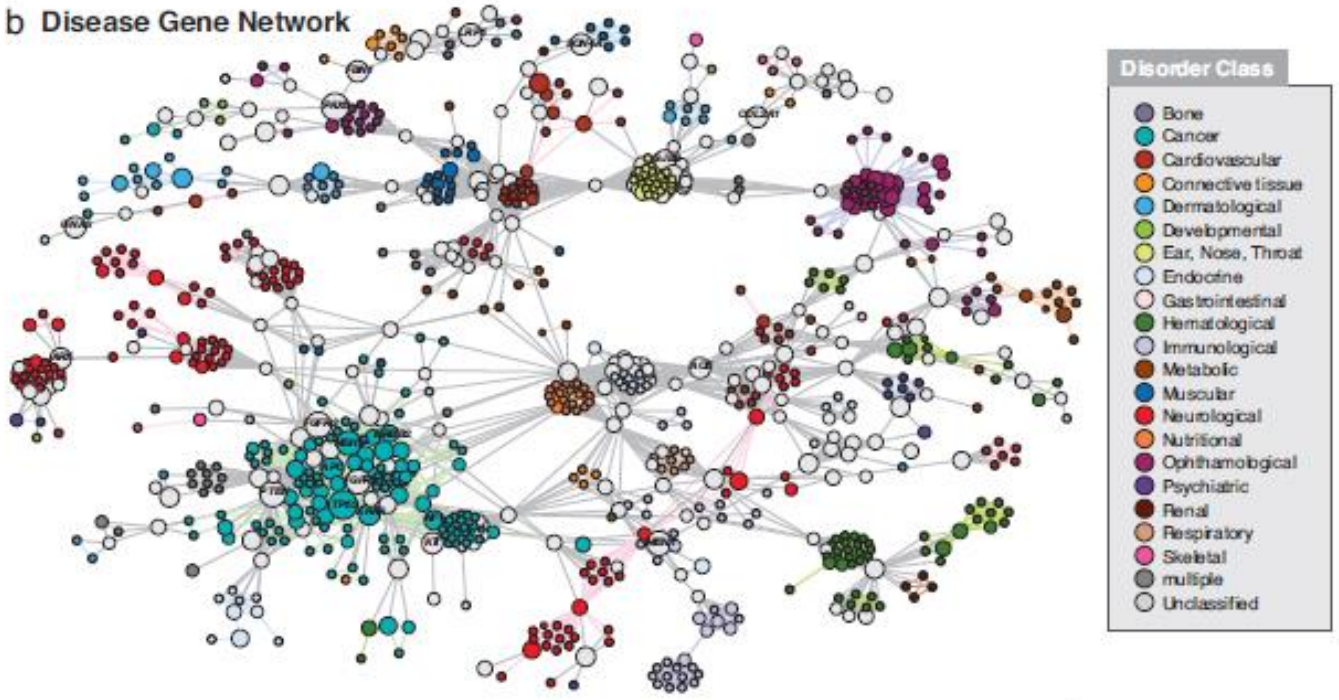
疾患
ネットワーク
(HDN)

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



疾患遺伝子
ネットワーク
(DGN)

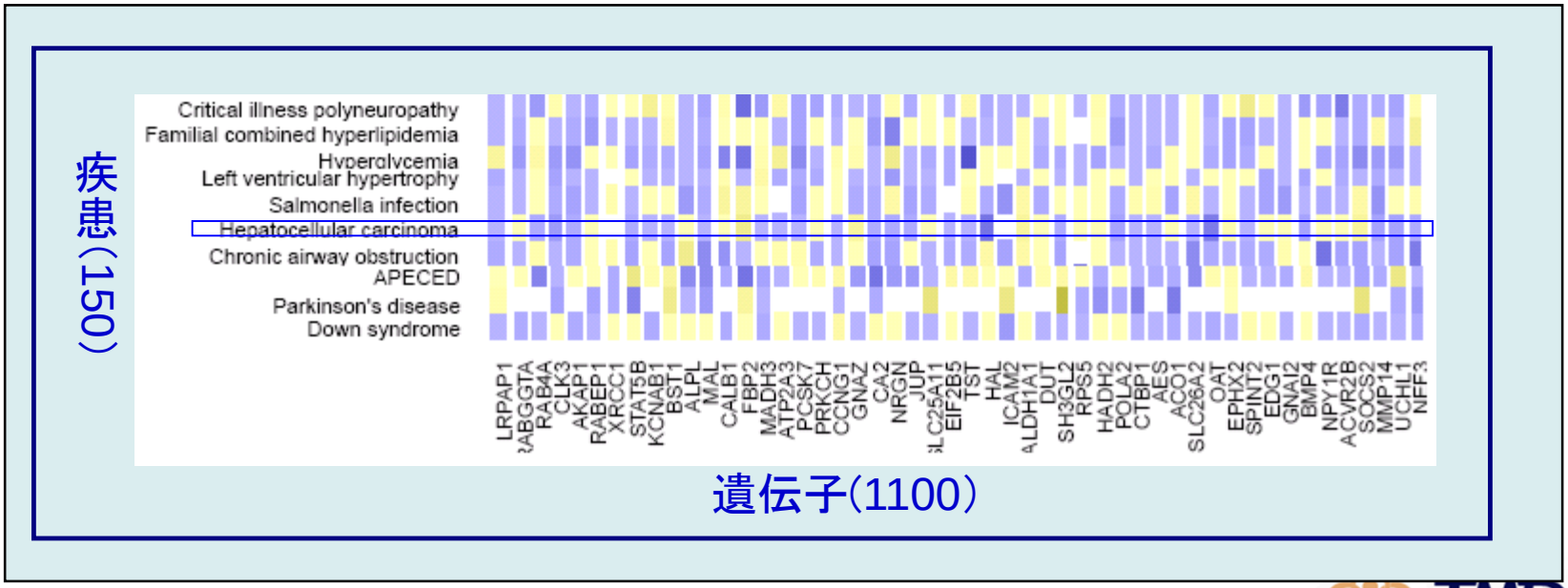
Nodeの径
その遺伝子を原因にしている疾患の数に比例
2つ以上の疾患に関与していると明灰色の遺伝子ノード



第２世代型

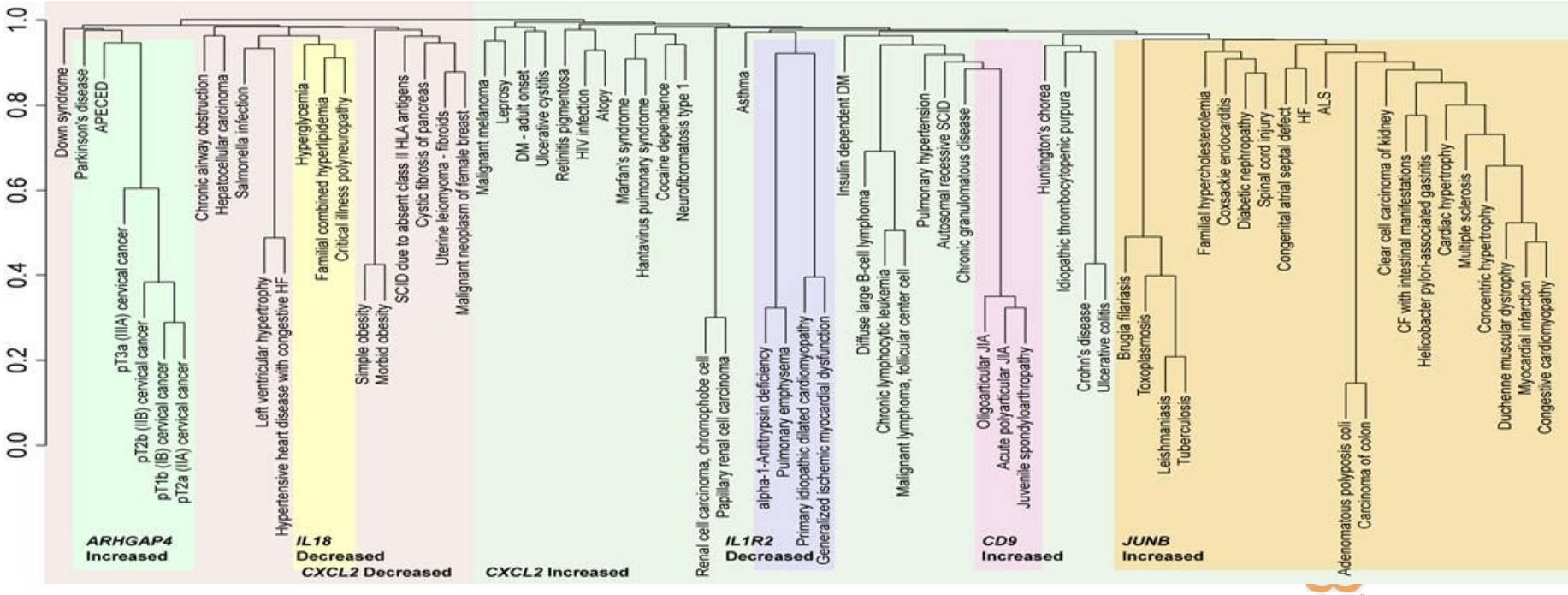
GENOMED (A.Butte et al)

- 遺伝子発現DBのGEO（Gene Expression Ominibus)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

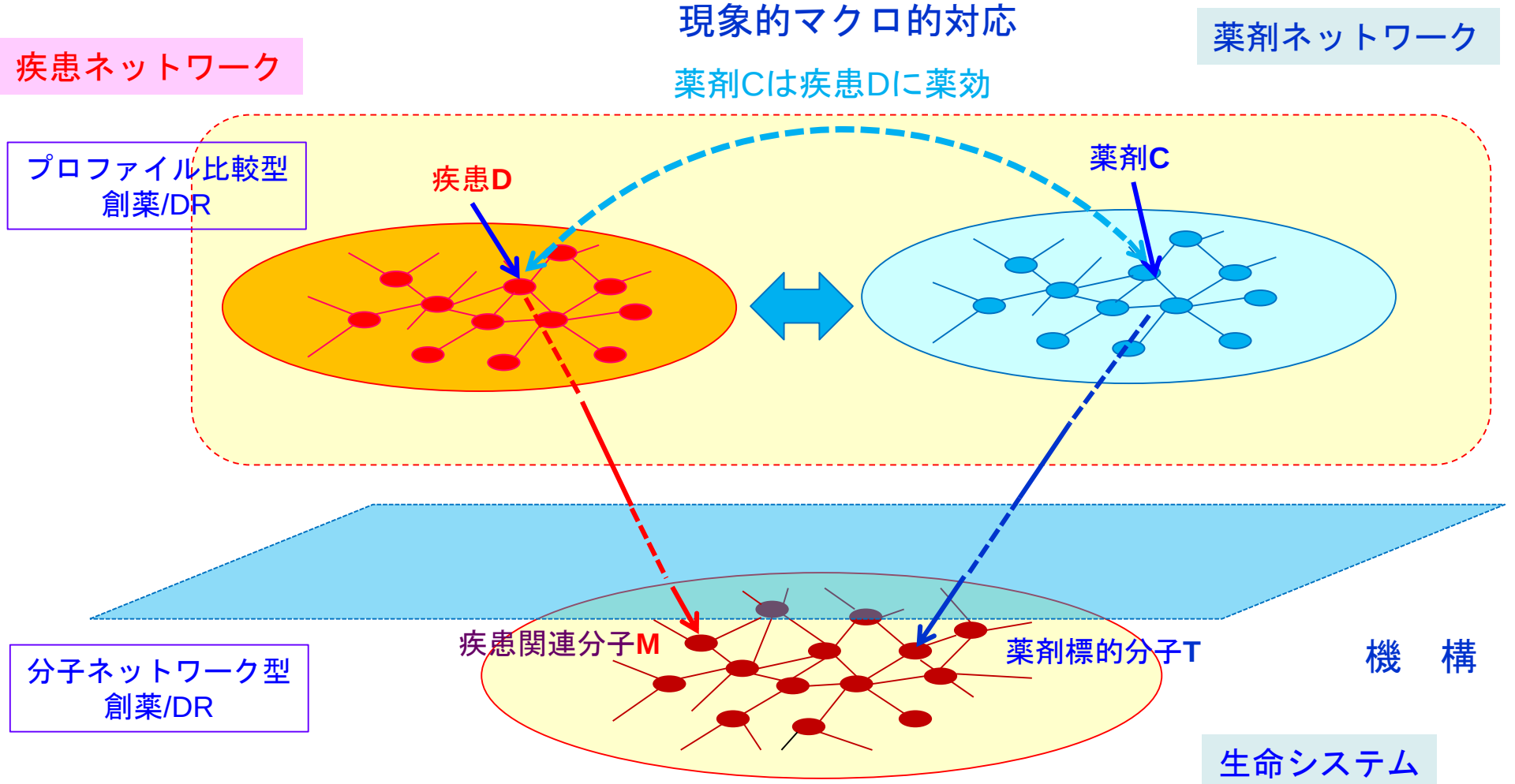
- 疾患を平均遺伝子発現パターンよりクラスター分類
  - 臓器別疾患分類では予想できない疾患間の親近性
  - 分類項目はサイトカインの遺伝子発現と相関
  - 疾患の再体系化に基づいた医薬の repositioning
- さらに656種類の臨床検査を結合した分析
- 心筋梗塞・デュシャンヌ型筋ジストロフィーが近い



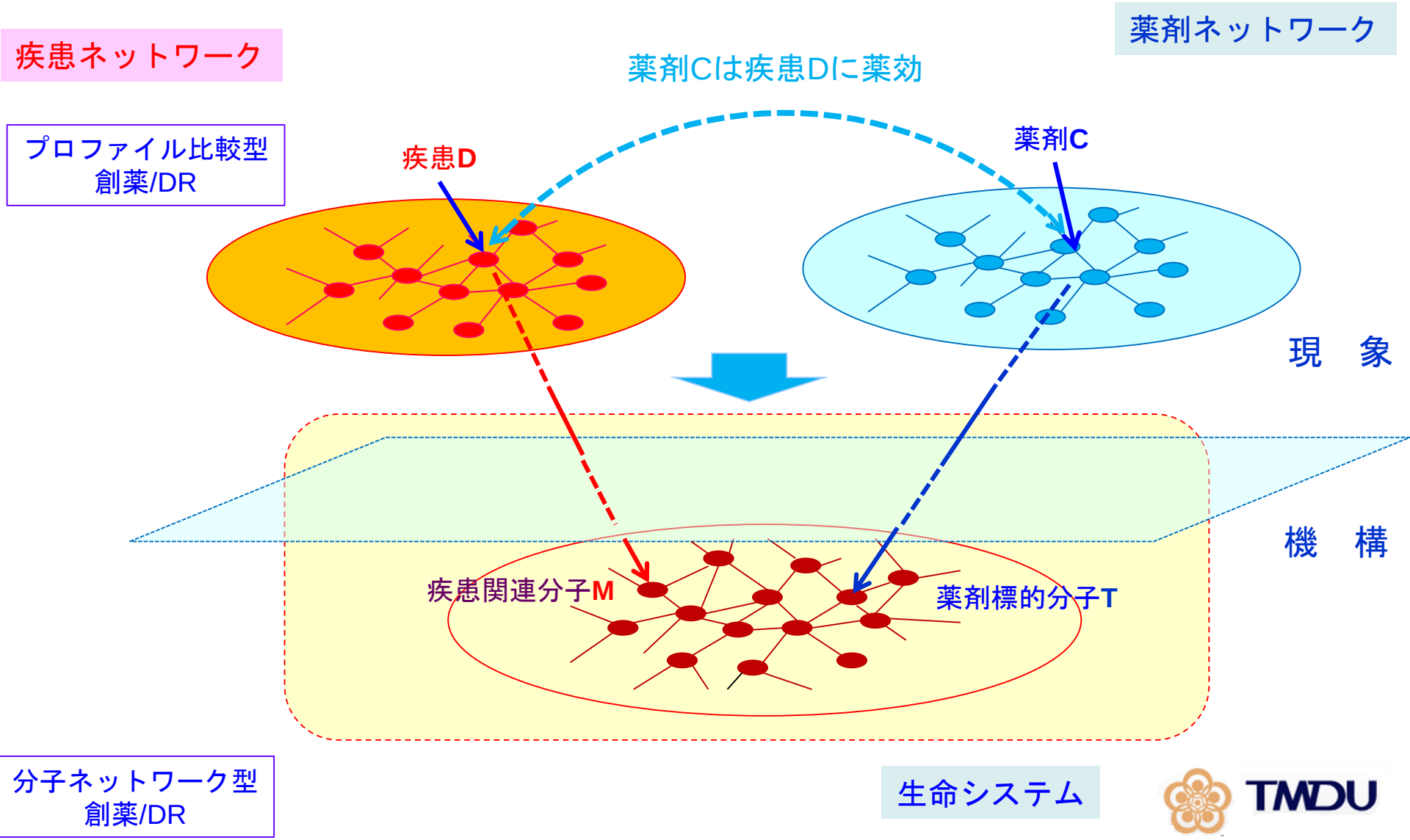
3. 階層的ネットワークによる創薬/DR

＜疾患-薬剤-標的＞の多層ネットワーク
生体分子ネットワークを基盤とする創薬・DR
ビッグデータ創薬/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

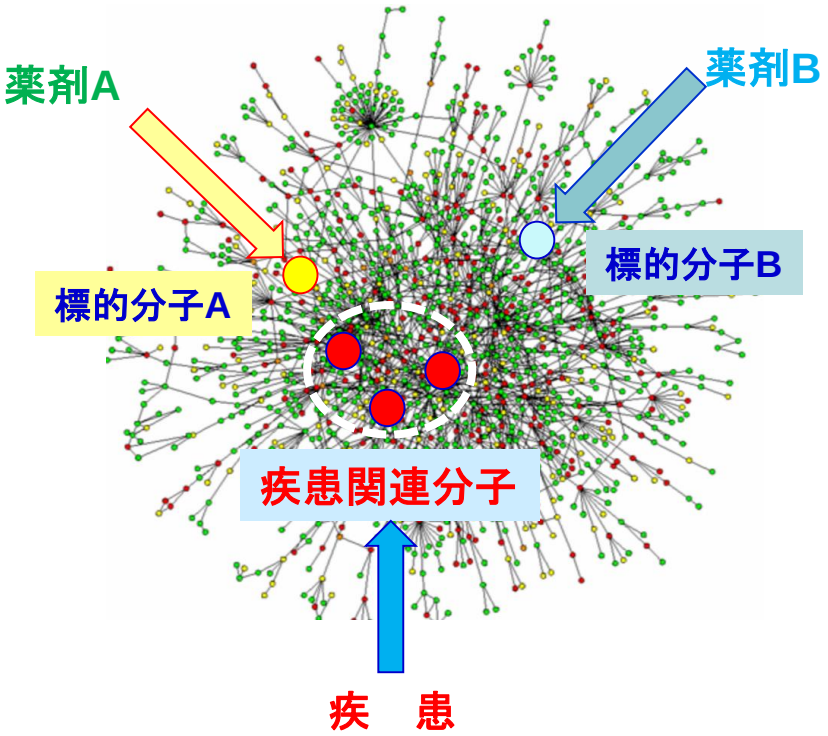


3層の生体・薬剤のネットワーク間の関係図式



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）

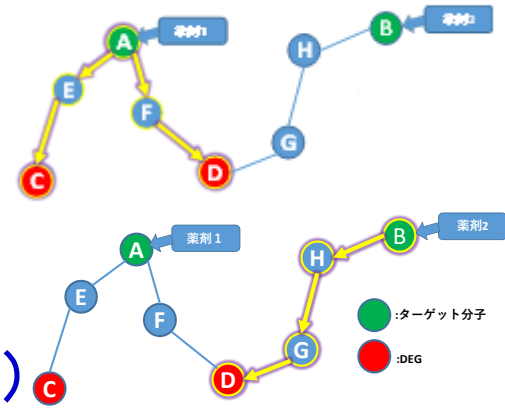
- 薬剤ネットワークと疾患ネットワークを媒介する第３の生体ネットワーク
- タンパク質相互作用ネットワーク（PPIN）での創薬/DR戦略
- PPIネットワーク場を基礎にして距離（類似性）を検討
- **薬　剤**：**薬剤の標的分子**（タンパク質）によって　PPI場と繋がる
- **疾　患**：**疾患特異的発現遺伝子**を疾患要因分子（**タンパク質**）へ翻訳、
- PPIN場内での**薬剤標的分子と疾患の「代理人(疾患遺伝子）」**の**距離・親近性**を基準に、薬理作用のインパクト力を評価



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 の薬剤ー標的分子組 （4,452薬剤, 1,617タンパク質）
- **疾患遺伝子の差異的遺伝子発現データ（DEG）**
  - TCGA (The Cancer Genome Atlas)より差異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 差異的発現遺伝子**
- **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

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

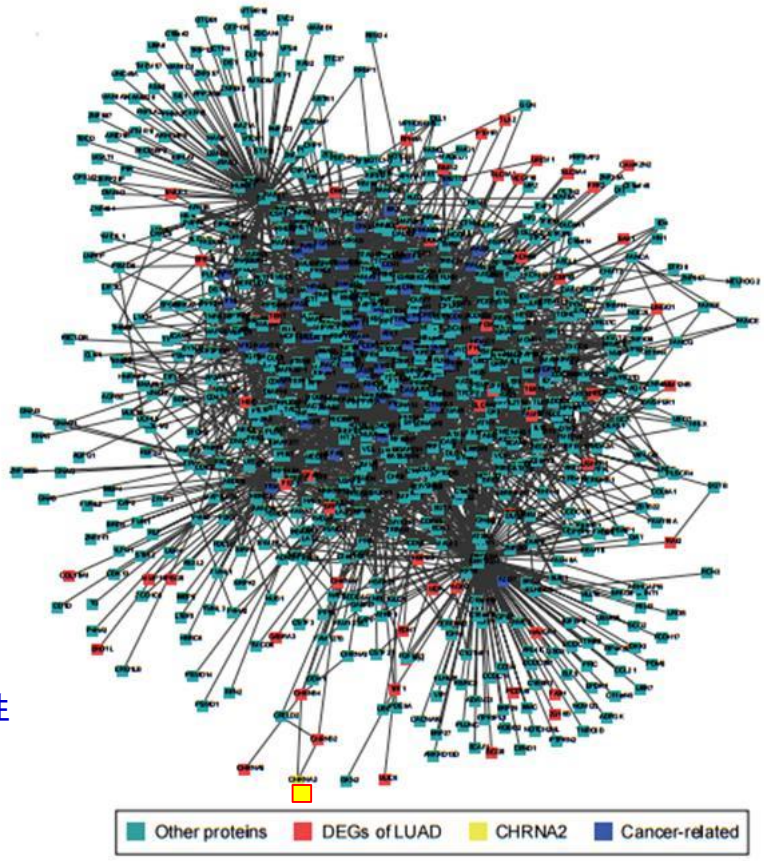
$\mathbf{P}^{s}$：時点sでの各ノードでの滞在確率　$\mathbf{M}$ ：各ノードへの遷移確率　γ：再出発確率

TMDU

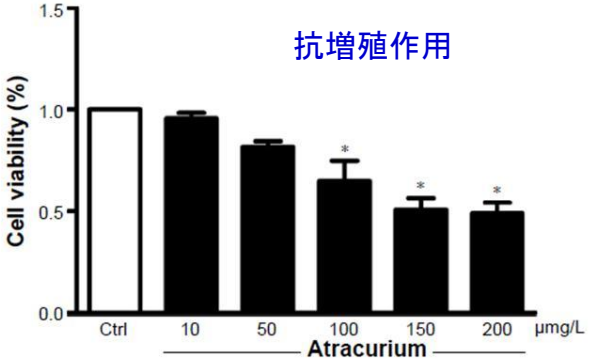
Interactome DR 結果の検証

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

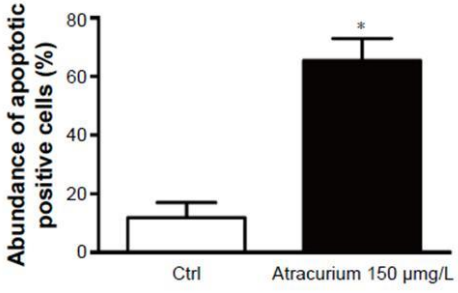
- **HPRD**と**STRINGS**の両方のPPINのランダム歩行でtop5%で**共通**な**145薬剤**を同定
- 最高スコアを挙げた**Atractium**を選択
- 薬剤標的は**CHRNA2**(Cholinergic Receptor Nicotinic Alpha 2）**でアポトーシス経路である**
- **培養細胞A549**（ヒト肺胞基底上皮腺癌）胞）の抗増殖作用を確認



抗増殖作用



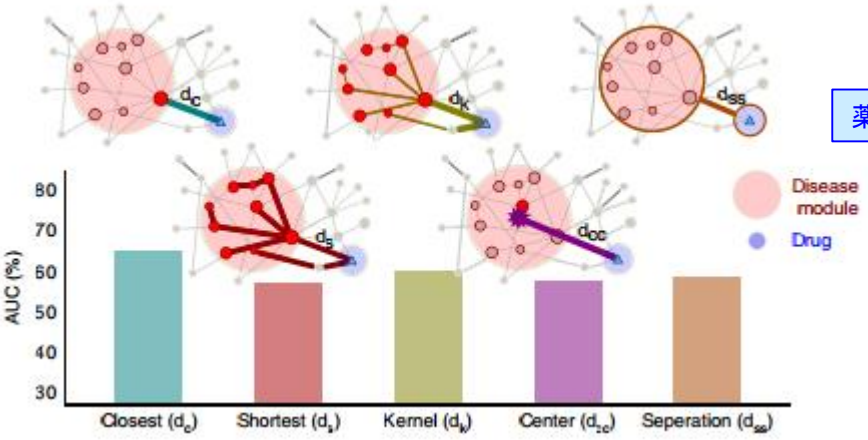
アポトーシス活動性



TMDU

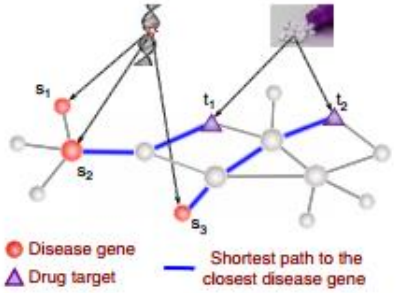
タンパク質相互作用ネットワークでの近接性によるDR

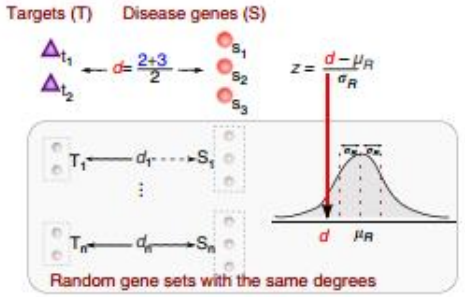
相対近接指標d_c：
①最近接の疾病関連分子との最短経路長の平均
②同じサイズで度数の分布より近接指標を計算して規格化⇒zスコア
（z < - 0.15⇒近接）
②様々な近接指標の中ではclosest measure d_c が一番薬効を予測する

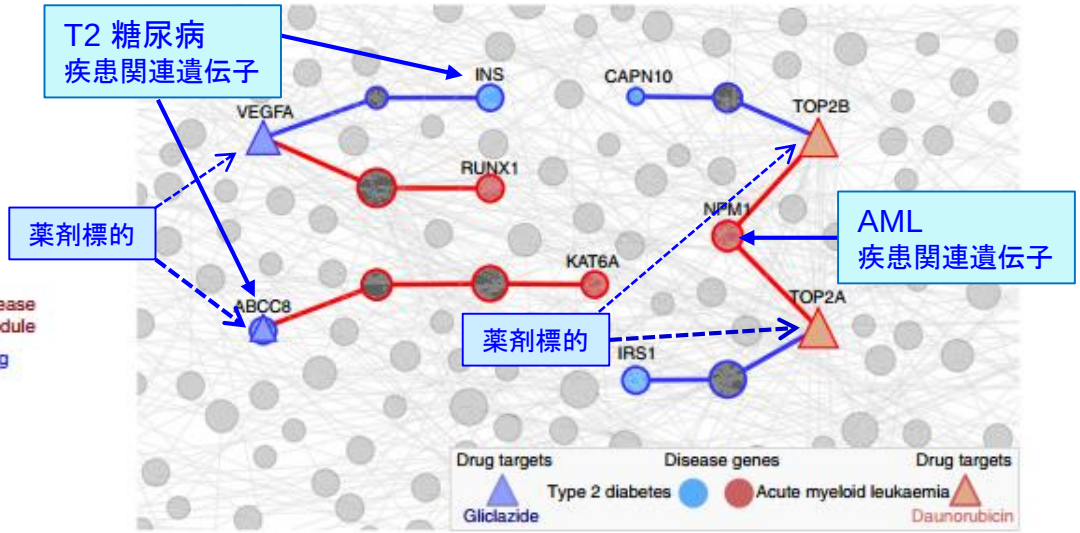


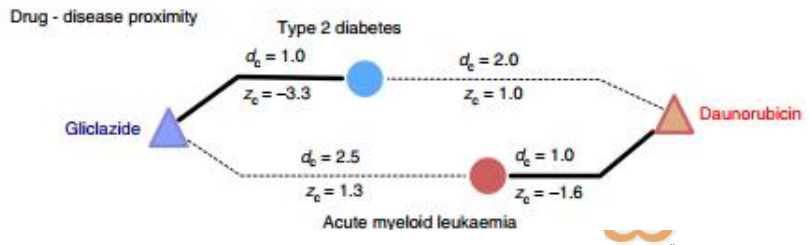
大半の薬剤は標的と疾患関連分子
２リンク離れている

(Guney, Barabasi,2016, Nat. Com)



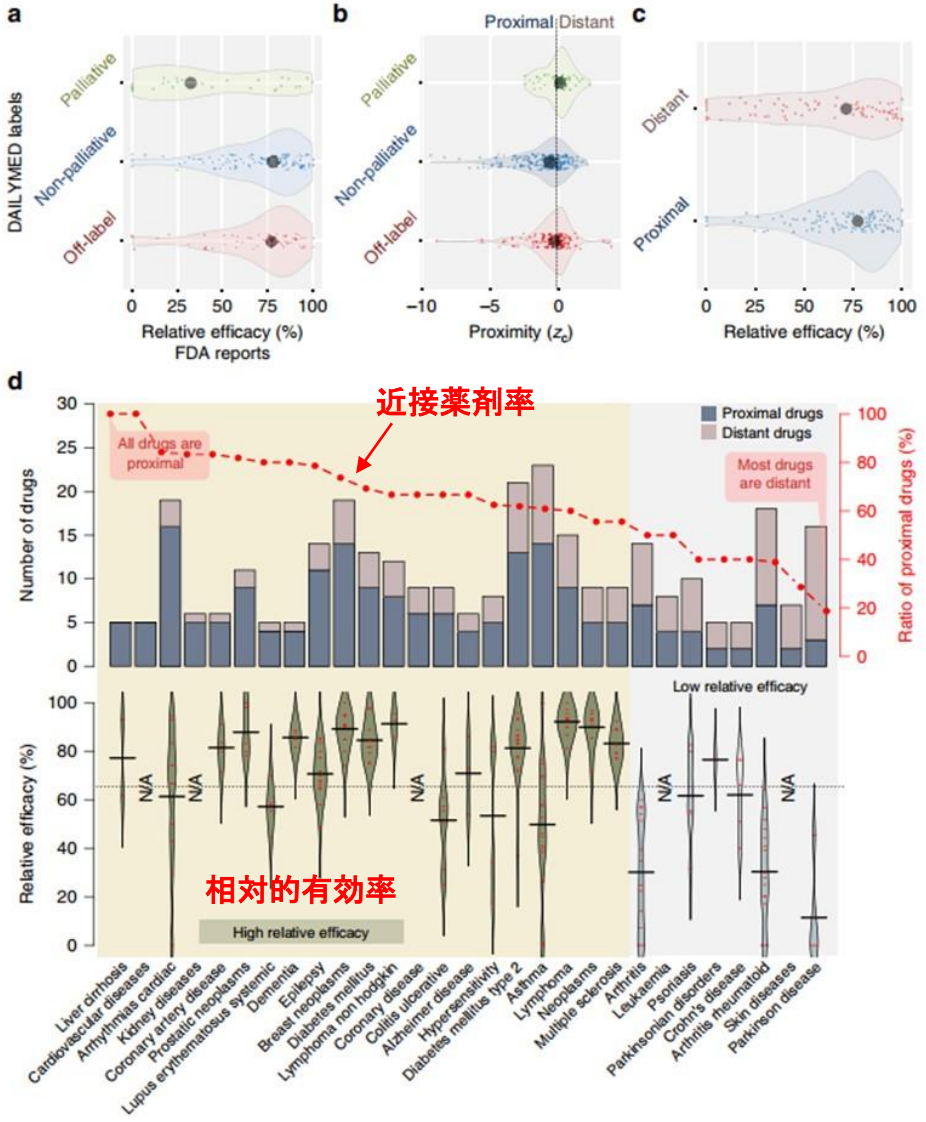






相対近接性による薬効予測

- 疾患モジュールの内部/近接に標的分子を持つ必要がある
- これまでの研究では疾患関連分子と標的分子の距離が大
  - 対症療法・緩和療法：疾患原因ではなく症状を標的としている
  - 標的分子が疾患関連分子の数は少ない（402対のうち62）
- 既成の薬は疾患と近接的である
- 緩和療法は遠隔的である
- Off-labelは緩和より近接的である
- 近接薬剤の治験の頻度は高い
- 薬剤は選択的であるが排他的ではない
- 相対的有効性と近接指標は相関する
- 平均の標的分子の数は3.5個である

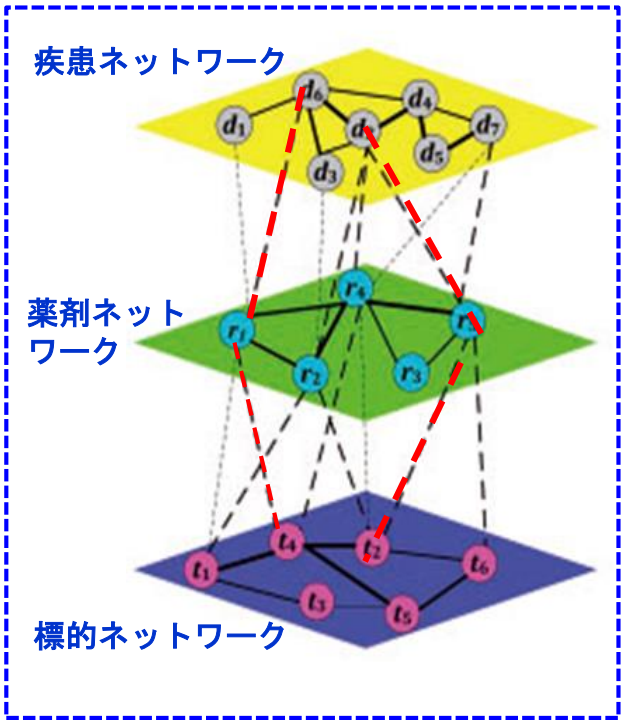


３階層生命ネットワークでの創薬/DR

- 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

創薬/DRとは

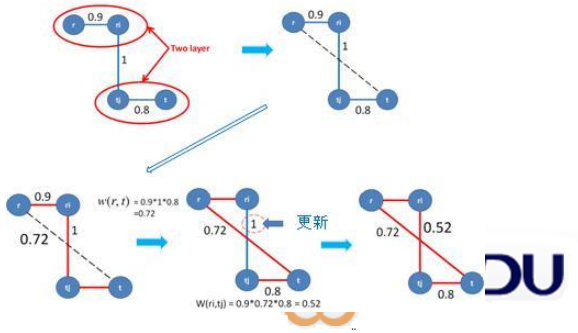
未発見の階層間リンクを
既存の階層間リンクの事実と
各層のネットワークから推測



(Wang et al. 2014)

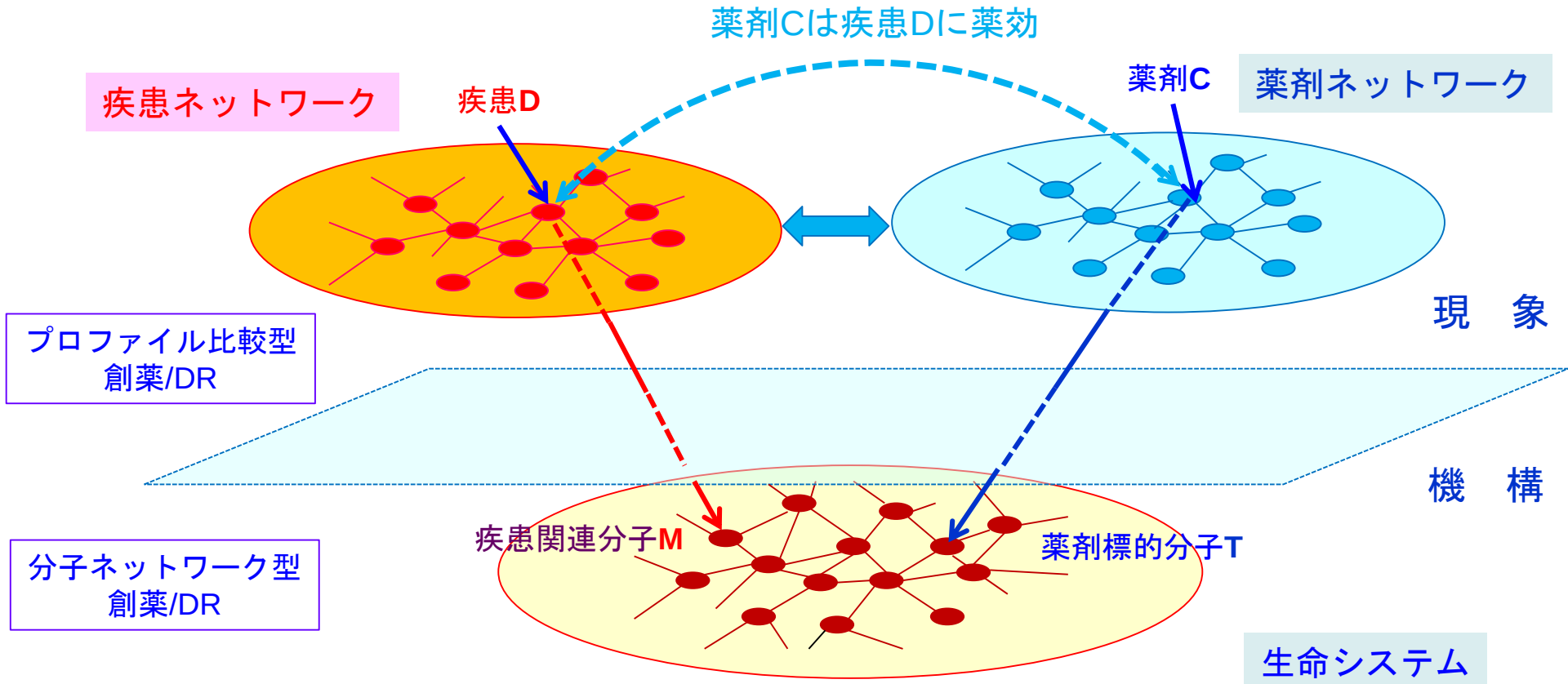
Wang et al. 2014は

- 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



プロファイル型計算創薬の原理

3層生体・薬剤ネットワークのFramework



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

期待消滅！

1990

しばらく停滞！

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

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

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

自己紹介と医療人工知能の歴史

人工知能(AI)を医療・創薬へ応用

田中　博
東京医科歯科大学
生命医療情報学
東北大学
東北メディカル・
メガバンク機構

1980から1995
第１期の
AIブームの時
医療AI研究に従事

私の人生

東大医学部

医療AI研究

東京医科歯科大

ゲノム医療

BigData医学
AI創薬

記号知識処理

問題解決の探索法（GPS）

医学「知識」を計算機に格納
医療診断システム（MYCIN）
知識工学：大ブーム
政府：第５世代コンピュータ
知識の移植問題

ブーム消滅！

医療機械学習
診断知識のDBからの学習

診療支援
医学の用語や
概念体系の基礎理論

ニューロネット(NN)

単純NN
パーセプトロン
判別能力の限界

1970
以前

1980

多層NN
バックプロ
パゲーション
重み修正の限界

1990

ブーム消滅！

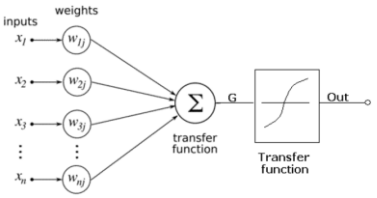
2000

Deep Learning
多層NN
「教師なし」特徴学習

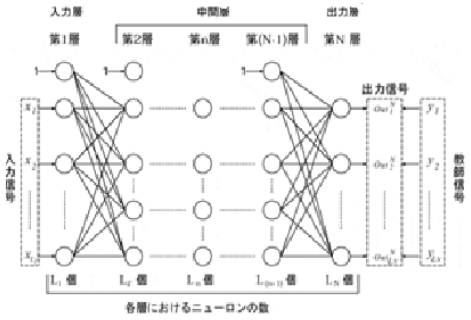
Deep Learning 型人工知能の革命性

Deep Learning による人工知能革命

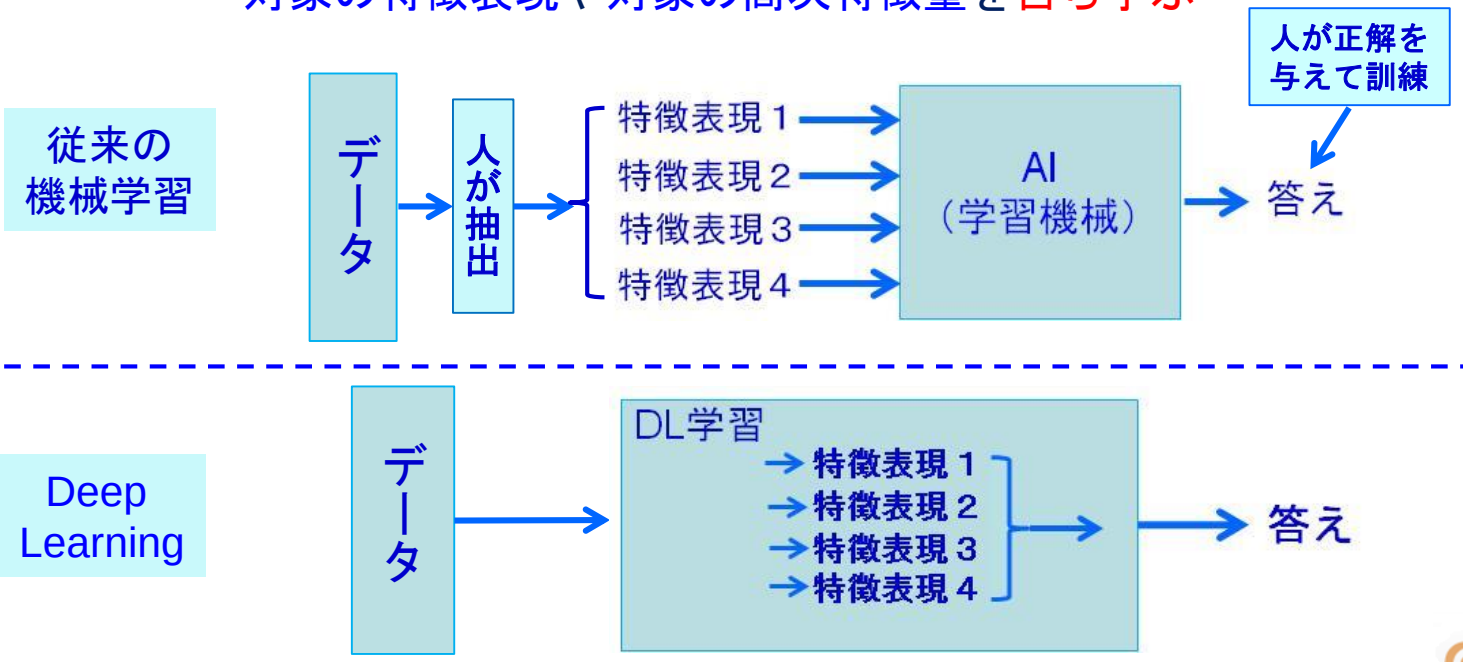
- 機械学習のこれまでの限界
  - 「教師あり学習」
    - 分類対象の特徴と正解を与え学習機械（AI）を構築
- Deep Learningの革命性
  - 「教師なし学習」
    - 対象の特徴表現や対象の高次特徴量を自ら学ぶ



神経情報素子

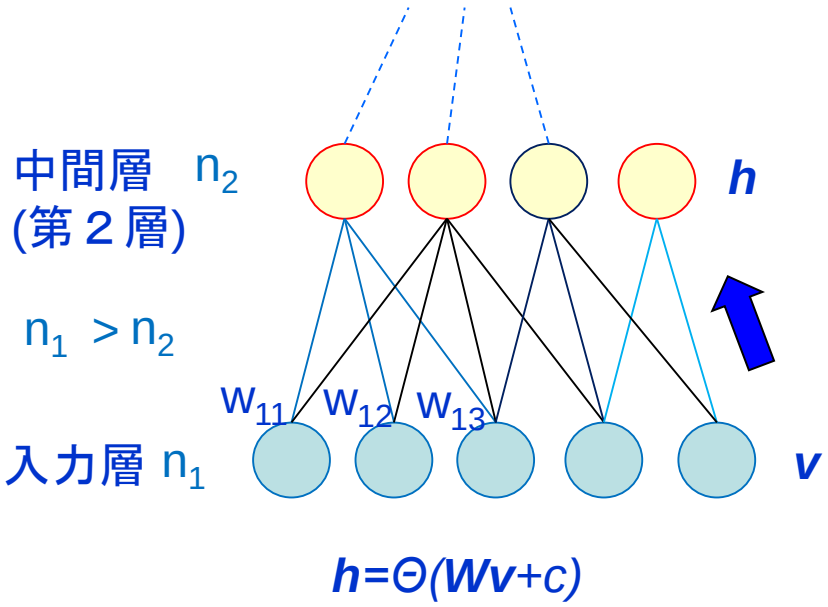


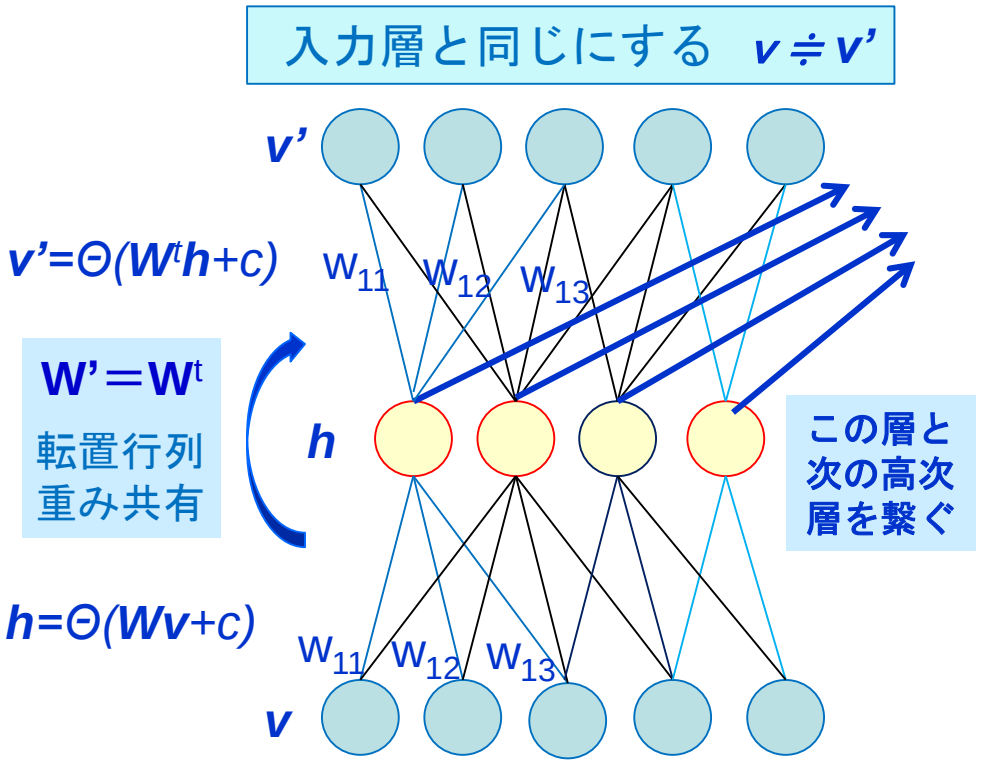
多層ニューロネットワーク



DLの革命点 Autoencoder 1

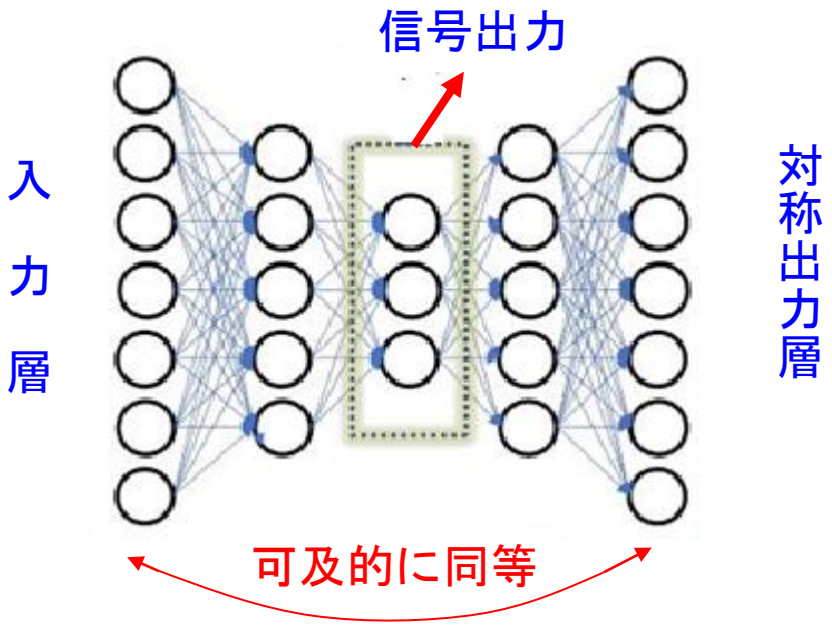
- 対象に固有な内在的特徴を学ぶ自己符号化の原理
- 格段ごとに入力の少ない中間層を入力へ逆投影して復元できるか
- 次元を圧縮され可及的に復元する（1000 nodes ⇒100nodes=?⇒1000nodes）
  →できるだけ復元に効果的な特徴量を探索する
  →内在的な特徴量を見出す





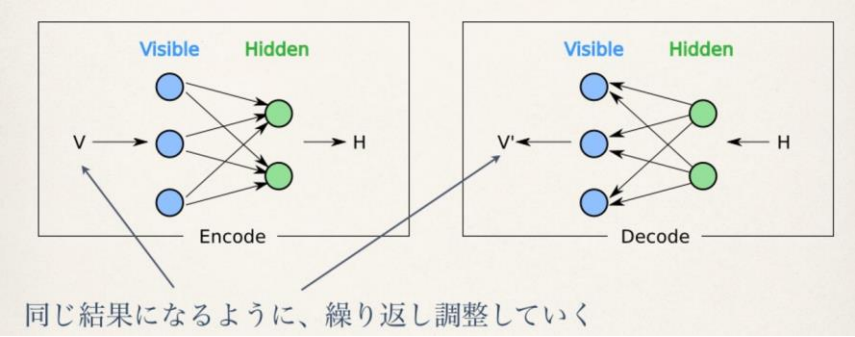
DLの革命点 Autoencode２

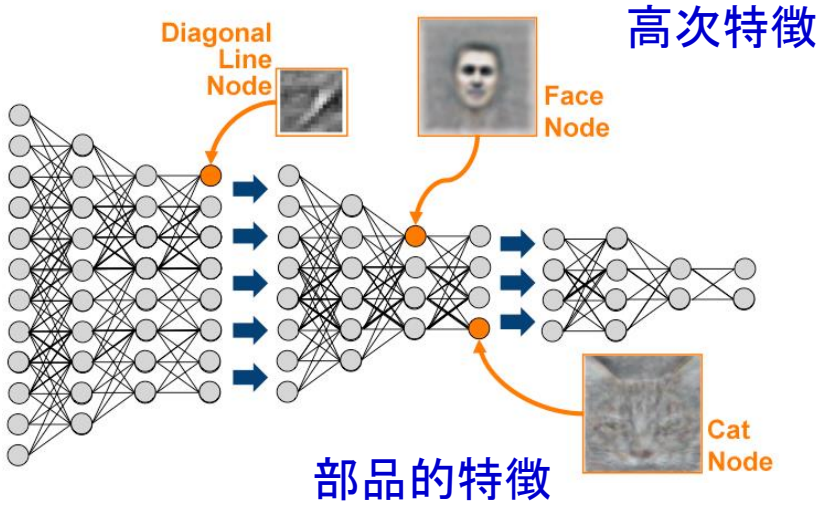
- 自己符号化器を多層に構成する
  - 積層自己符号化器（stacked autoencoder）
- 入力層と出力層を対称に層構成する
  - 深層自己符号化器（deep autoencoder）



DLの革命点 Autoencoder 3

- 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**





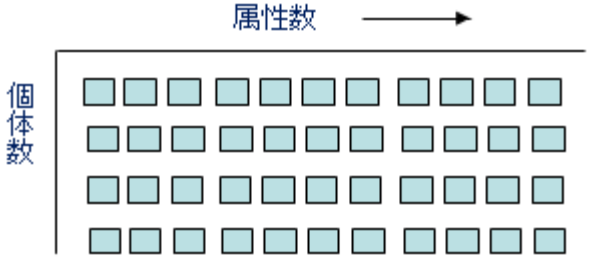
Deep Learningの医療・創薬へ応用

TMDU

「ビッグデータ」のData 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＜網羅的分子情報ー臨床表現型情報＞
  - ゲノムコホートにおける＜遺伝子情報ー環境（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

TMDU

タンパク質相互作用ネットワークでの疾患-薬剤-標的分子の学習

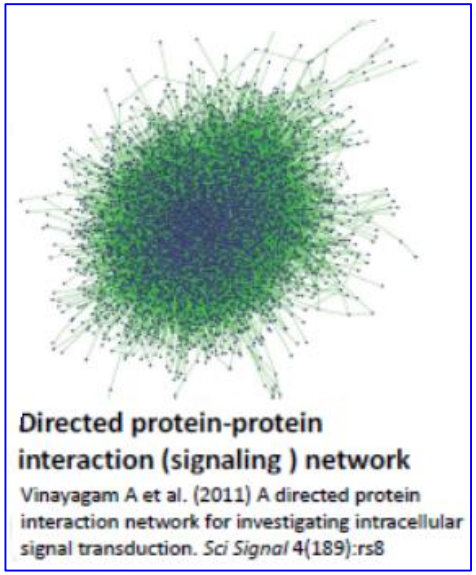
- ビッグデータ創薬/DR
  - タンパク質相互作用ネットワーク上での有効性予測
  - 基準指標：疾患関連分子と薬剤標的分子の距離
  - ネットワーク上のランダム歩行による総合距離（Sun, 2015）
  - 疾患関連遺伝子モジュールと標的分子の標準化近接指標
  - 判定情報量が不足
- AI創薬/DR
  - ビッグデータ創薬/DRの限界（情報の不足）をAI学習で補完
  - 既成の疾患-薬剤-標的分子の正例を学習（DrugBank)
  - 疾患関連分子と標的分子のタンパク質相互作用ネットワークにおけるトポロジカルな関係性を学習
  - 人工知能（AI）によって学習
  - 学習された疾患関連分子と標的分子の関係性のトポロジー特性により各分子の標的分子としての有効性を判定
  - 有力な標的分子を推測

特徴的ネットワーク基底への分解

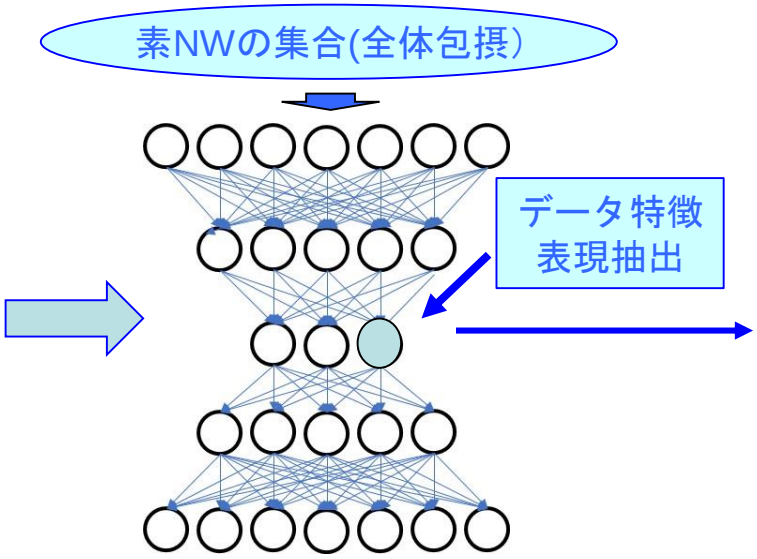
特徴的ネットワーク基底の和に縮約

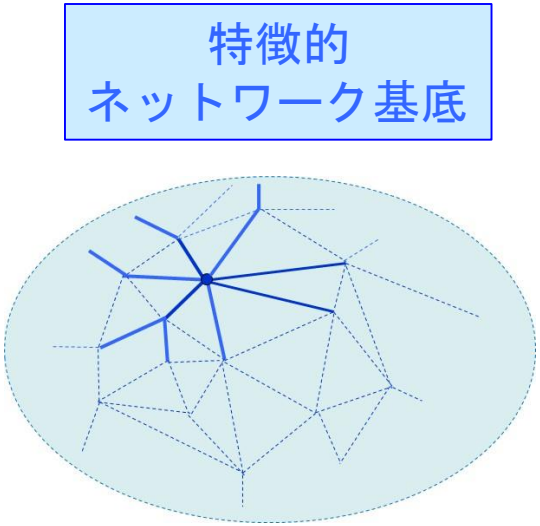
特定のノードを起点とした素NW（部分NW）の集合
全体NWを包摂する集合にDL反復自己学習

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





DL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

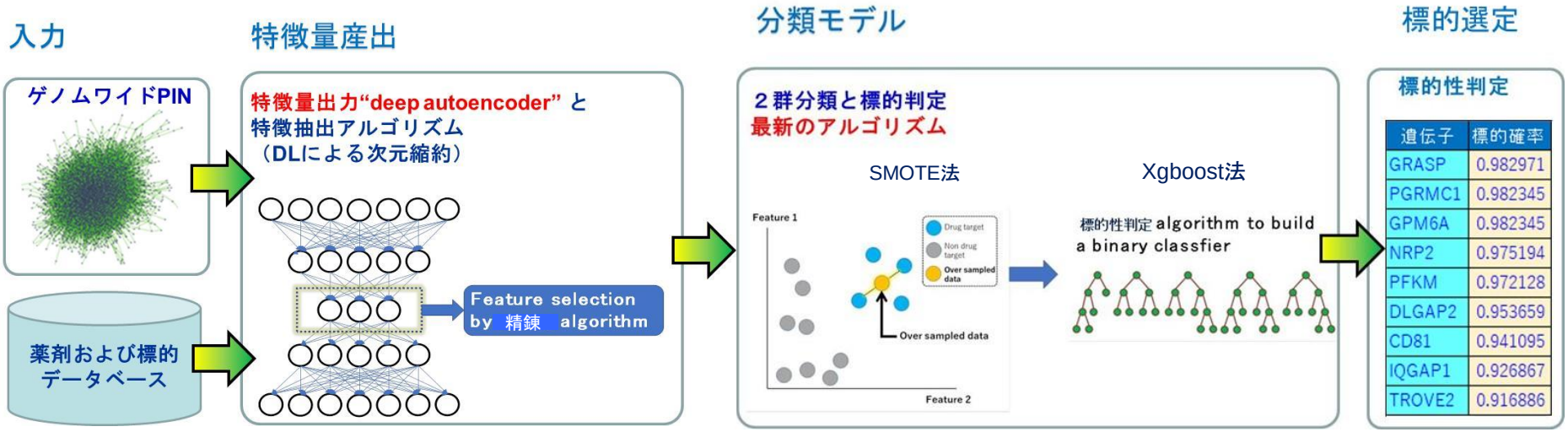
TMDU

DL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

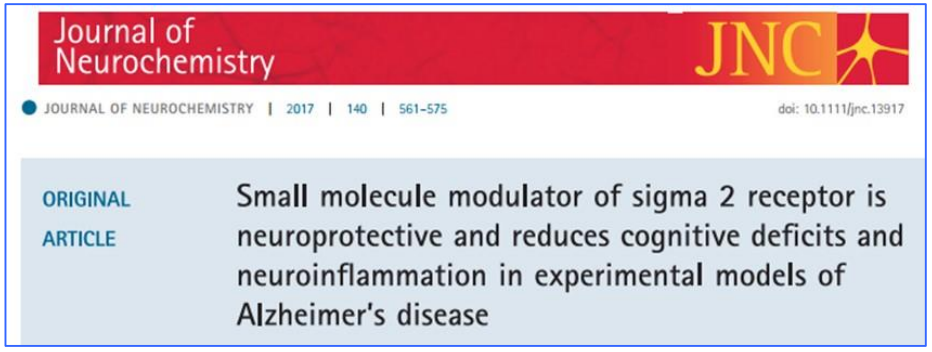


遺伝子	標的確率
GRASP	0.982971
PGRMC1	0.982345
GPM6A	0.982345
NRP2	0.975194
PFKM	0.972128
DLGAP2	0.953659
CD81	0.941095
IQGAP1	0.926867
TROVE2	0.916886

従来の機械学習（Random Forrest)と同じ成果は得られている

実験的研究との付合 1

PGCM1：progesterone receptor membrane 1

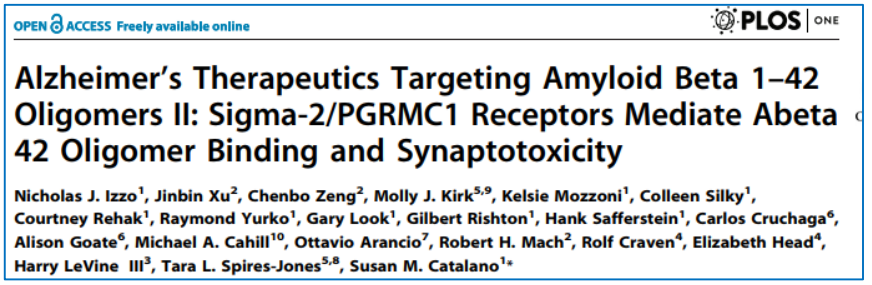

Journal of Neurochemistry

JNC

JOURNAL OF NEUROCHEMISTRY | 2017 | 140 | 561–575 doi: 10.1111/jnc.13917

ORIGINAL ARTICLE

Small molecule modulator of sigma 2 receptor is neuroprotective and reduces cognitive deficits and neuroinflammation in experimental models of Alzheimer's disease

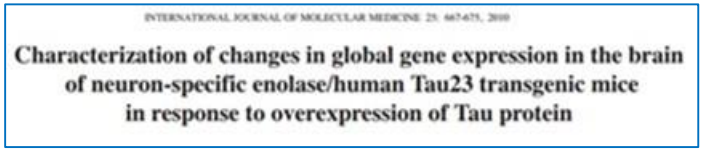

OPEN ACCESS Freely available online

PLOS ONE

Alzheimer's Therapeutics Targeting Amyloid Beta 1–42 Oligomers II: Sigma-2/PGRMC1 Receptors Mediate Abeta 42 Oligomer Binding and Synaptotoxicity

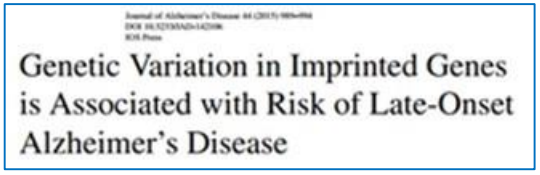
Nicholas J. Izzo[1], Jinbin Xu[2], Chenbo Zeng[2], Molly J. Kirk[5,9], Kelsie Mozzoni[1], Colleen Silky[1], Courtney Rehak[1], Raymond Yurko[1], Gary Look[1], Gilbert Rishton[1], Hank Safferstein[1], Carlos Cruchaga[6], Alison Goate[6], Michael A. Cahill[10], Ottavio Arancio[7], Robert H. Mach[2], Rolf Craven[4], Elizabeth Head[4], Harry LeVine III[3], Tara L. Spires-Jones[5,8], Susan M. Catalano[1]*

GPM6A：Glycoprotein M6A

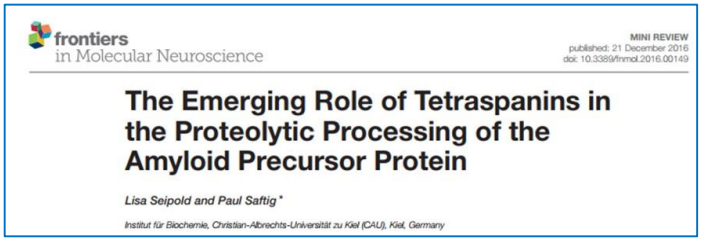

INTERNATIONAL JOURNAL OF MOLECULAR MEDICINE 25: 667-675, 2010

Characterization of changes in global gene expression in the brain of neuron-specific enolase/human Tau23 transgenic mice in response to overexpression of Tau protein

DLGAP2：DLG-Associated Protein 2


Genetic Variation in Imprinted Genes is Associated with Risk of Late-Onset Alzheimer's Disease

CD81:Tetraspanins family


frontiers in Molecular Neuroscience

MINI REVIEW

published: 21 December 2016

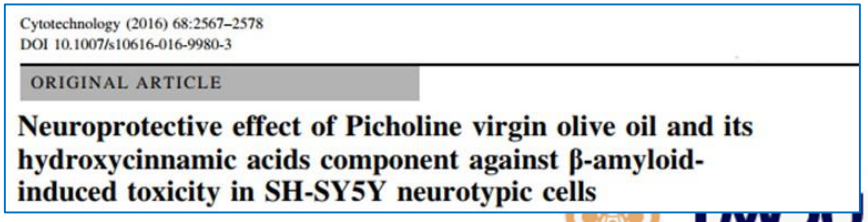
doi: 10.3389/fnmol.2016.00149

The Emerging Role of Tetraspanins in the Proteolytic Processing of the Amyloid Precursor Protein

Lisa Seipold and Paul Saftig*

Institut für Biochemie, Christian-Albrechts-Universität zu Kiel (CAU), Kiel, Germany

PFKM: Phospofructokinase


Cytotechnology (2016) 68:2567–2578

DOI 10.1007/s10616-016-9980-3

ORIGINAL ARTICLE

Neuroprotective effect of Picholine virgin olive oil and its hydroxycinnamic acids component against β-amyloid-induced toxicity in SH-SY5Y neurotypic cells

実験的研究との付合 2

WISP-2/CCN5：WNT1 inducible signaling pathway protein 2

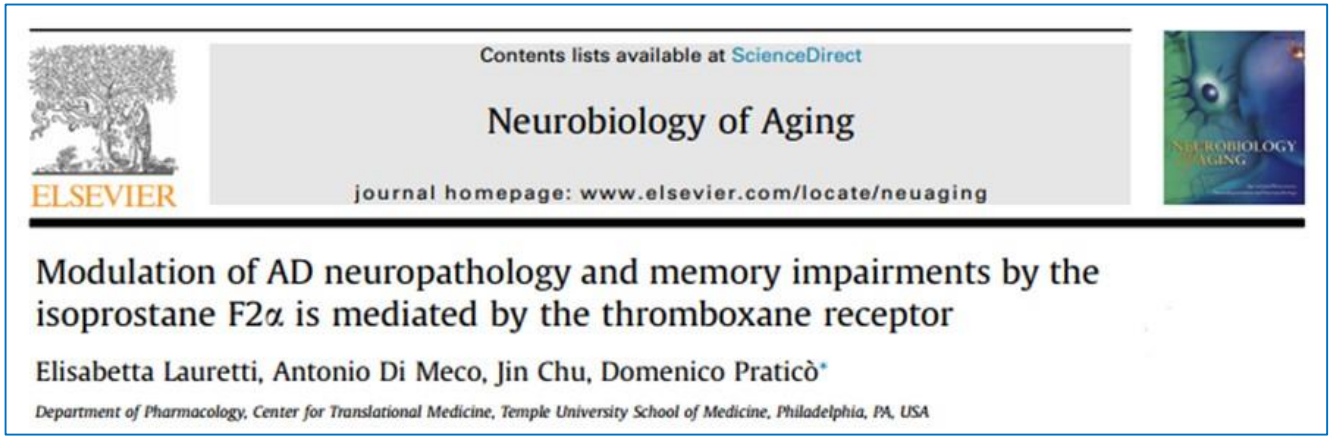
frontiers in Cellular Neuroscience

REVIEW
published: 24 June 2015
doi: 10.3389/fncel.2015.00237

Matricellular proteins of the Cyr61/CTGF/NOV (CCN) family and the nervous system

*Anna R. Malik, Ewa Liszewska and Jacek Jaworski**

TBXA2R: thromboxane A2 receptor



Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

ELSEVIER

Modulation of AD neuropathology and memory impairments by the isoprostane F2α is mediated by the thromboxane receptor

Elisabetta Lauretti, Antonio Di Meco, Jin Chu, Domenico Praticò*

Department of Pharmacology, Center for Translational Medicine, Temple University School of Medicine, Philadelphia, PA, USA

TMDU

DL型NNへの期待と困難点

- 医療・創薬の応用は大きく期待される
  - 本質的に「教師なし学習」:人間が思いつかない解を提示
  - **現状では**、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
    - 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
  - **期待される本質的な寄与**
  - **超多次元（生命医学）ネットワークから革新的知の発見**
- DL型ニューラルネットは困難点もある
  - 特徴表現を自己学習するが基本的にはBlack Boxで解析が必要
  - 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。

TMDU

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

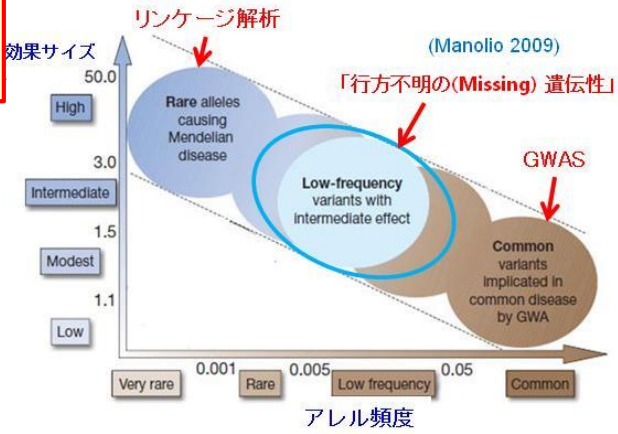
多因子疾患の機序/発症予測は無着手である

- 「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 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



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

環境によるエピゲネティク修飾

オランダ
飢饉 (1944)

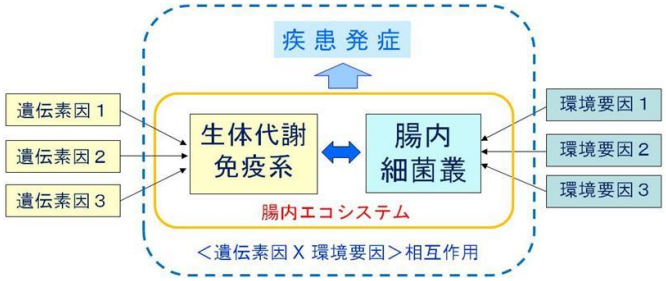
DOHaD(Developed Origin of Health and Diseas) 学説
オランダ飢饉のとき、母親の胎内にいた人々
出生30年後、肥満、糖尿病、心疾患、高罹患率
過度な低栄養：肝臓のPPARα/γ（倹約遺伝子）メチル化低下・遺伝子発現がオン
エピジェネティック変化は**可変**：短期的変化、**長期的「記憶」**次の世代も

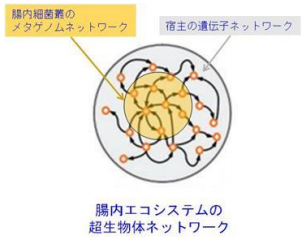
環境因子 → Epigenome**変化** ⇒ **遺伝子発現調節** → **疾病発症**

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
  - ビッグデータ医療への適応可能
  - ゲノム医療の〈網羅的分子情報ー臨床表現型〉の相関ネットワーク構造
  - バイオバンクの〈遺伝素因ー環境要因〉と発症
- AI創薬の「枠組み」実現方向は「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

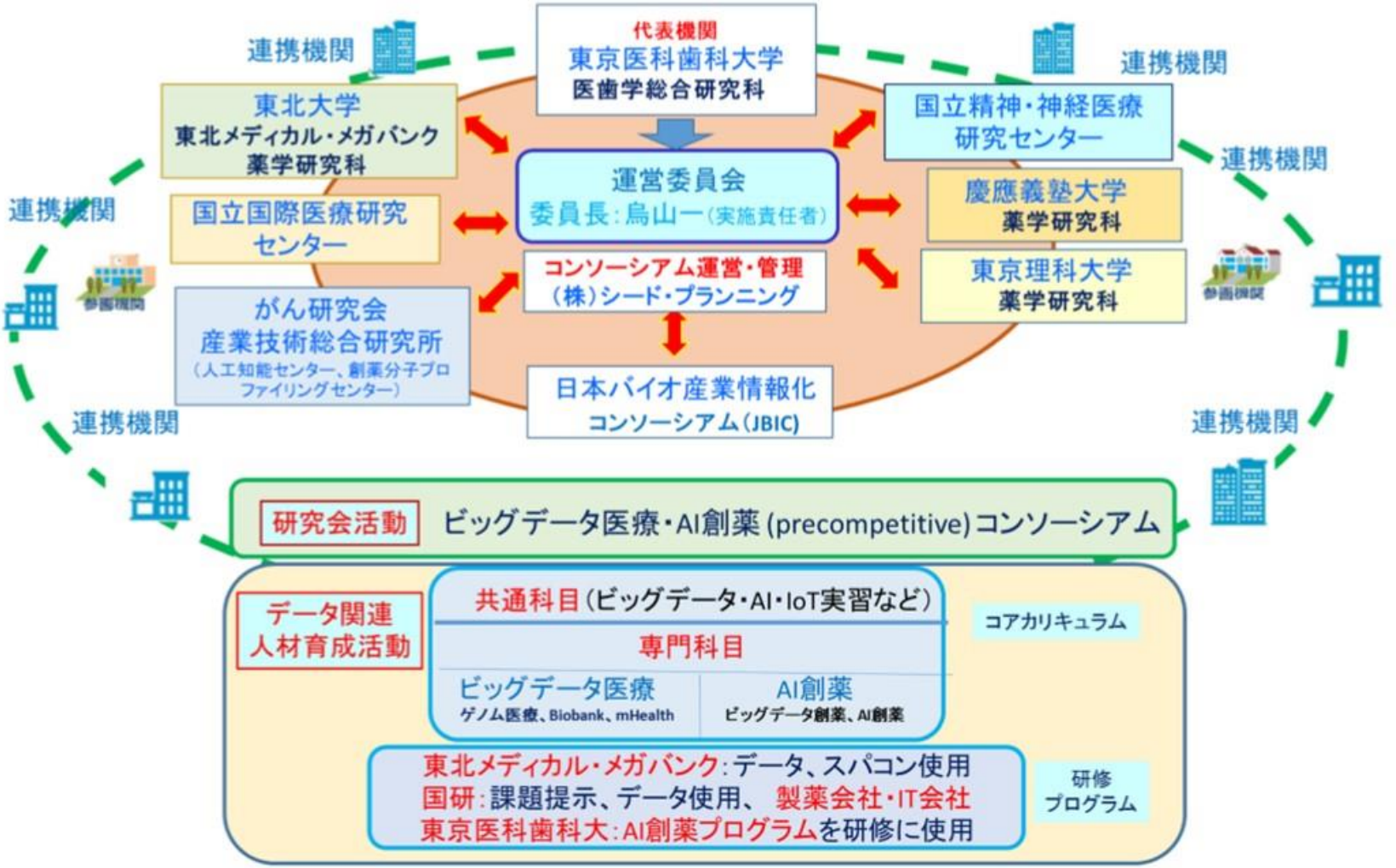
TMDU

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