

AI創薬

ビッグデータから人工知能を用いて
創薬・DRを行う

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

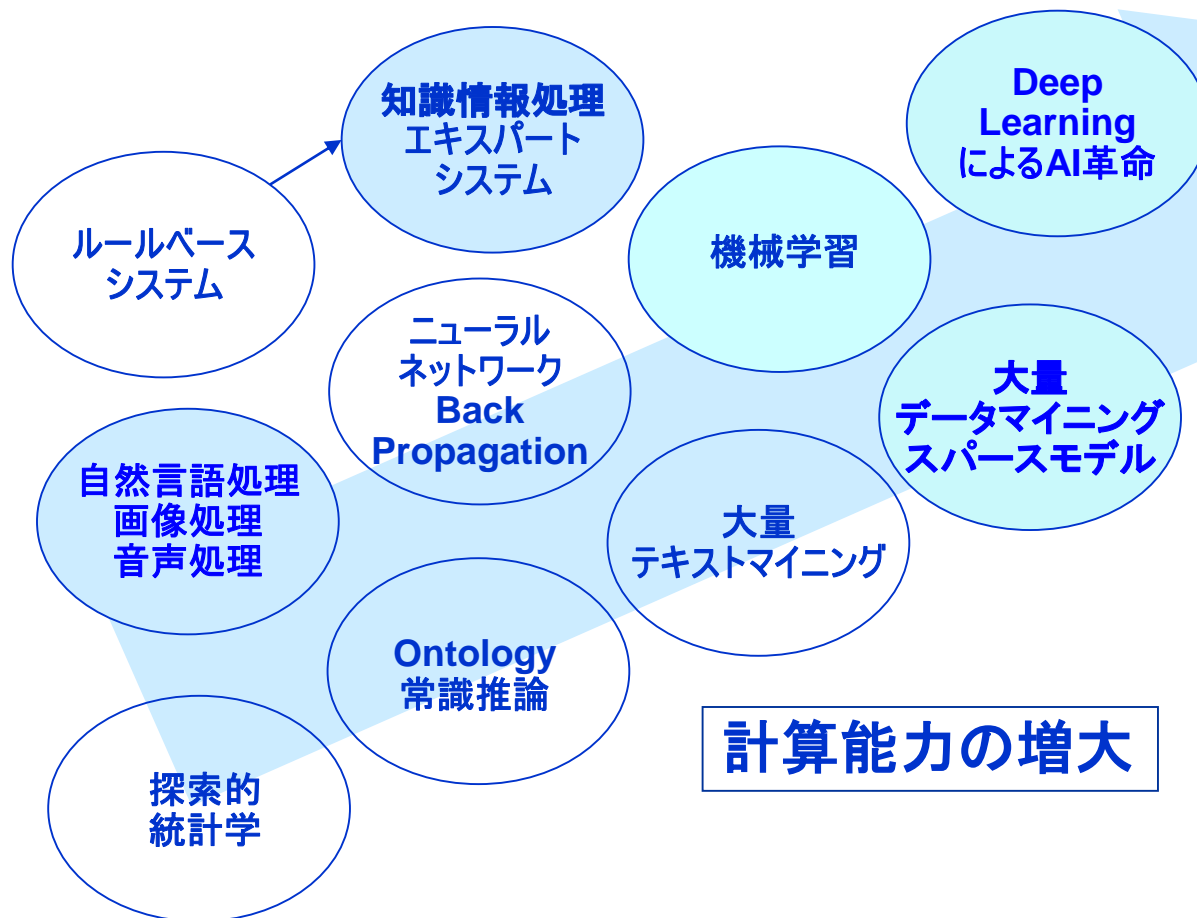


人工知能への期待

人工知能 (AI) の分野

データの増大

ビッグデータ
人工知能による
知的処理

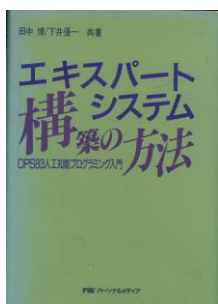


自己紹介と医療人工知能の歴史

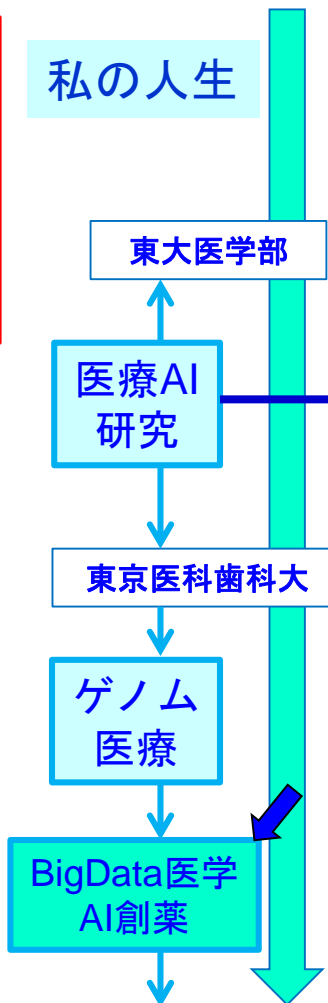
人工知能(AI)を医療・創薬へ応用

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



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

多層NN

バックプロ
 パゲーション
 重み修正の限界

1980

1990

ブーム消滅！

Deep Learning

多層NN
 「教師なし」特徴学習

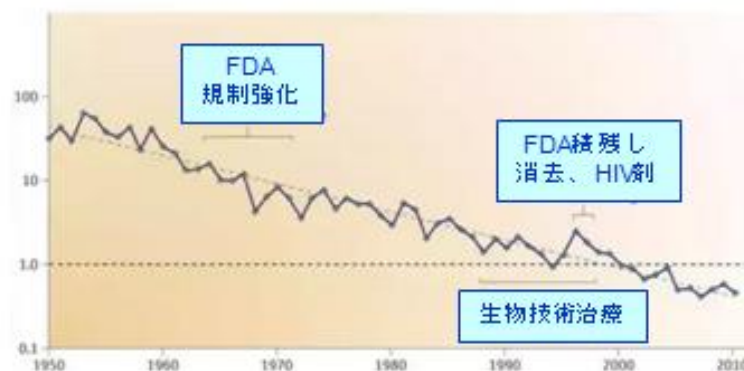
2000

ビッグデータ計算創薬

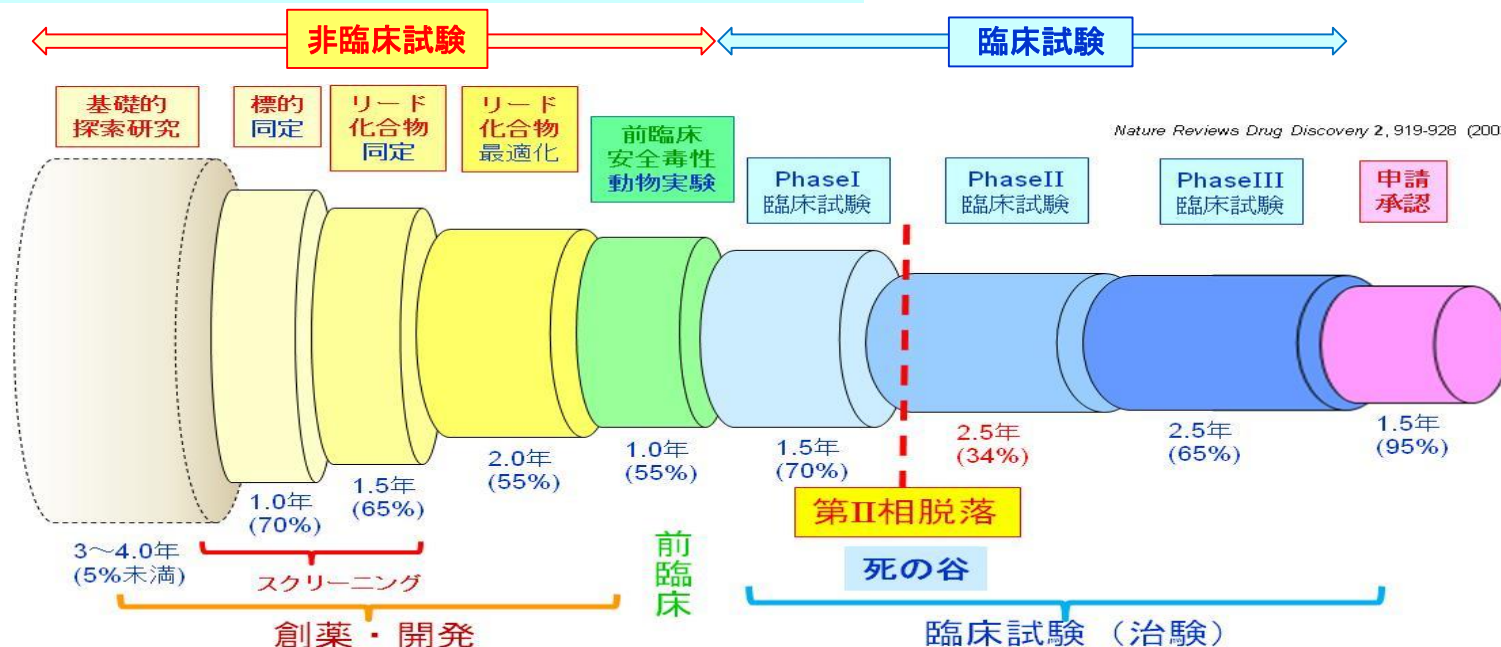
創薬をめぐる状況

- 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 11, 191-200 (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。

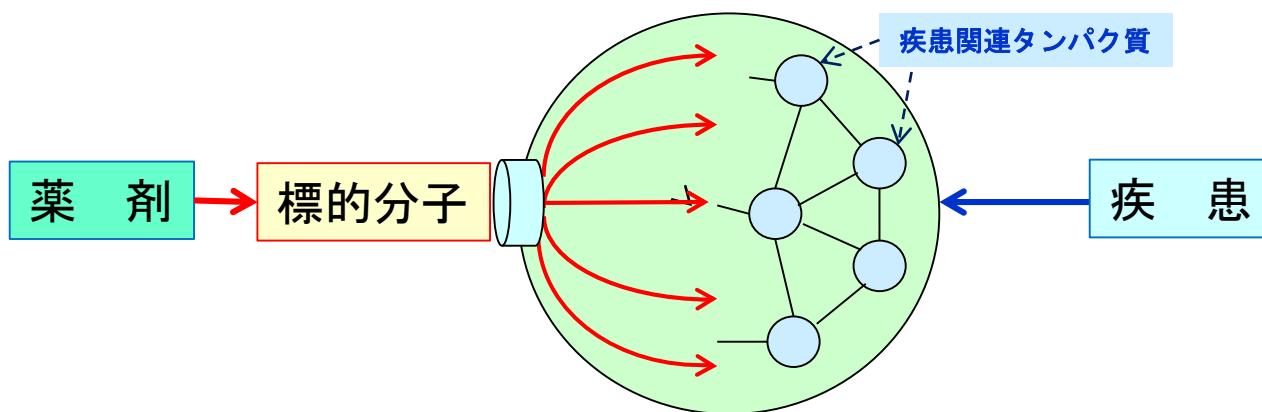
疾患・薬剤・標的の関係

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

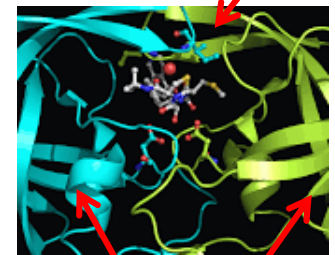
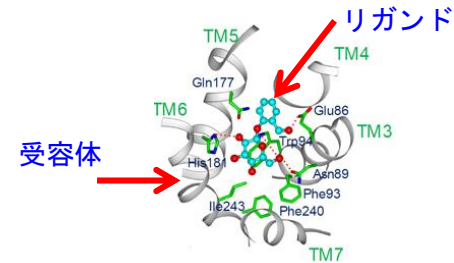
これまでの計算創薬

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

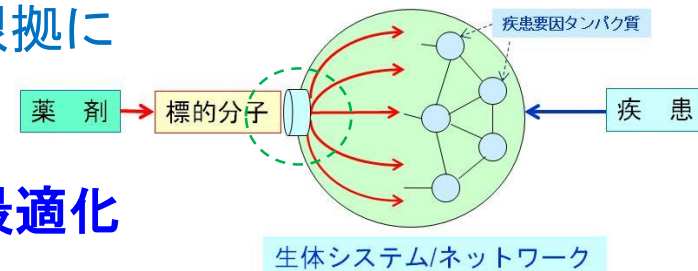
定量的構造活性相関(QSAR)

- 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



薬剤
(ポケット内)

酵素



構造(化合物)

生体システム

活性(反応)

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(*molecular profiling* 創薬)

疾患罹患状態における

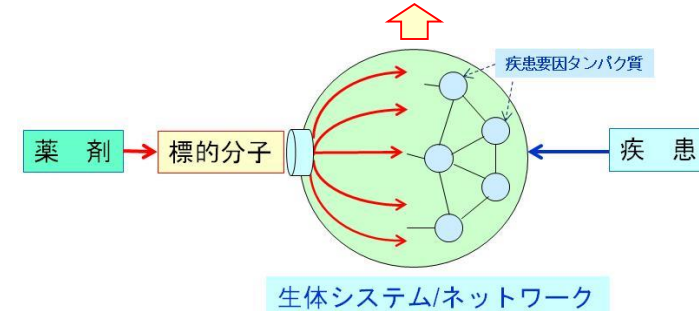
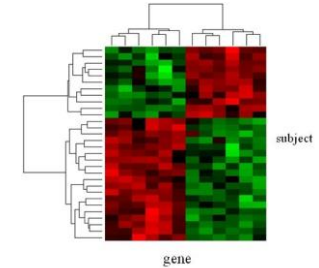
疾患要因遺伝子/タンパク質に起因し決定される
疾患時の生体のゲノムワイドな特異状態
疾患特異的な網羅的分子プロファイル変化状態



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>を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

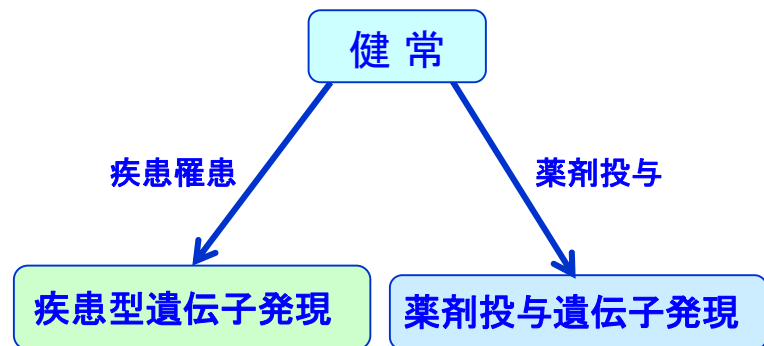
ビッグデータ計算創薬の基本原則

発現プロファイル創薬原理

- 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：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

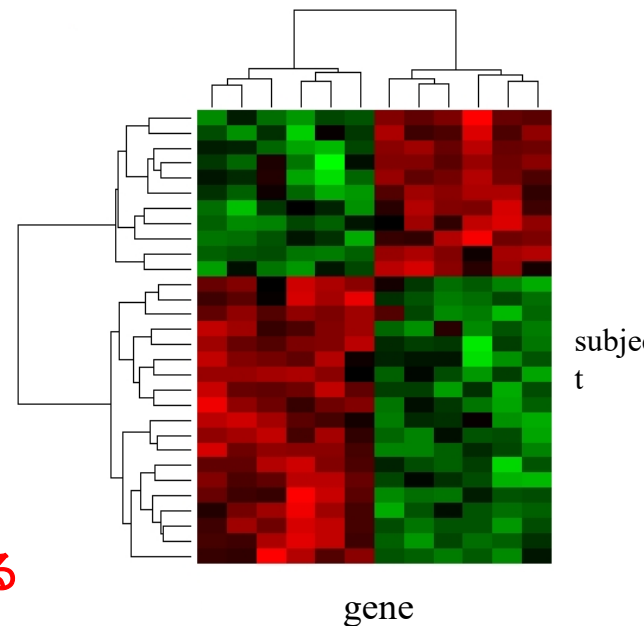
- GEO (Gene Expression Omnibus),

- 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/薬剤特異的变化

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



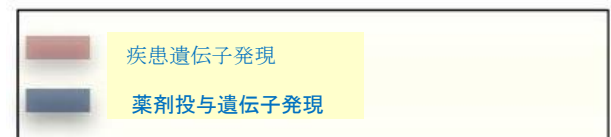
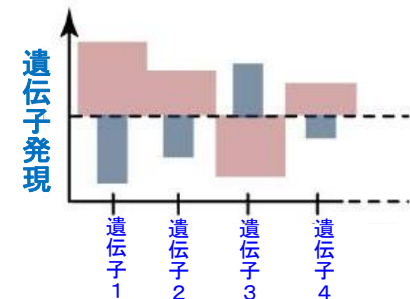
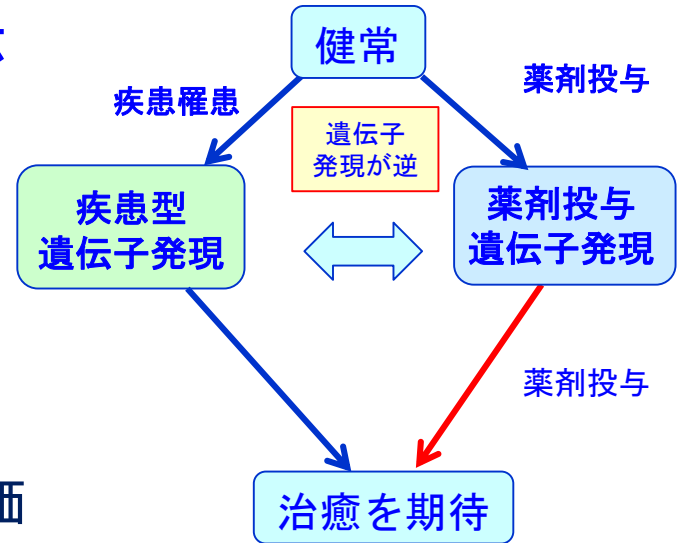
遺伝子発現プロファイルによる有効性予測

● 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**

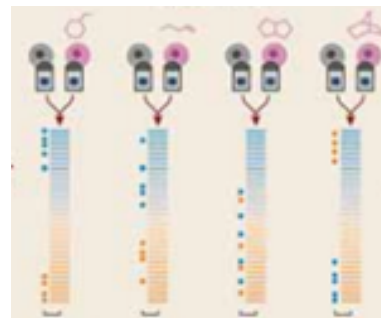
● 効果が相加的なら**有効性**が期待される

- 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

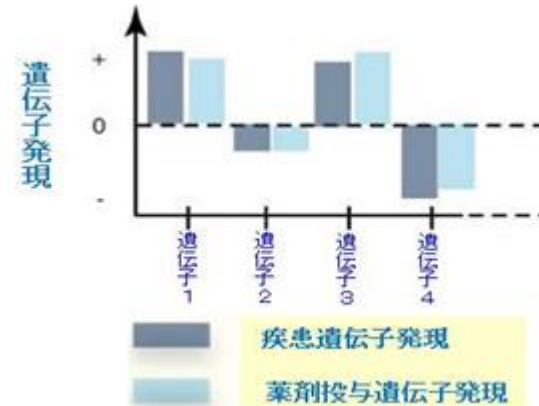
青は発現が**上昇**した遺伝子
赤は発現が**下降**した遺伝子



強正 弱正 弱負 強負

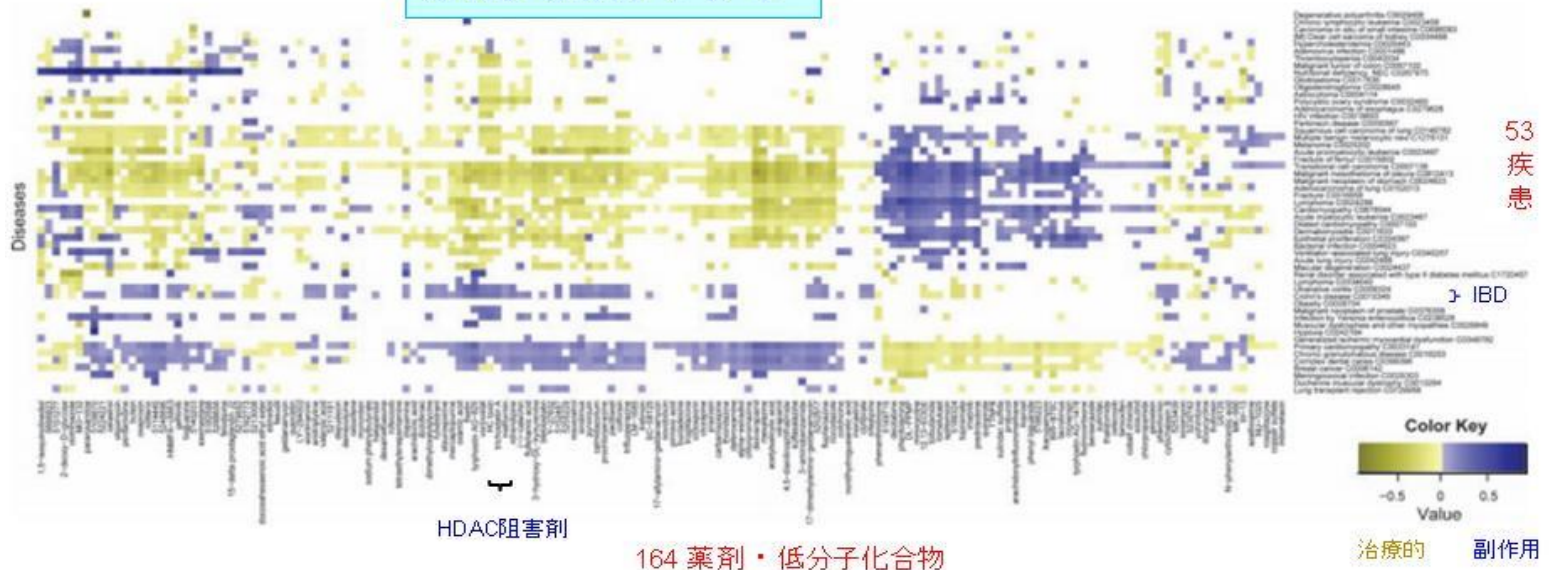
遺伝子発現プロファイルによる毒性予測

- 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

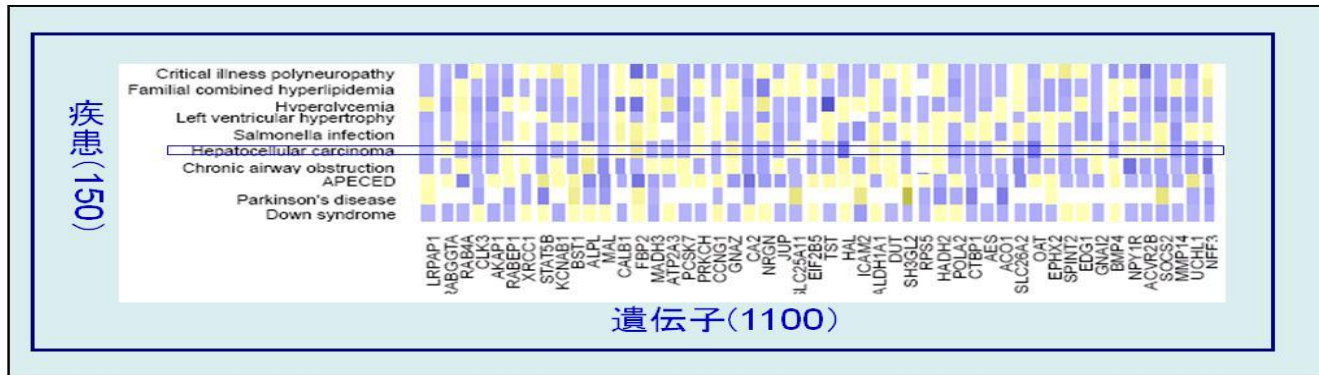


疾患-薬剤マップ

(Sirota, Butte 2011)



疾患ネットワークとDR

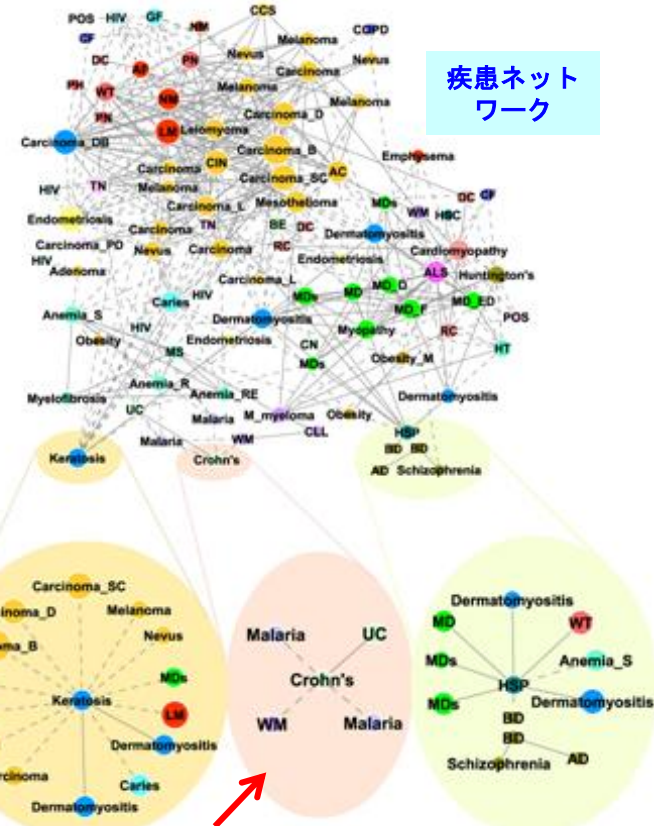


疾患を平均遺伝子発現パターンより疾患樹形図

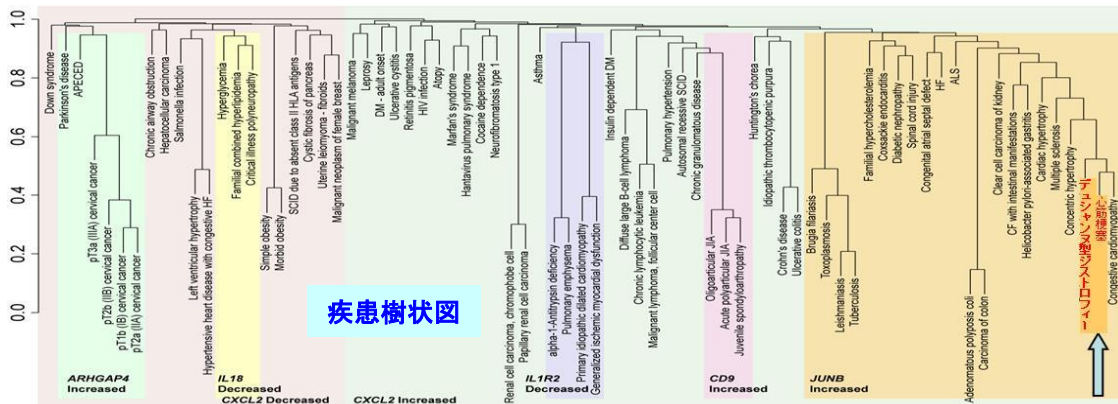
- 臓器別疾患分類では予想できない疾患間の親近性
- 分類項目はサイトカインの遺伝子発現と相関

疾患遺伝子発現パターンの近親性より疾患ネットワーク

- 内在的機序による「疾患ネットワーク」
- 従来にない親近性 例：クローン病とマラリア



疾患ネットワーク



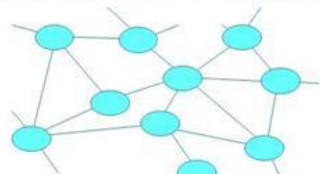
疾患樹状図

創薬方法論と人工知能（AI）

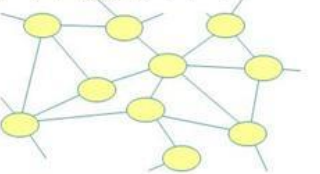
疾患ネットワークとDrug Projection Map

薬剤/疾患ネットワークの構築

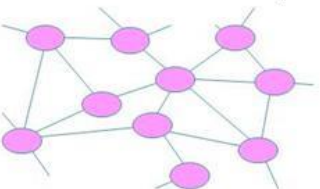
第1世代疾患ネット (原因遺伝子親近性)



第2世代疾患ネット (OmicsProfile親近性)



第3世代疾患ネット (Pathway親近性)



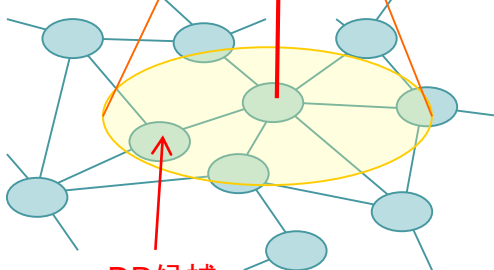
統合化



薬剤



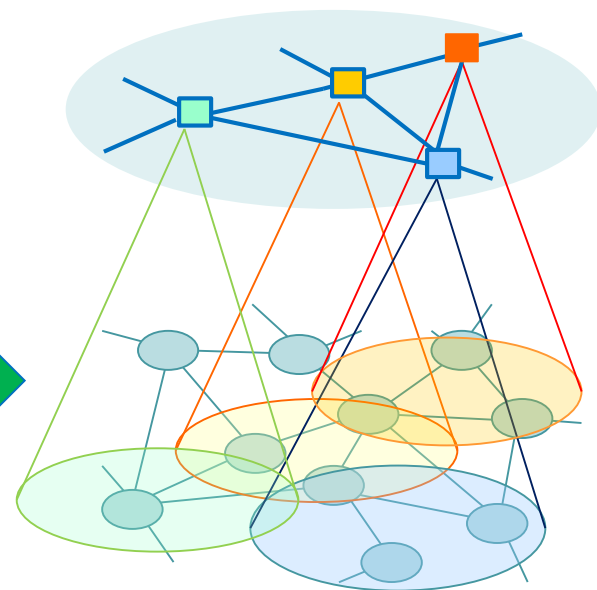
既存薬
関係



DR候補

疾患ネットワーク

薬剤ネットワーク

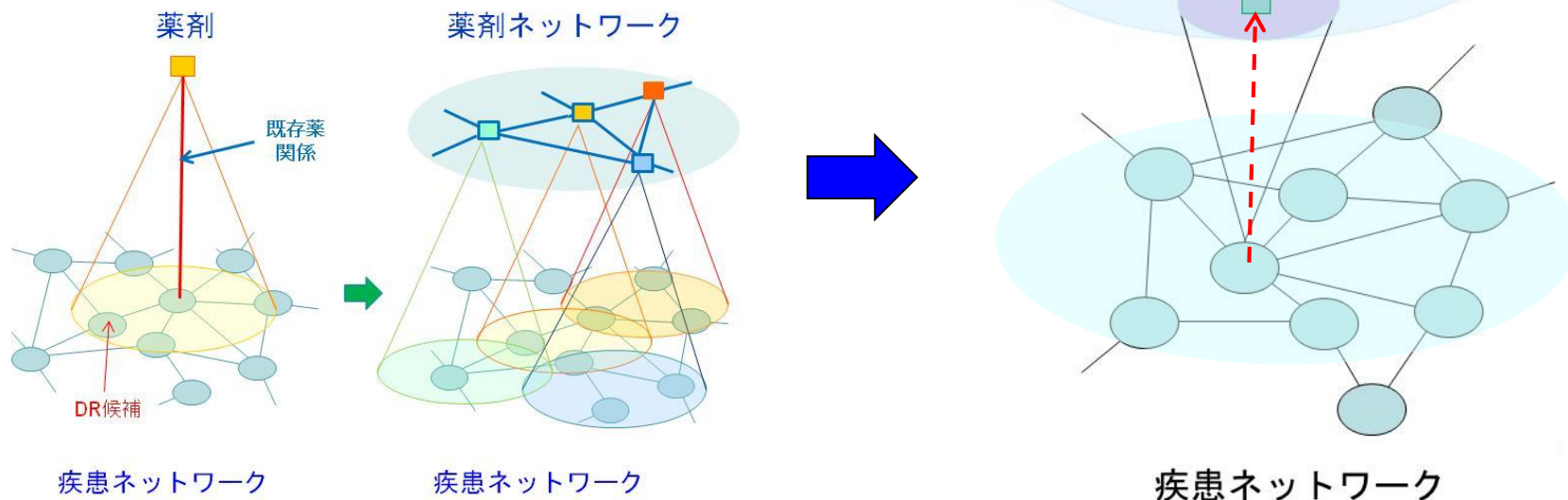


疾患ネットワーク

DRの方法論から創薬方法論へ

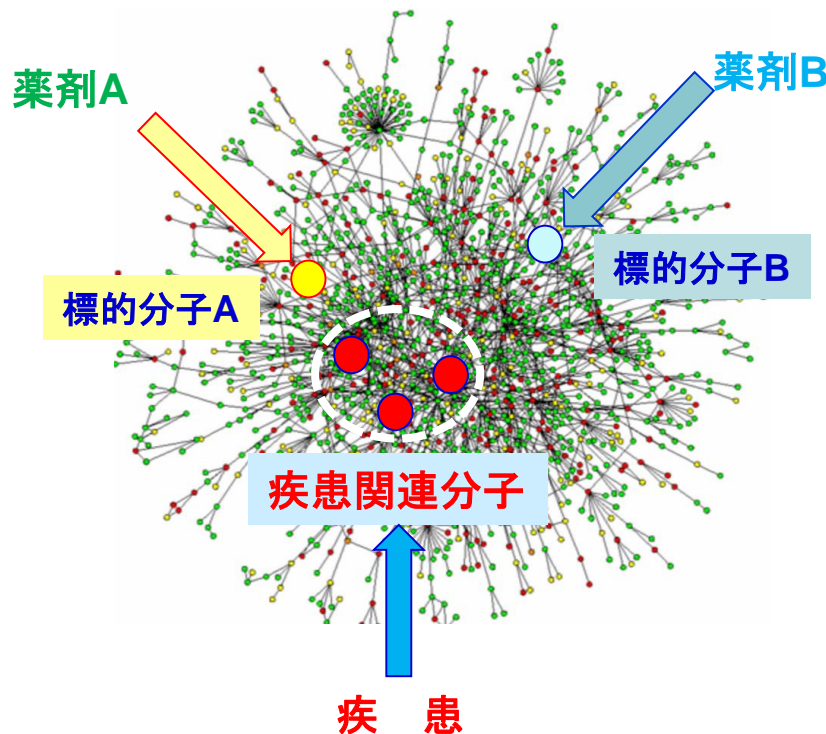
- 疾患ネットワークの十全な形成
 - 内在的機序の近親性から疾患ネットワーク
 - 医薬品の有効性・毒性の近傍 Projection
⇒ DRにおける**有効性**はすでに確立
- 創薬への展開
 - 薬剤階層のネットワークは既に確立
 - 投与時生体反応の近親性だけではなく
 - 化合物の構造的近親性(finnger print) からも作成
 - 疾患から逆投影。創薬の可能性探索
- 疾患ネットワークと薬剤ネットワーク間写像
 - **双方向性・対等性**

疾患から薬剤ネットワークへの**逆投影**
Multi-Topology 双対写像 創薬方法論



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)

3階層生命ネットワークでの創薬/DR

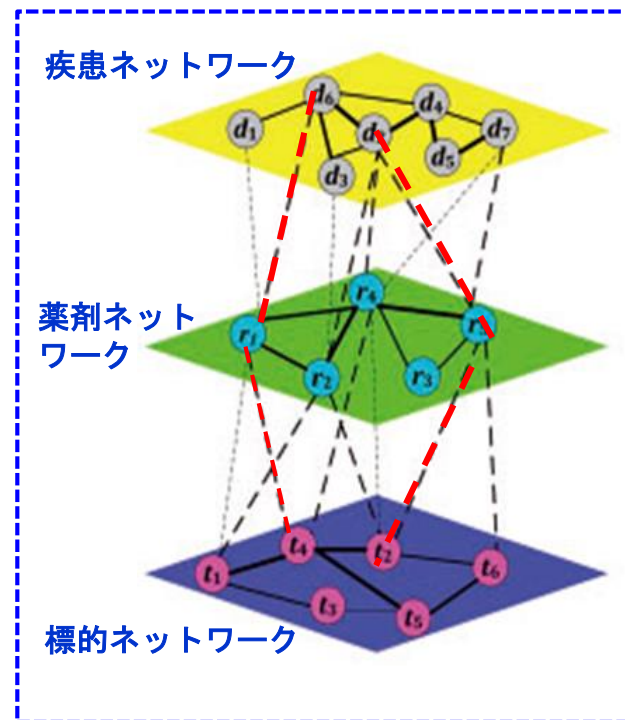
- 3階層の生体ネットワーク
 - 疾患ネットワーク：網羅的分子による内在的機序
 - 薬剤ネットワーク：化学構造によってネットワーク
 - 標的ネットワーク：薬剤と標的（DrugBank参照）
- 各層のネットワーク内結合
 - 稠密に自己完結的に構築可能
- 各層ネットワーク間のリンク
 - 成功した<疾患-薬剤>の事実の根拠のみ
 - 階層間はスパースな結合である

創薬/DRとは

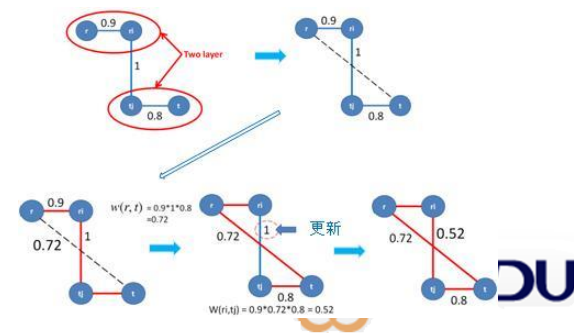
未発見の階層間リンクを
既存の階層間リンクの事実と
各層のネットワークから推測

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



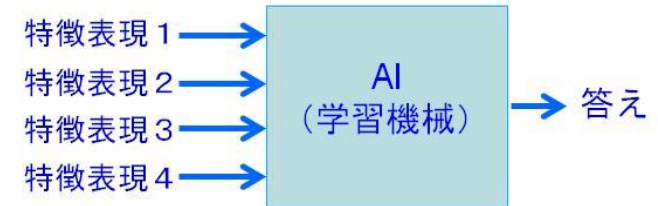
(Wang et al. 2014)



Deep Learning による 人工知能革命

- **機械学習のこれまでの限界**

- 分類・判別する学習機械
- 「教師あり学習」
 - 対象の特徴表現と正解を与えて学習
 - 与え方に関して細かな技法にとられる
- 人間の知性を超えることができない



- **Deep Learning の内在的表現学習**

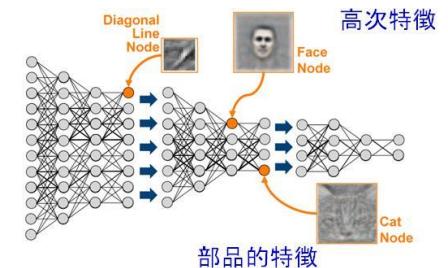
- 対象の固有の構造を記述する**特徴表現**や対象の**高次特徴量**を自ら学ぶ「**教師なし学習**」を行う
- 最終層で、人間の概念との連関をとるため「教師あり学習」

- **各層間で「自己符号化」の積上げ (autoencoder stack)**

- 各層ごとに前段階の層を、次元を圧縮された次階層から可及的に復元するよう重みを決定する
- 可及的に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

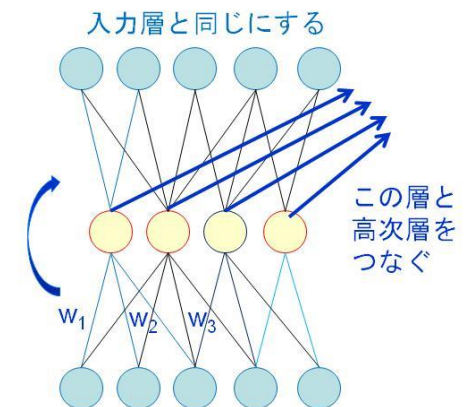
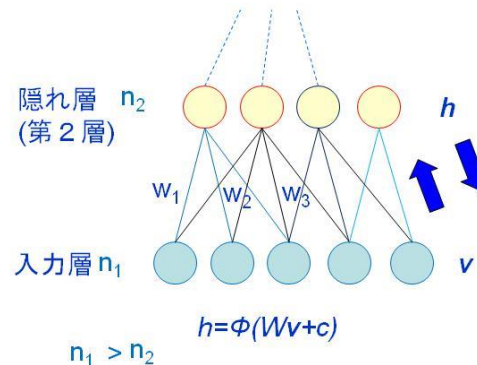
- 人間の「思考の枠組み」を超えた正解の低次

- 「アルファGo」が定石にない手で碁の名人に勝つ



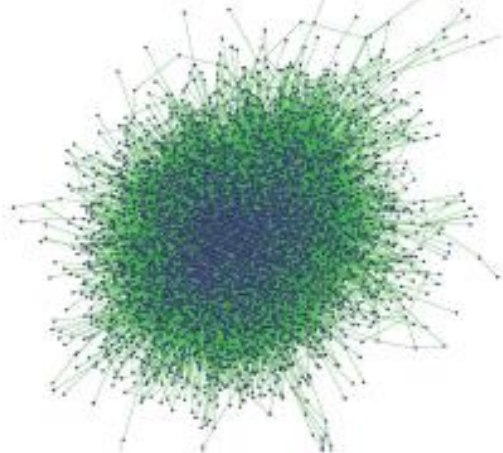
仮説駆動 (hypothesis-driven)から

データ駆動 (data-driven)な
「教師なし」学習へ

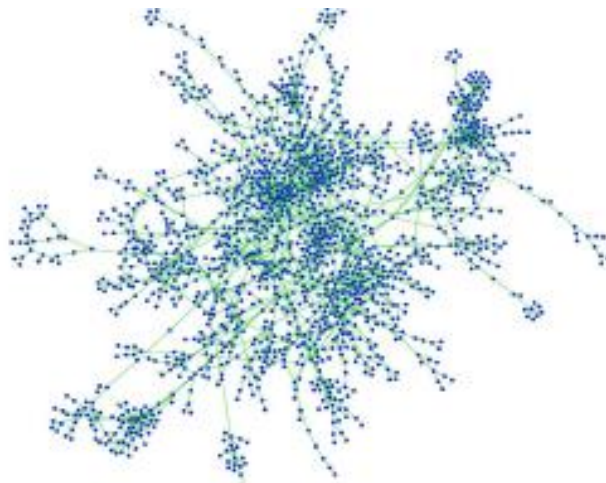


3階層ネットワークの有効性例

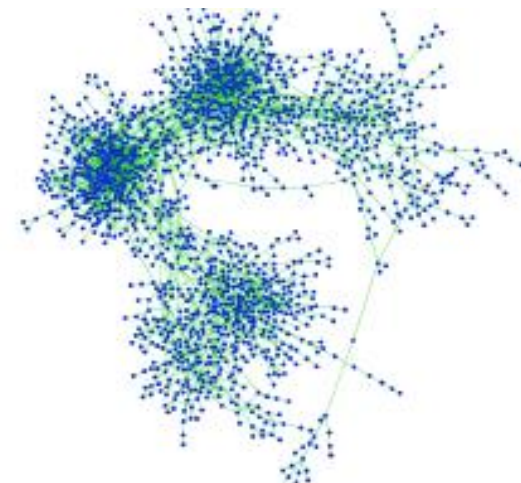
使用したデータベース



タンパク質相互作用ネットワーク



合成用量致死ネットワーク



合成致死ネットワーク

3種のネットワークと、薬剤と標的分子の情報を、訓練データ・試験データの構築に用いた。



Step 1 : ネットワーク解析を行い、機械学習モデルを生成するための**特徴量**を抽出する。

Step 2 : 機械学習を用いて**対象疾患（疾患A）**に対する**新規標的分子**を予測する

Step 3 : 予測された**新規標的分子**をベースにして、この疾患Aに対する**新しいDRの薬剤**を推測する

Step 1 : Network解析による各遺伝子の特徴量の抽出

そのノードのネットワーク特徴量を表す21種のネットワーク統計量を、各遺伝子に対して、各ネットワークを解析して算出した合計63種のネットワーク統計量を各遺伝子の特徴量として予測モデルの構築に用いた。

1. **次数中心性**（他のタンパク質との結合次数が多い）

2. **近接中心性**

他の全ノードとの距離の平均、ネットワークの中心に位置する程小さい、すなわち他の位置ノードと全般的に「近い」

3. **媒介中心性**

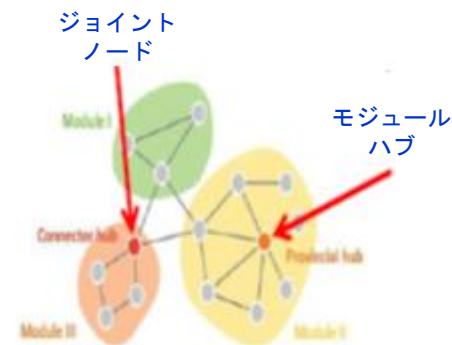
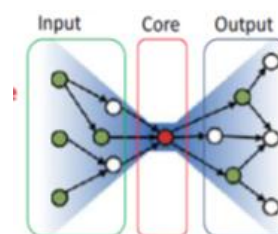
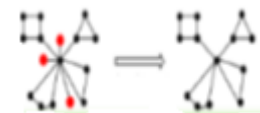
ネットワーク内の各ノード間を繋ぐ様々なパスが対象となるタンパク質をどれだけの頻度で通過するか、通過する頻度の割合が高い程、中心性がある

4. **蝶ネクタイ構造指標**：

ネットワークに様々なノードから入力信号が入ってくると、中心的なコアで一旦集約されて様々な遺伝子へと拡散する構造（bow-tie）においてコア部にいるか、両末端部にいるかを表す指標である。

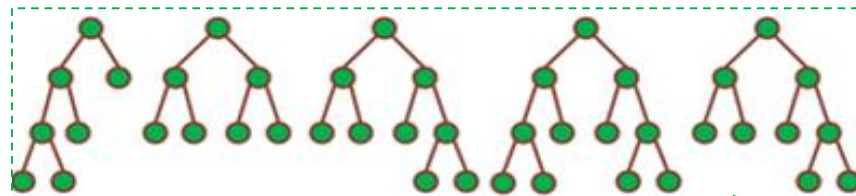
5. **モジュール指標**

ネットワークが幾つかのモジュールに分かれる（積み木方式）ときモジュール間を繋ぐノードか、各モジュールのハブかなどの指標である。



Step 2 機械学習モデルによる決定木 Random Forest

63 個の特徴量からランダムに
いくつかの特徴量を選んで判別
ルールの決定木を**1000**個作成



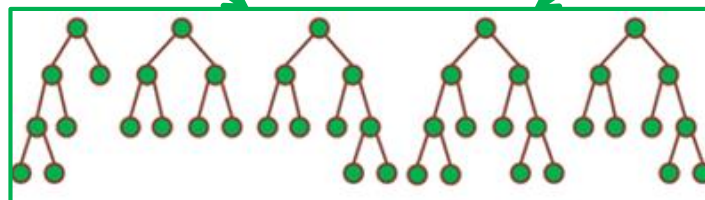
遺伝子A

ネットワーク特徴量1=0.5
ネットワーク特徴量2=1.3
...

遺伝子B

ネットワーク特徴量1=2.3
ネットワーク特徴量2=4.0
...

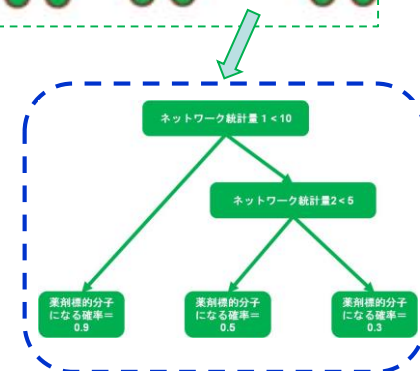
63種
の特徴量



1000個
決定木

薬剤標的でない
遺伝子A

薬剤標的である
遺伝子B



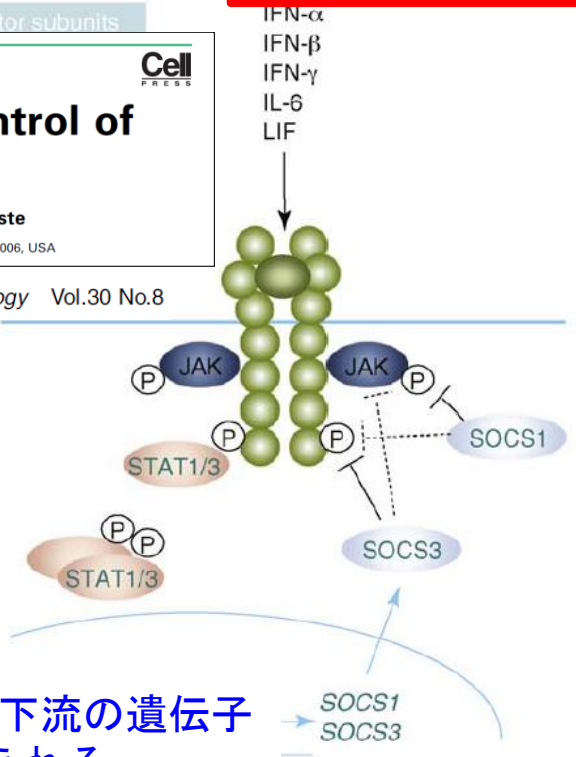
<疾患- 標的分子> 予測結果

Target disease	Top 25 genes for potential novel drug target for a target disease
Anti-Alzheimer's disease	SOCS1 ; <i>EPHA2; PRR38; 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; NCF1; C6orf47; NR4A1</i>
Anti-Breast Cancer	<i>SHC1; NFKB1; RELA; ID2; RAC1; SRC; JUNX1; HDAC2; I</i>
Anti-Colorectal	<i>SHC1; GHR; HRAS; PLCG1; NFKB1; LYN</i>
Anti-Pancreatic	<i>IL2; DSCC1; TAF9; AIF1; IL2; 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。

Step 3 : 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

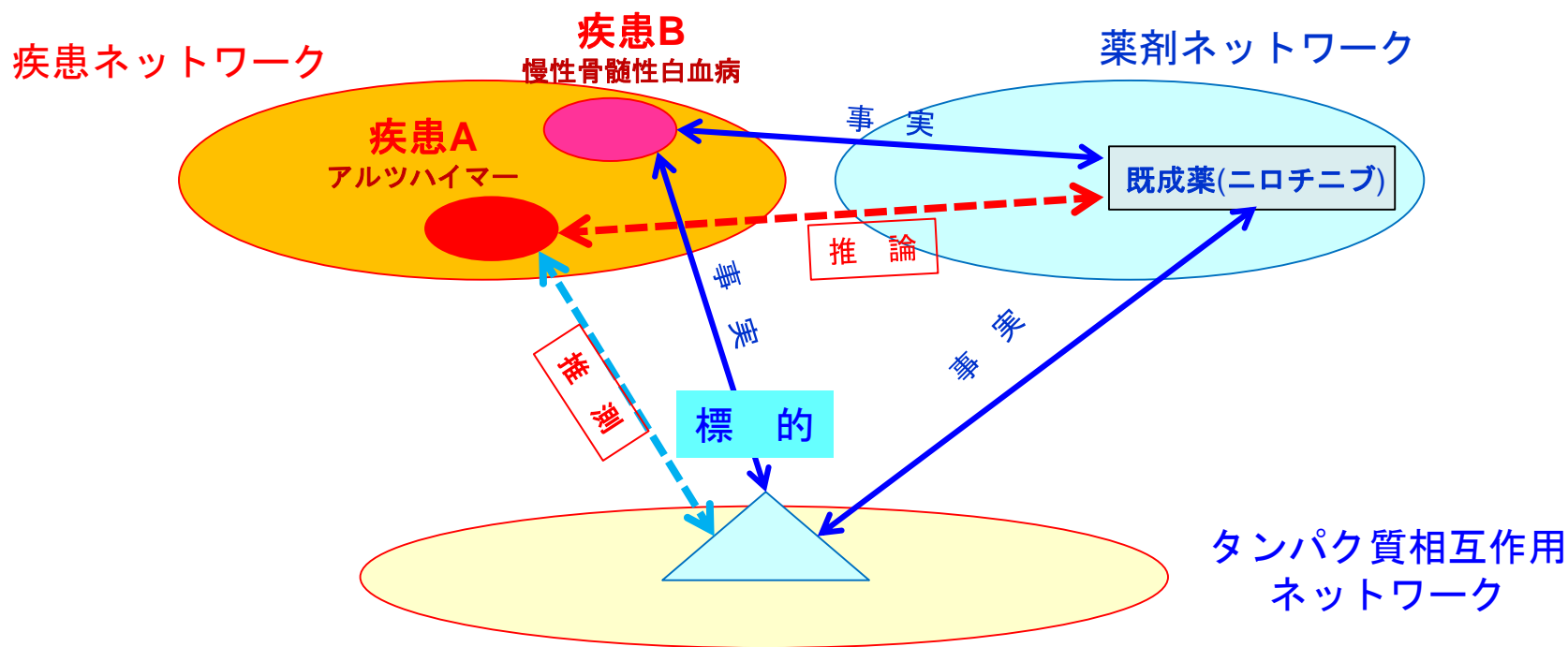


Table 3. Predicted repositionable drug candidates

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; Testosterone; Tetraethylammonium; Theophylline; Thiamine;

Neuroscience 304 (2015) 316–327

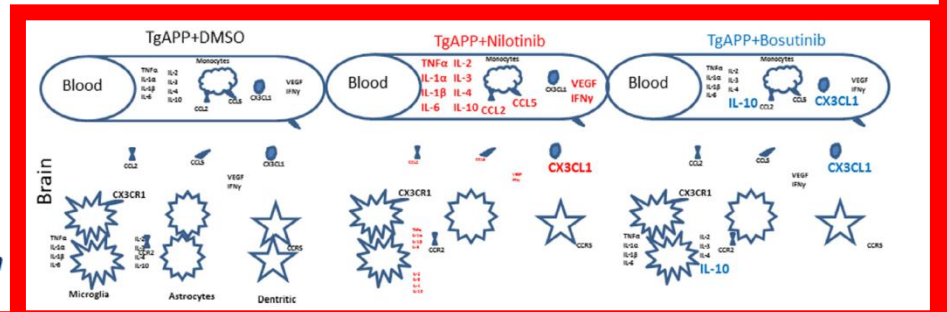
慢性白血病の抗がん剤であるニロチニブがアルツハイマーのDR薬剤として選定

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*}

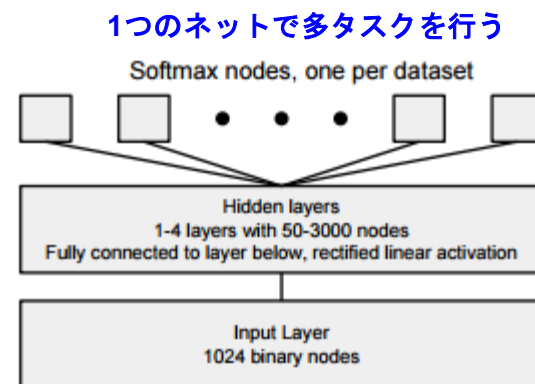
^a Department of Neurology, Laboratory for Dementia and Parkinsonism, Georgetown University Medical Center, Washington D.C. 20007, USA

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



Deep learning: 創薬からの注目

- 創薬を巡る状況
 - 平均14年、約 1 000億円を超える費用
 - 市場化された新薬の減少
 - 創薬に費やす期間・コストを低減したい
- **Kaggle** (データサイエンス競技会)に**Merck社**が出題
Molecular Activity Challenge (2012).
 - 15データセットから異なった**構造活性相関のデータ**を学習して構造から分子の生物学的活性を予測するモデルの開発コンテスト
 - 勝利したモデルはdeep learning を用いたモデル
- Google in collaboration with Stanford (2015)
 - Stanford 大学の Pande 研究室と共同研究
バーチャルドラッグスクリーニングに対する deep learningによるツール開発
"Massively Multitask Networks for Drug Discovery"



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
- 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。

AI創薬/DRの方向性

- 我々の研究の「ネットワーク解析と機械学習に基づいたDR」
 - 機械学習の時点で、DLも適応したが特徴量抽出後であったためか、精度はRandom Forrestの方が高かった
 - ネットワーク特徴量を与えるのではなく、**ネットワークの結合情報の全ての情報**を与えStacked Autoencoderで特徴量を学習させる
- AI創薬の基本方針
 - 創薬・DRが根拠とする3階層ネットワーク枠組みのもとで、Deep Learningを中心とする機械学習・知識発見の戦略が期待される

今後の戦略・方向

1. AI創薬の「枠組み」・実行方向は「見えてきた」
2. 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
3. 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「AI創薬コンソーシアム」を設立する必要がある

ご清聴有難う御座いました

