ビッグデータと人工知能に基づく創薬
～とくにリード化合物探索自動化の現状と展望～

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

医療・創薬分野への
ビッグデータ時代の到来

(1) 次世代シーケンサ（Clinical Sequencing)による
「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)

急激な大量データの出現
コストレス化かつ高精度化 ⟵ !!

ゲノム：13年→1日(1/5000) 3500億→10万円(1/350万)

個別化医療・医療の国民レベルの向上
医療・生命情報の適確性の飛躍的な増大

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

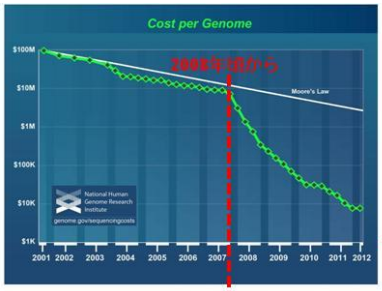
欧州でのゲノム医療

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



千ドル
ゲノム

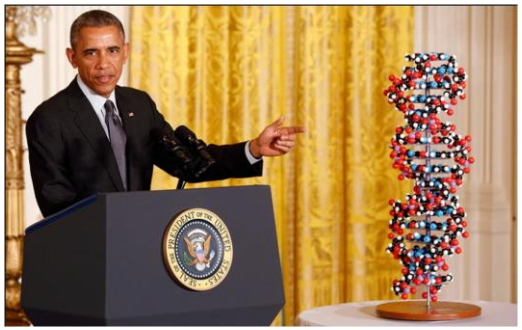
DNA Sequencing Cost: the National Human Genome Research Institute

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



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

国家政策の時代

精密医療普及期

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

2007年
2009年
2010年
2011年
2012年

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

2013年

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

2014年

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

2015年

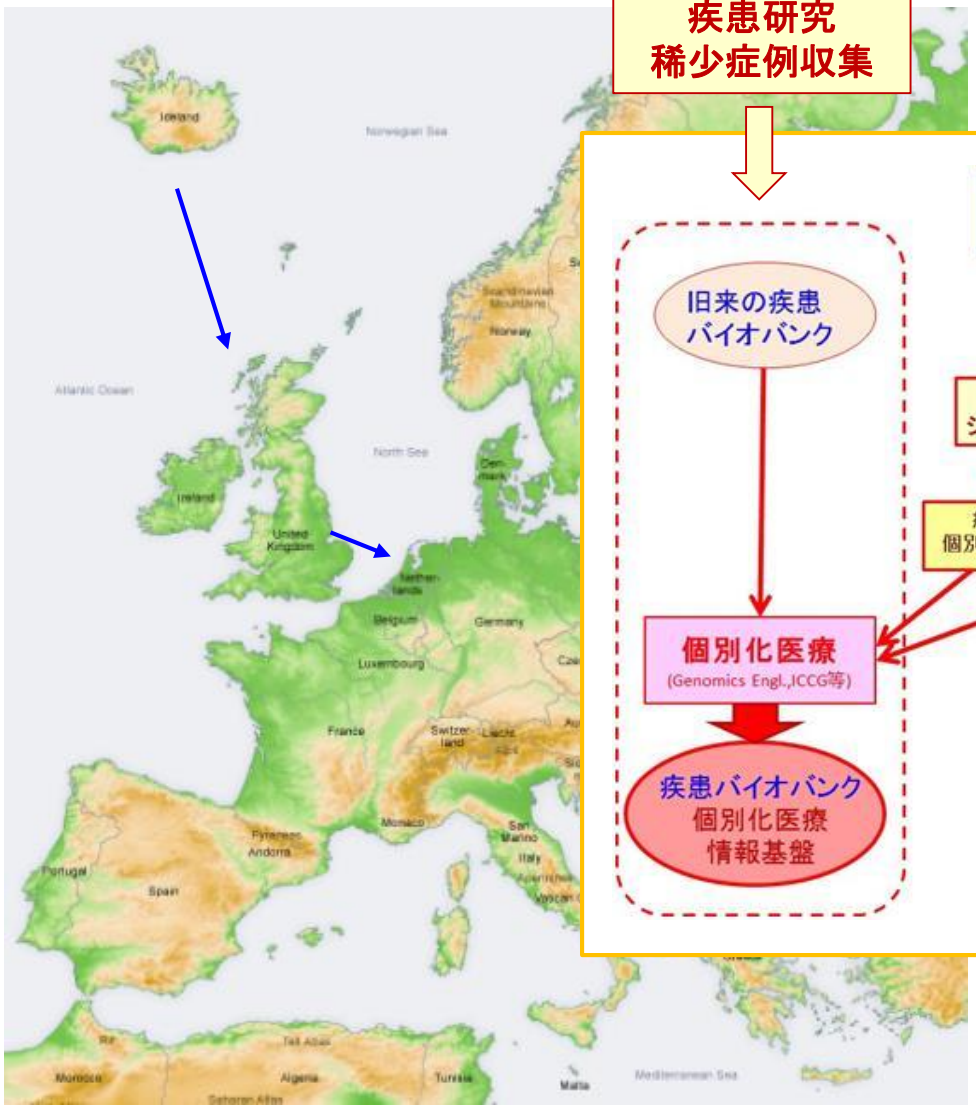
第三期

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)

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

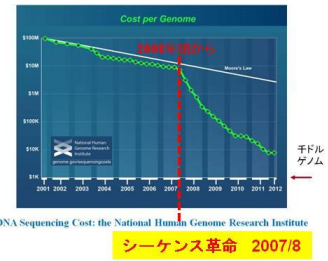
- ヒトゲノム解読計画以降急速に進展
  - Hapmapプロジェクト,1000 genome,がんICGC,TCGA、TopMED
  - ゲノム変異・多様体
    - dbSNP, HGMD, **Clinvar**, **Clingen**, OMIM,GWAS catalog
    - 表現型との対応: dbGaP, EGA
  - 遺伝子発現プロファイル
    - 疾患特異的transcriptome: **GEO, ArrayExpress**,
    - 薬剤特異的trancriptome: **c-Map, LINCS**
  - タンパク質
    - ３次元構造：PDB, 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

医療の「新しいビッグデータの革命性」

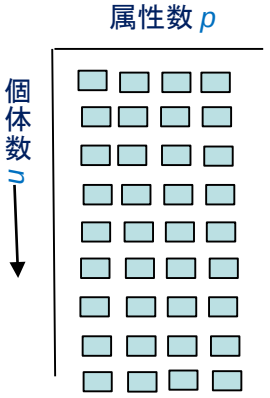
～ゲノム・オミックスデータの基軸的な特徴～

＜**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)

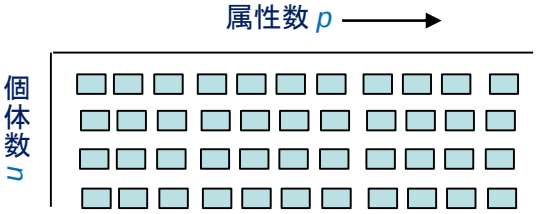
医療情報・疫学調査では **属性数**：**数十項目程度**
個体数：近年電子化の流れ⇒**個体数：膨大**

– 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

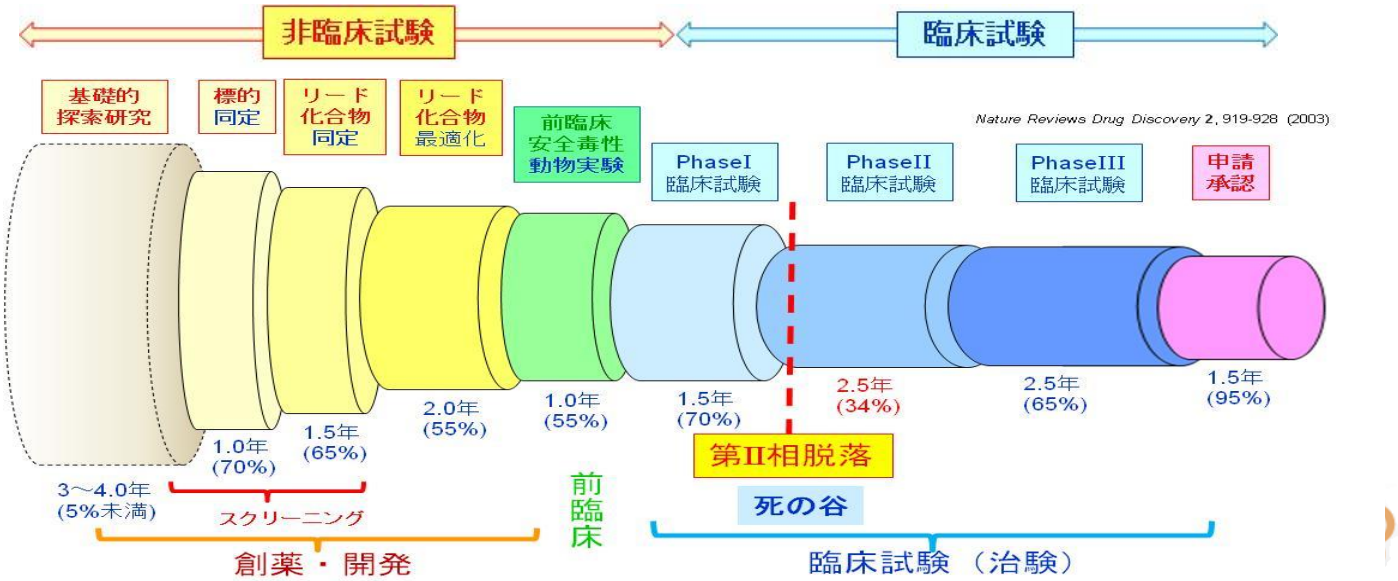
生命情報ビッグデータを利用した創薬戦略の進展へ

創薬をめぐる状況と解決の方向

- 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**

Nature Reviews Drug Discovery (2012)

ヒトの＜薬剤－疾患－生体系＞のビッグデータを早期R&D段階で使う



ドラッグ・リポジショニング（DR）

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

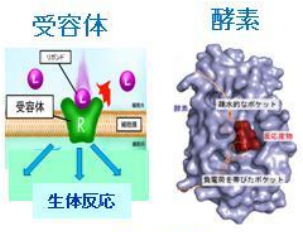
疾患・薬剤・標的の関係

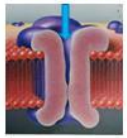
病気の主要な要因

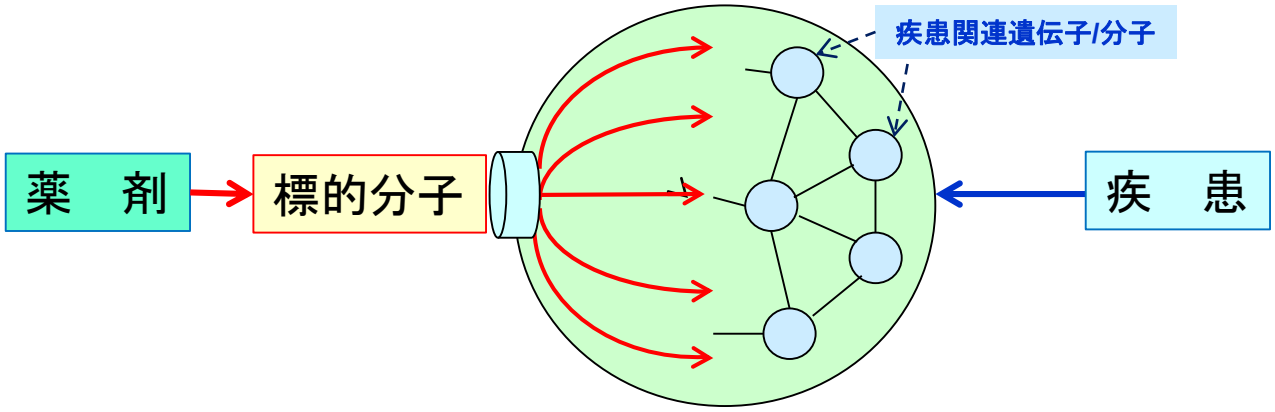
疾患関連遺伝子/分子(複数)

薬：薬剤**標的分子**を通して作用する
疾患関連遺伝子/分子に影響を持つ

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

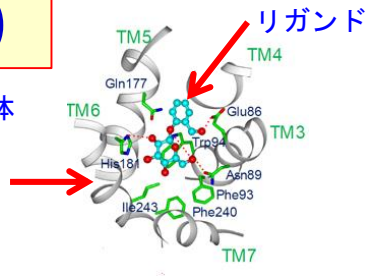
TMDU

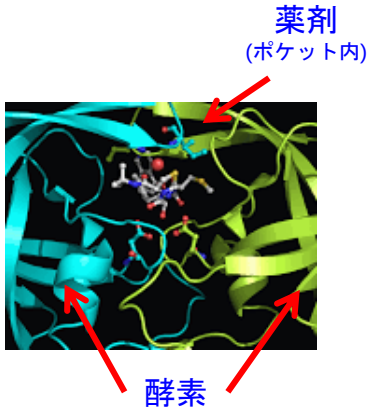
計算創薬・DRの基本枠組みと「生体分子プロファイル」型創薬・DR

生体プロファイル型計算創薬・DR

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)

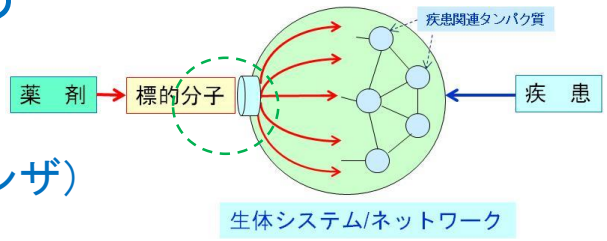




分子(結合構造)中心

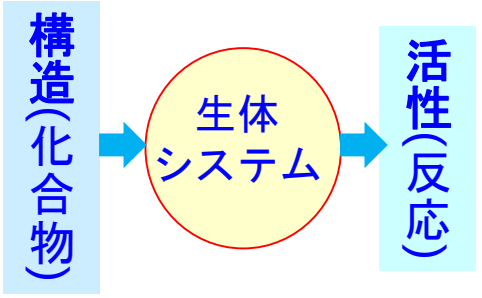
- **分子構造解析・分子設計**
- **Structure-based rational drug design**
- **標的分子**(受容体・酵素)**と薬剤**(リガンド)との**結合構造**（ポケット）の分子構造を根拠に
- **リガンドの分子設計(量子化学等を使う)**
  - 成功例：インフルエンザ薬　ザナミビル（リレンザ）
- 標的に結合する**リード化合物・構造最適化**
- **結合後の生体システム**の反応・振舞い

  ⟹ 明確な取扱いがない



定量的構造活性相関(QSAR)

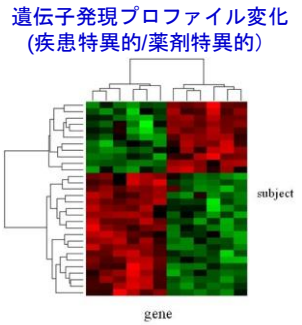
- 化合物の分子**構造**と生体**活性**の**関係**
- **しかし両者の間**には**生体システム**がある



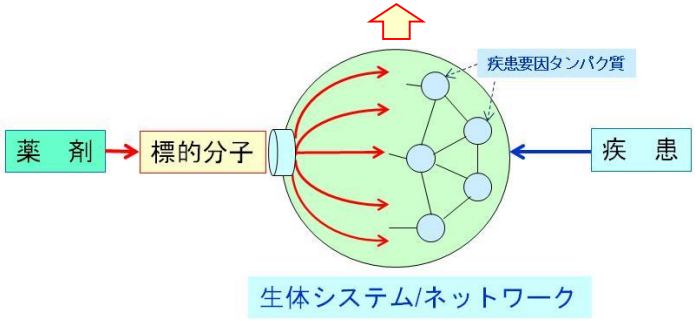
生体分子プロファイル型計算創薬

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



網羅的分子プロファイル⇒分子ネットワーク全体変化

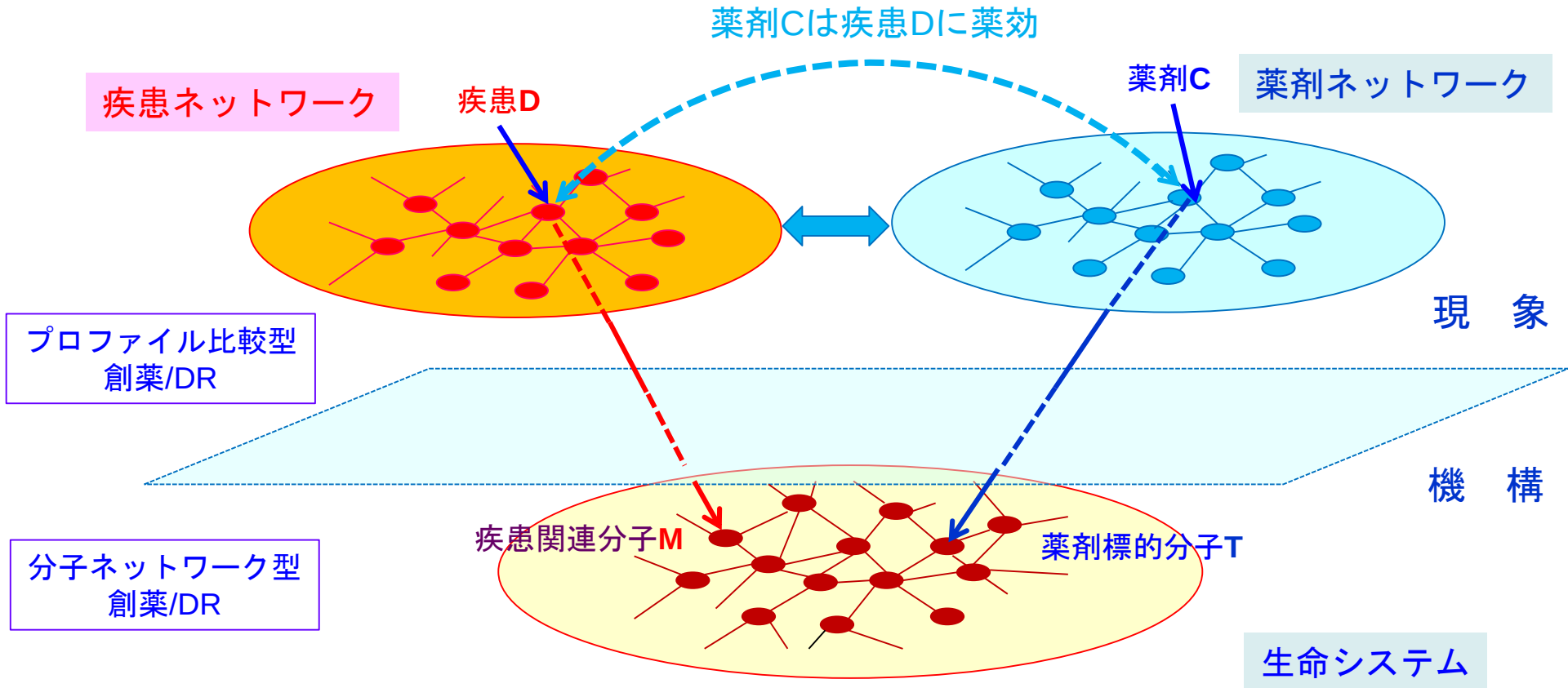
＜疾患状態の生体＞に**＜薬剤ー標的分子の結合＞**を引き起す作用によって
ゲノムワイドな 生体分子環境がどう変化するか「生命システム観点からの理解」

化合物, 標的分子, 疾患間の関係の「ビッグデータ」DBを利用

TMDU

創薬・DRの基本的枠組み

疾患・薬剤・生体系をネットワーク間の関係として認識



生体分子プロファイル型の計算創薬・DRの体系

計算創薬・DRの体系

計算創薬・DRの「非学習的」アプローチ

「ビッグデータ創薬・DR」

- **疾患ー薬剤プロファイル直接比較**
  - 「現象論的」アプローチ
  - 疾患罹患時と薬剤投与時の遺伝子発現プロファイル比較
  - 疾患ネットワークと薬剤（化合物）ネットワークの比較
- **生体分子ネットワーク準拠近接解析**
  - 「機構論的」アプローチ
  - 疾患ネットワーク上の比較
  - タンパク質相互作用ネットワーク疾患遺伝子と標的分子

計算創薬・DRの「学習的」アプローチ

「AI創薬」

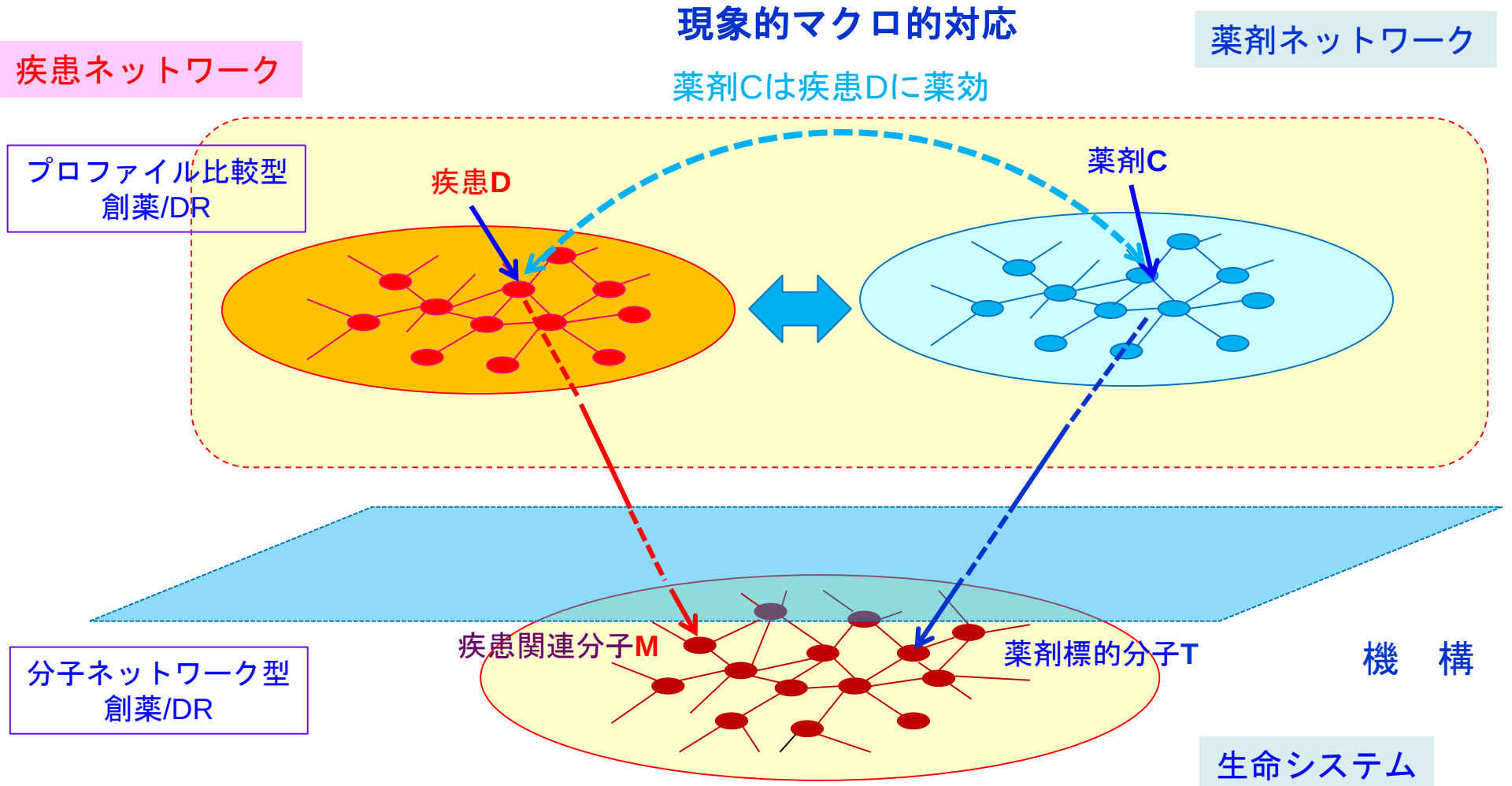
- **Virtual Screeningへの人工知能・機械学習の応用**
  - Ligand-based　AIーバーチャルス・クリーニング
  - Structure-based　AIーバーチャル・スクリーニング
- **標的分子探索に人工知能を用いた方法**
  - HaseーTanakaの多層Deep AutoEncoderを用いた標的分子探索法
- **その他**
  - 化合物自動設計、合成経路探索、毒性予測

TMDU

ビッグデータ創薬・DR
（非学習型アプローチ）

「疾患ー薬剤ネットワーク」相互関連型

3層の生体・薬剤のネットワーク間の関係図式



生体分子プロファイル型創薬/DR
非学習アプローチの深化

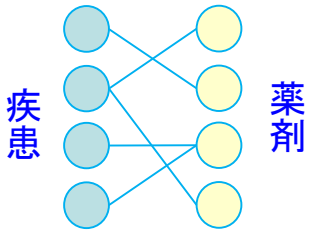
深化

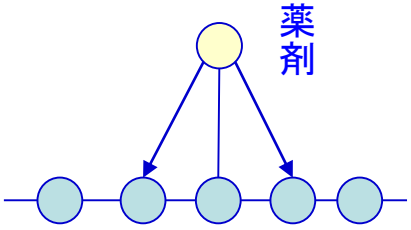
第１段階：疾患・薬剤プロファイル直接比較
- 疾患罹患時と薬剤投与時の生体反応の遺伝子発現プロファイルを比較。
- パターン正負相関性に基づく有効性毒性予測

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
- ネットワーク近接性に基き有効性・毒性予測

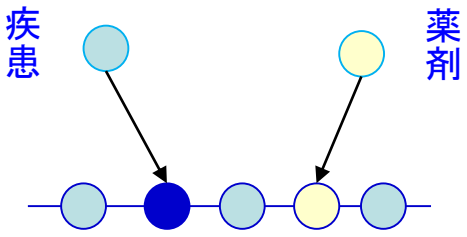
第３段階：生体ネットワーク媒介型比較
- 生体分子ネットワークを＜場＞として、疾患・薬剤の作用の足場分子を同定
- 足場分子間の相互作用（総合的距離）の評価に基づき有効性・毒性予測

生体分子プロファイル比較





疾患ネットワーク



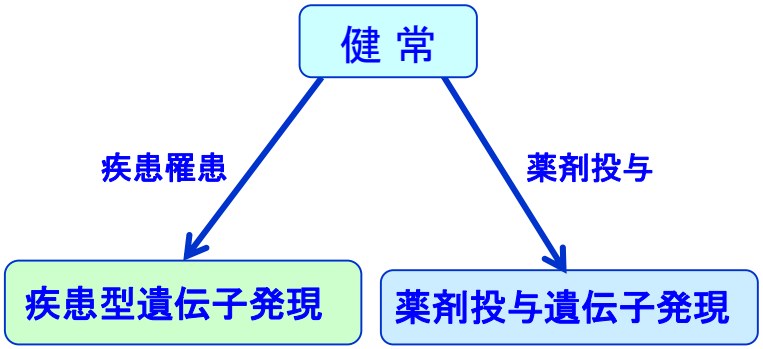
生体分子ネットワーク

TMDU

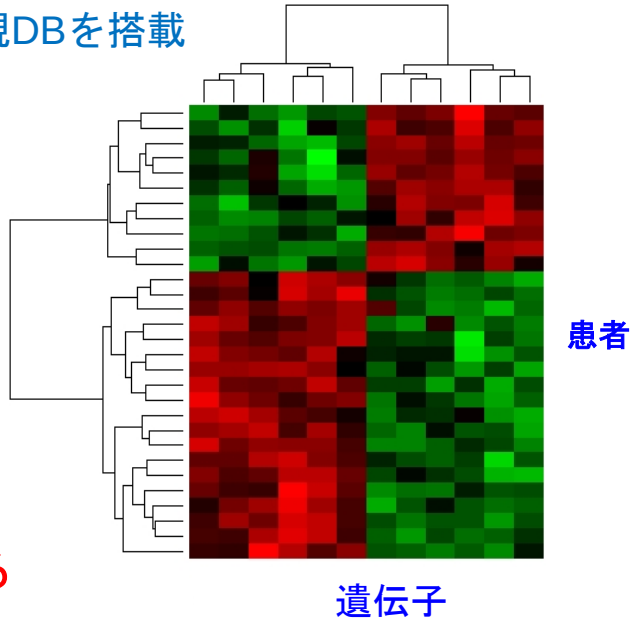
１．遺伝子発現プロファイル
直接比較型

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のプロファイルを提示
    - 最近はLINCSデータベース：100万種の薬剤特異的発現DBを搭載



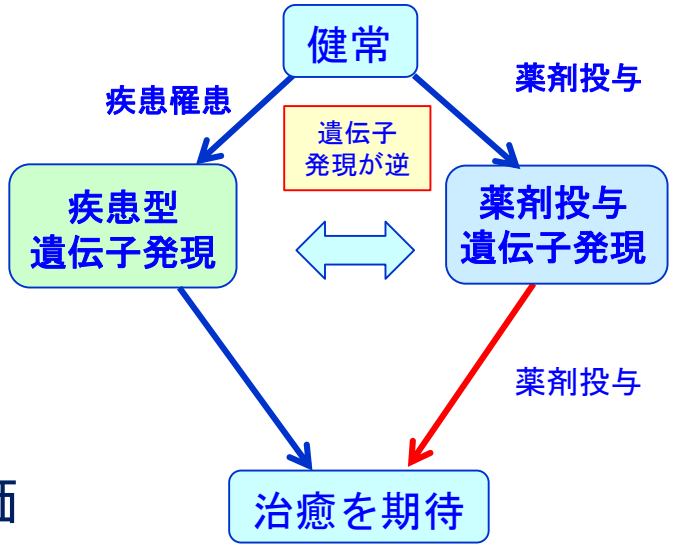
- 疾病特異的遺伝子発現
  - GEO（Gene Expression Omnibus),
    - 疾病罹患時の遺伝子発現プロファイルの変化
    - 米国NCBI作成・運用　2万5千実験,
      70万プロファイル（欧州　ArrayExpress）
    - EBIが作成、サンプル数同程度

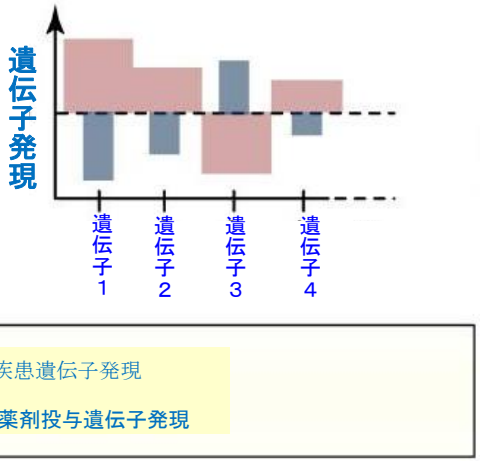


基礎には分子ネットワークの疾病/薬剤特異的変化
遺伝子発現プロファイル変化
≈　分子ネットワーク活性構造変化を反映する

遺伝子発現プロファイルによる有効性予測

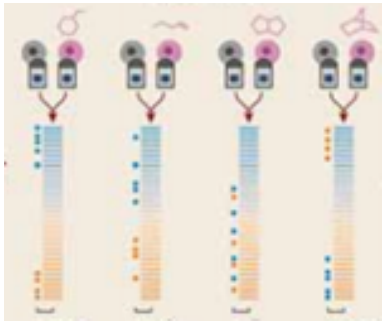
- 遺伝子発現シグネチャア逆位法
  - 疾患によって健常状態から変異
    「疾患特異的遺伝子発現プロファイル」
  - これに薬剤投与の変化を起こす
    「薬剤特異的遺伝子発現プロファイル」
  - 両者のパターンが負に相関する
  - ノンパラメトリックな相関尺度で評価
- 効果が相加的なら**有効性**が期待される
  - 例：炎症性腸疾患に 抗痙攣剤(topiramate),
    骨格筋委縮にウルソール酸





DB格納プロファイルは**薬剤**遺伝子発現
問合せ（横点）署名は **疾患**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

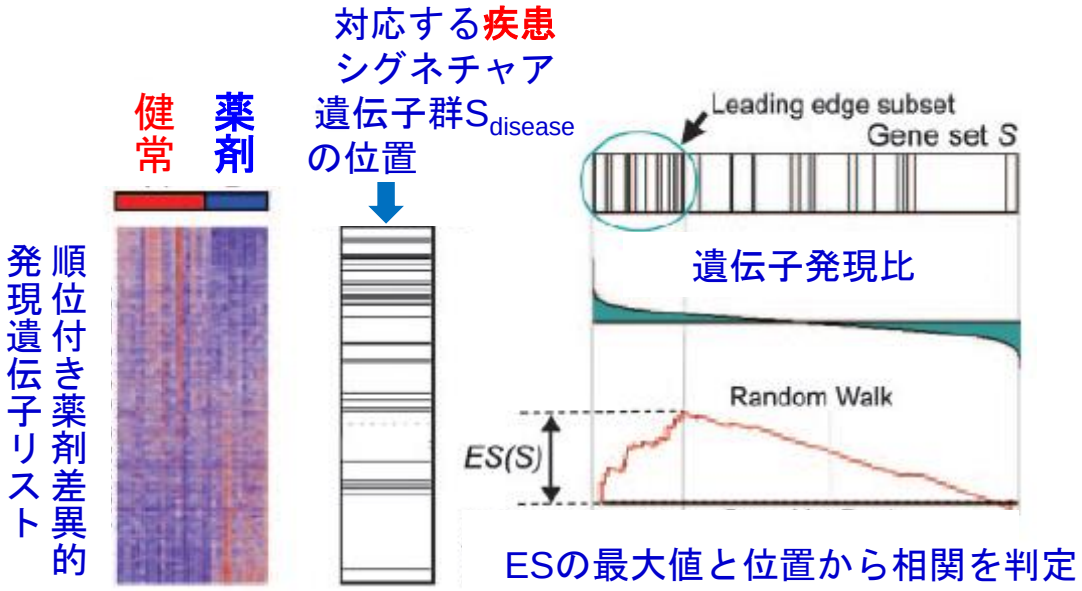


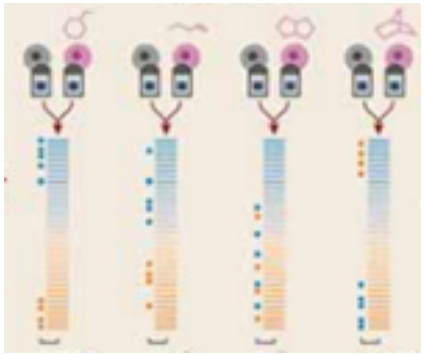
TMDU

Non-parametric な相関尺度で評価

Gene Set Enrichment Analysis（GSEA)

– 対照と比較して順位づけられた遺伝子リストの上位に密集しているかの尺度



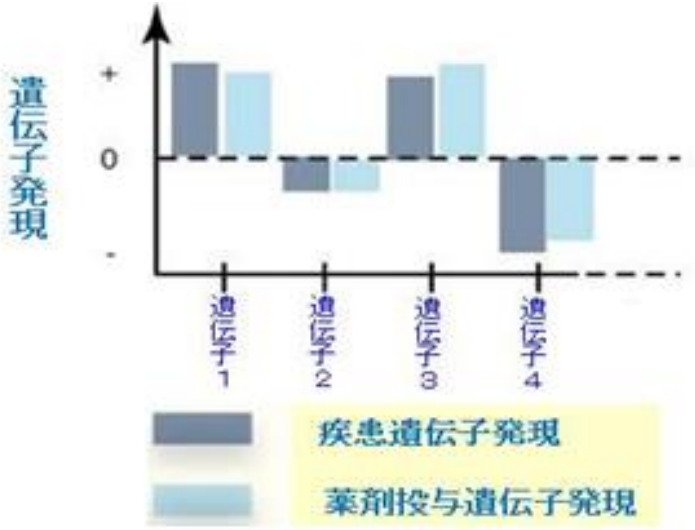


ESの最大値と位置から相関を判定

発現比ランクの高い順から遺伝子を調べ
遺伝子リスト$S_{disease}$中に**該当する遺伝子が存在したらES (Enrich Score)を加算**、無ければ一定値を減算

遺伝子発現プロファイルによる毒性予測

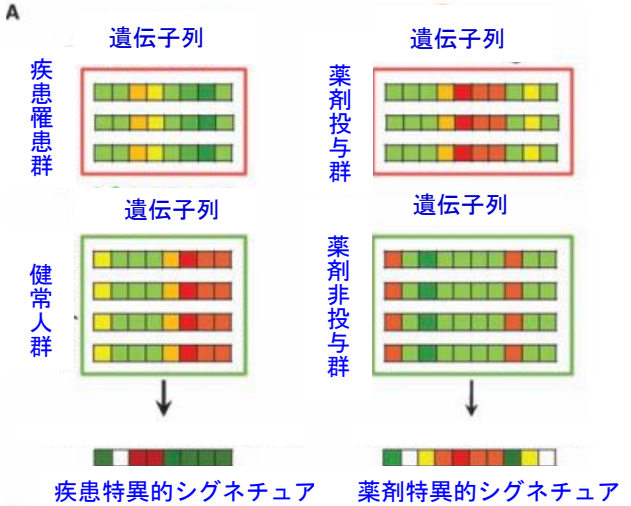
- 連座法 guilt-by-association ：
- **薬剤ー疾患間　副作用予測**
  - 薬剤特異的シグネチュアと
  - 疾患特異的シグネチャアが
  - ノンパラメトリック相関　正
  - 毒性・副作用の予測
- 薬剤ー薬剤間
  - 薬剤ネットワークからのDR
  - Connectivity map から薬剤特異的遺伝子発現の薬剤間の親近性をノンパラメトリック親近性尺度（GSEA)で評価
  - この類似性のもとに薬剤ネットワーク構築
    近隣解析によりDR
  - 例：抗マラリア剤をクローン病に適応



遺伝子発現プロファイルの正・負のパターン相関性に基づく計算DR

(Sirota, Butte 2011)

- NCBI・**GEO**から**100疾患**のシグネチャアを取得
- **c-Map**より得た**164の薬剤・化合物**の
  薬剤特異的遺伝子発現プロファイル
  疾患-薬剤間で類似性スコアを計算
- 約16000組の疾患-薬剤間の2664組が
  有意、半数以上が治療的関連(負)あり
- 100疾患内, 53疾患が有意に164薬剤と関連



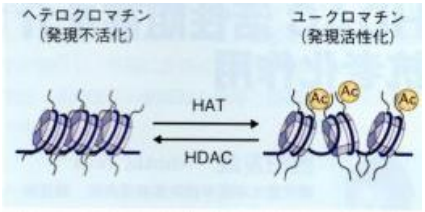
適応の多い薬剤と疾患

Drugs with most indications		Diseases with most indications	
Vorinostat **HDAC阻害剤**	21	Transitional cell carcinoma	95
Gefitinib	18	Melanoma	79
HC toxin	18	Cardiomyopathy	73
Colforsin	17	Adenocarcinoma of lung	72
17-Dimethylamino-geldanamycin	16	Multiple benign melanocytic nevi	68
Trichostatin A	16	Squamous cell carcinoma of lung	67
3-Hydroxy-DL-kynurenine	15	Malignant neoplasm of stomach	66
5114445	15	Dermatomyositis	63
Dexverapamil	15	Malignant mesothelioma of pleura	53
Prochlorperazine	15	Primary cardiomyopathy	48

(Sirota, Butte 2011)

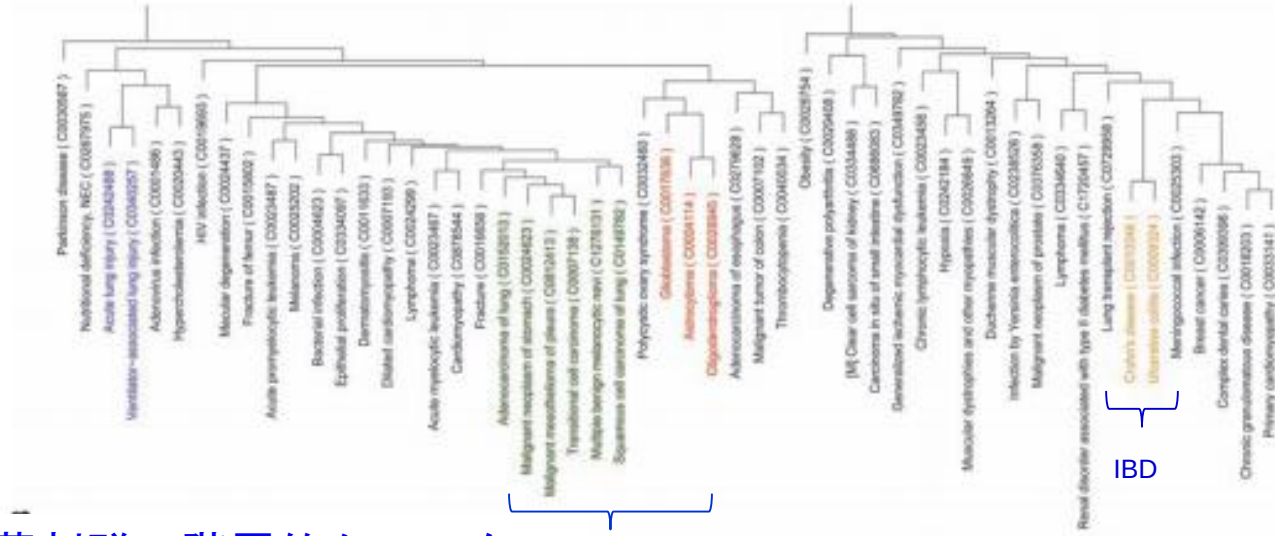
Drug group	Drugs
PI3K inhibitors	LY-294002 and wortmannin
HSP90 inhibitors	Geldanamycin, raloxifene, monorden, and sodium phenylbutyrate
HDAC inhibitors	Vorinostat, HC toxin, and trichostatin A
Salicylate anti-inflammatory agents	Sulfasalazine, mesalazine, and acetylsalicylic acid

Canonical	Noncanonical
Cancers	Crohn's disease and lung transplant
Ulcerative colitis and Crohn's disease	Polycystic ovary and glioblastoma
	Cardiomyopathy and cancer



遺伝子発現プロファイル比較による
疾患－薬剤関係に基づく計算/DR

疾患群の階層的クラスター

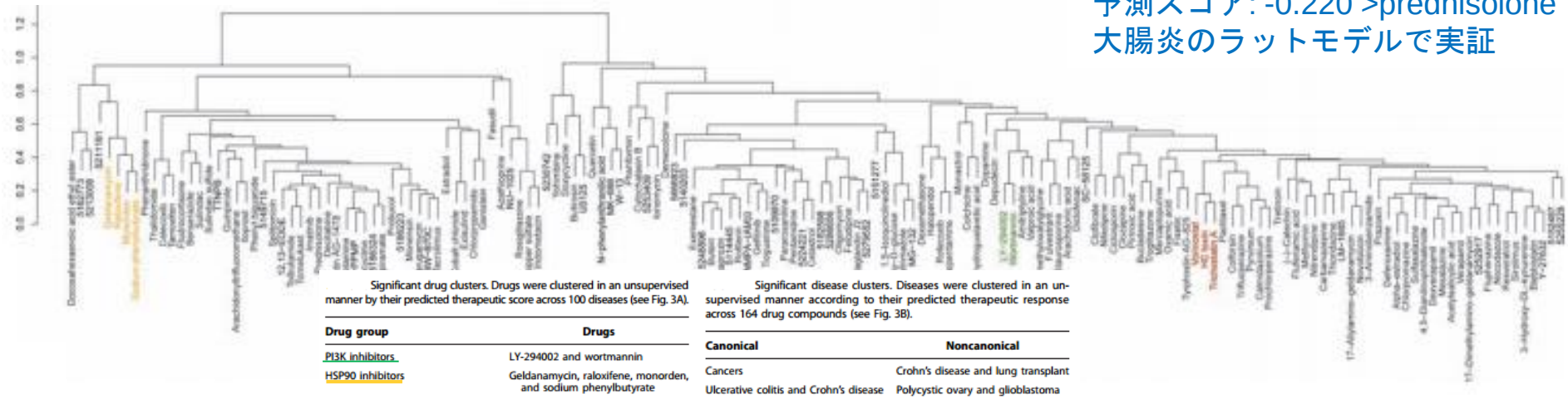


疾患のクラスター解析
- がんの大クラスター
- IBD: 潰瘍性大腸炎、クローン病のクラスター

薬剤のクラスター解析
- HDAC阻害剤
  HC toxin, trichostatin
- PI3K, 抗炎症剤クラスター
- HSP90関連薬剤

DR候補：
Topiramate(抗痙攣剤) IBDに有効
予測スコア: -0.220 >prednisolone
大腸炎のラットモデルで実証

薬剤群の階層的クラスター



Significant drug clusters. Drugs were clustered in an unsupervised manner by their predicted therapeutic score across 100 diseases (see Fig. 3A).

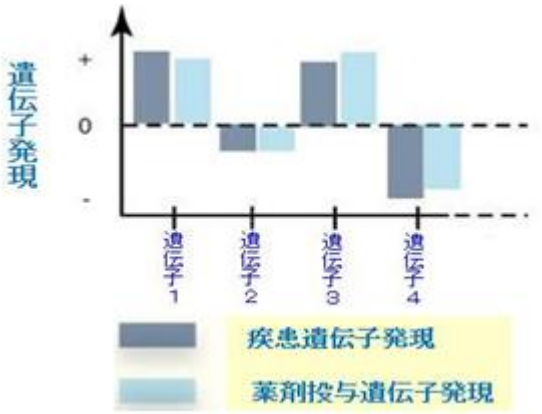
Drug group	Drugs
PI3K inhibitors	LY-294002 and wortmannin
HSP90 inhibitors	Geldanamycin, raloxifene, monorden, and sodium phenylbutyrate
HDAC inhibitors	Vorinostat, HC toxin, and trichostatin A
Salicylate anti-inflammatory agents	Sulfasalazine, mesalazine, and acetylsalicylic acid

Significant disease clusters. Diseases were clustered in an unsupervised manner according to their predicted therapeutic response across 164 drug compounds (see Fig. 3B).

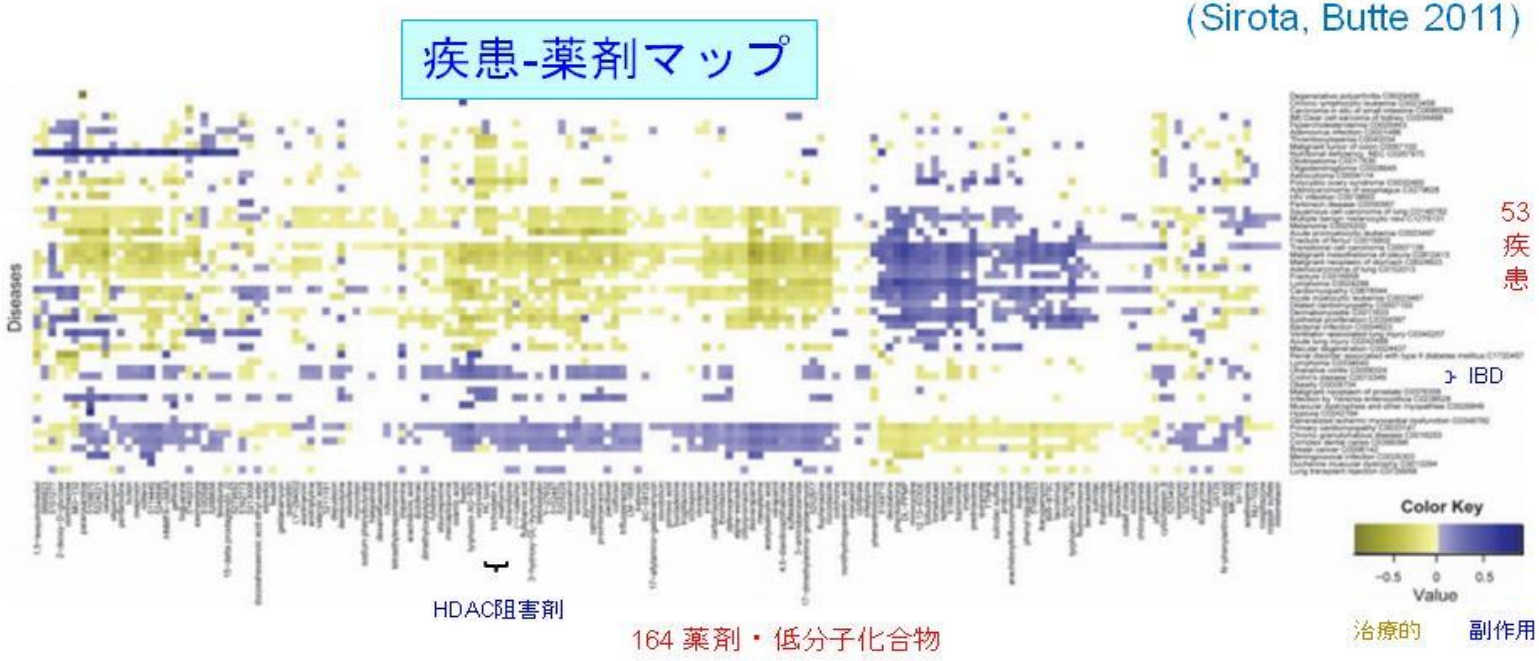
Canonical	Noncanonical
Cancers	Crohn's disease and lung transplant
Ulcerative colitis and Crohn's disease	Polycystic ovary and glioblastoma
	Cardiomyopathy and cancer

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)



遺伝子発現Profiling による疾患ー薬剤ネットワーク（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 - 疾患（disease-disease）645 組
疾患- 薬（disease-drug）5008 組
薬 - 薬（drug-drug) 164,374 組

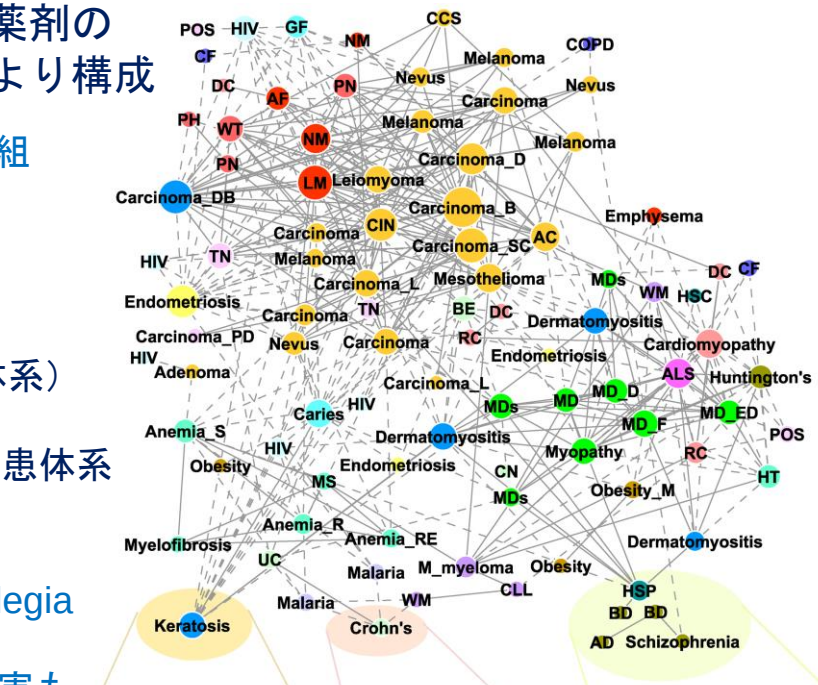
結果

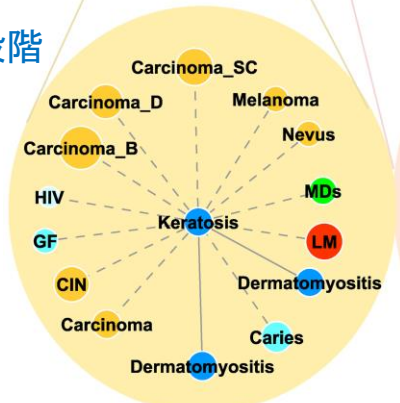
①疾患関連の60%はMeSH (既知体系）
その他は**分子レベル疾患分類学**
Transcriptomeの類似性による疾患体系

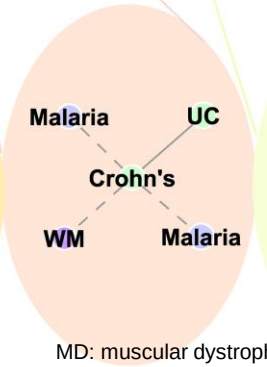
②主な発見

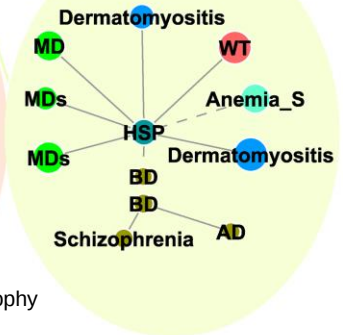
＜疾患 - 疾患＞
HSP（Hereditary Spastic Paraplegia
（遺伝性痙攣性対麻痺）
⇒bipolar 双極性障害 --精神障害も
Solar keratosis 日光性角化症
⇒ cancer(squamous) --前癌段階

＜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









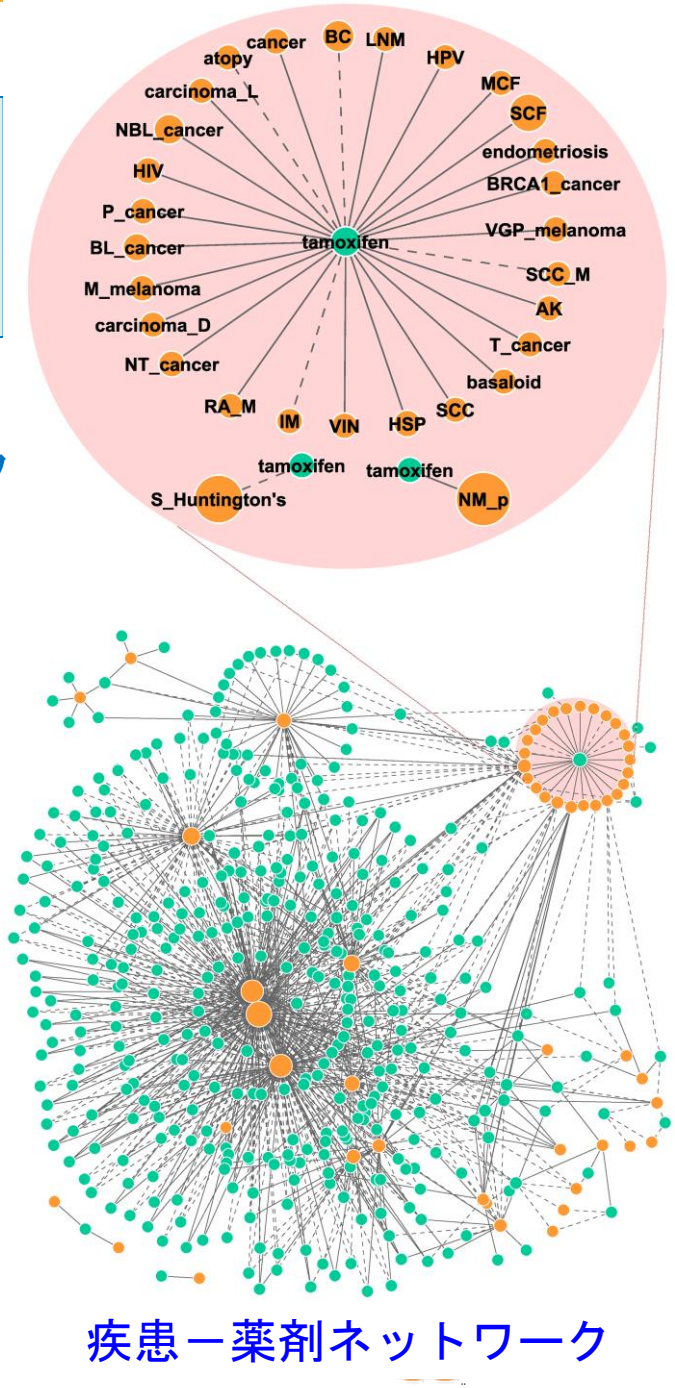
MD: muscular dystrophy

カラーはMeSH同一カテゴリー
実線はMeSH内
破線はMeSH外

遺伝子発現 Profiling による Drug-Diseaseネットワーク

疾患－薬剤および薬剤－薬剤ネットワーク
（Disease-drug network:右図）
橙色節　49 疾患,　緑色節　213 薬剤
906　疾患－薬剤結合
実線　正値　破線　負値

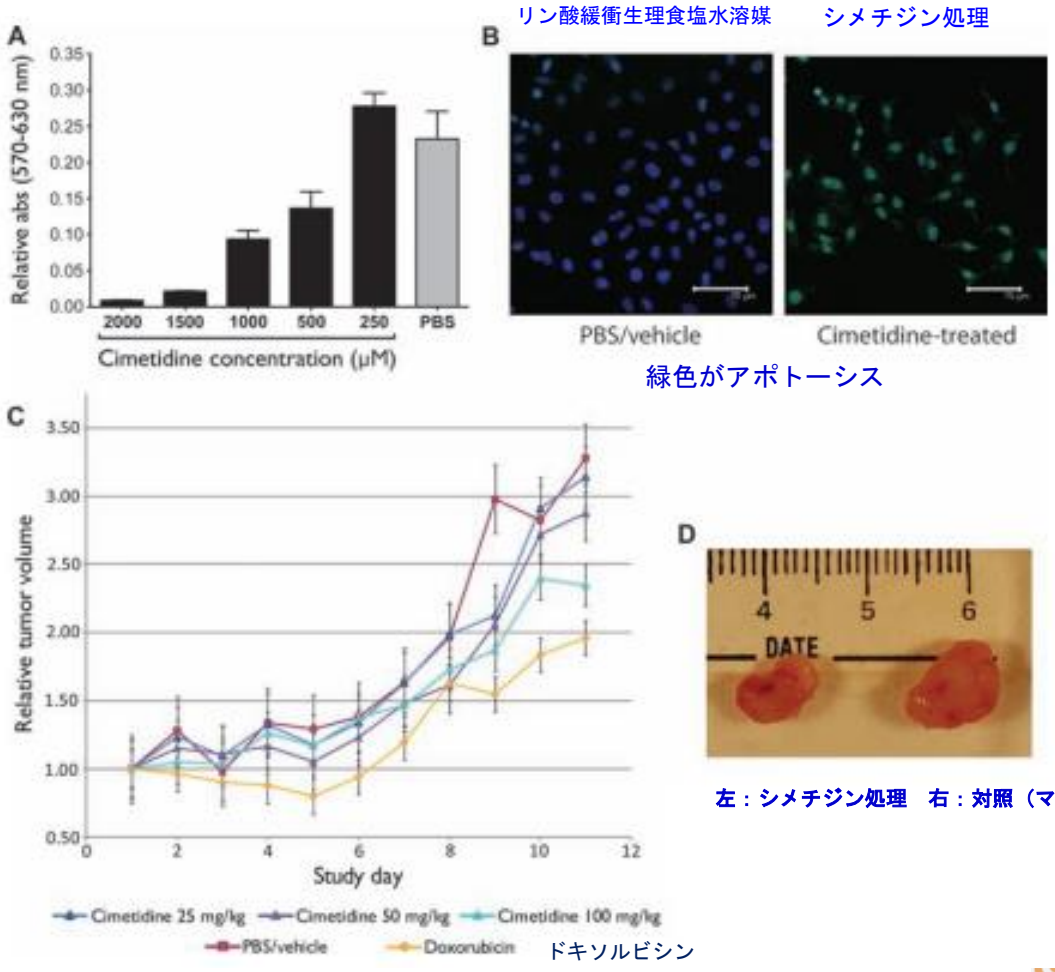
Tamoxifen (乳がん医薬)
有効性　負の値をもっている
⇒アトピー,
⇒マスト細胞分泌抑制、
アレルギー抑制
Hunting病に多数のDR薬
副作用　正の値をもっている
副作用の予測
⇒ 発癌性



疾患－薬剤ネットワーク

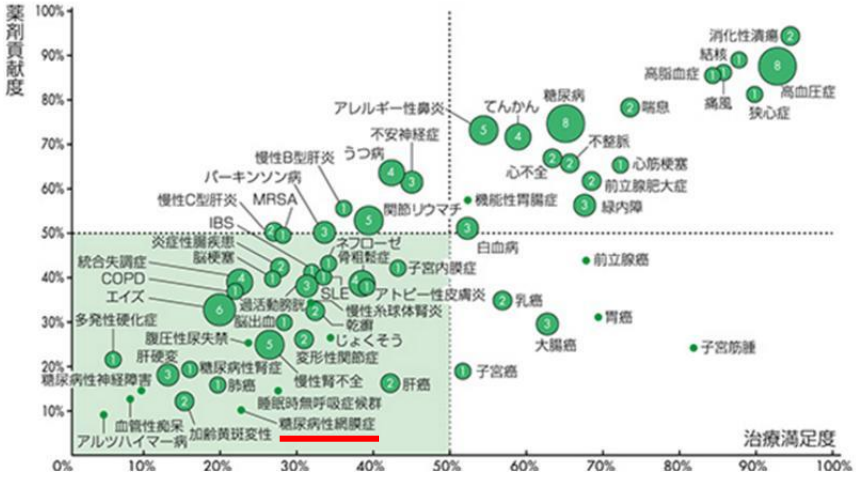
動物実験での実証

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



我々の研究室での成果

- 対象疾患　(Shibata et al. 2016)
  - 薬剤貢献度と治療満足度がともに低い糖尿病性網膜症 (diabetic retinopathy) の薬剤探索
- 方法
  - Signature revision法を適用
  - **疾患特異的遺伝子発現**
    - GEOから糖尿病性網膜症の遺伝子発現プロファイルを収集 (GSE53257)
    - 対照：16サンプルの健常例
    - 206遺伝子 疾患signature を確定
      - 130 up-regulated
      - 76 down-regulated genes
  - cMAPより疾患と負値ESの薬剤特異的発現を提示する有意な薬剤を探索
- 結果
  - 1600組のなかで37組の＜疾患 - 薬剤＞が有意、その中でも**11剤が負値のES**
  - FDR（*q*値）＜0.005
  - thapsigargin (score -0.983, p-value 0.00002),
    タプシガルギン（小胞体ストレス誘導）
  - alprenolol (score -0.892, p-value 0.00026),
    ionomycin (score -0.896, p-value 0.00208),
    phenylpropanolamine (score -0.814, p-value 0.00219)
    など
- 考察
  - thapsigargin ：endoplasmic reticulum (ER) ストレスに関与。ER stress は NF-kB を活性化
  - 糖尿病性網膜症は本質的には炎症反応
  - NF-kB は the unfolded protein response (UPR) で制御されている .
  - ER stress がこの炎症の制御に役立つ可能性がある



ライフサイエンス振興財団

糖尿病性網膜症のDR候補化合物

	薬剤	SCORE	p 値
1	thapsigargin	-0.983	0.00002
2	alprenolol	-0.892	0.00026
3	ionomycin	-0.896	0.00208
4	phenylpropanolamine	-0.814	0.00219
5	etiocholanolone	-0.621	0.00961
6	kinetin	-0.72	0.01249
7	triflupromazine	-0.706	0.0155
8	vanoxerine	-0.681	0.02274
9	cicloheximide	-0.657	0.03185
10	khellin	-0.579	0.03975
11	rotenone	-0.625	0.04852

TMDU

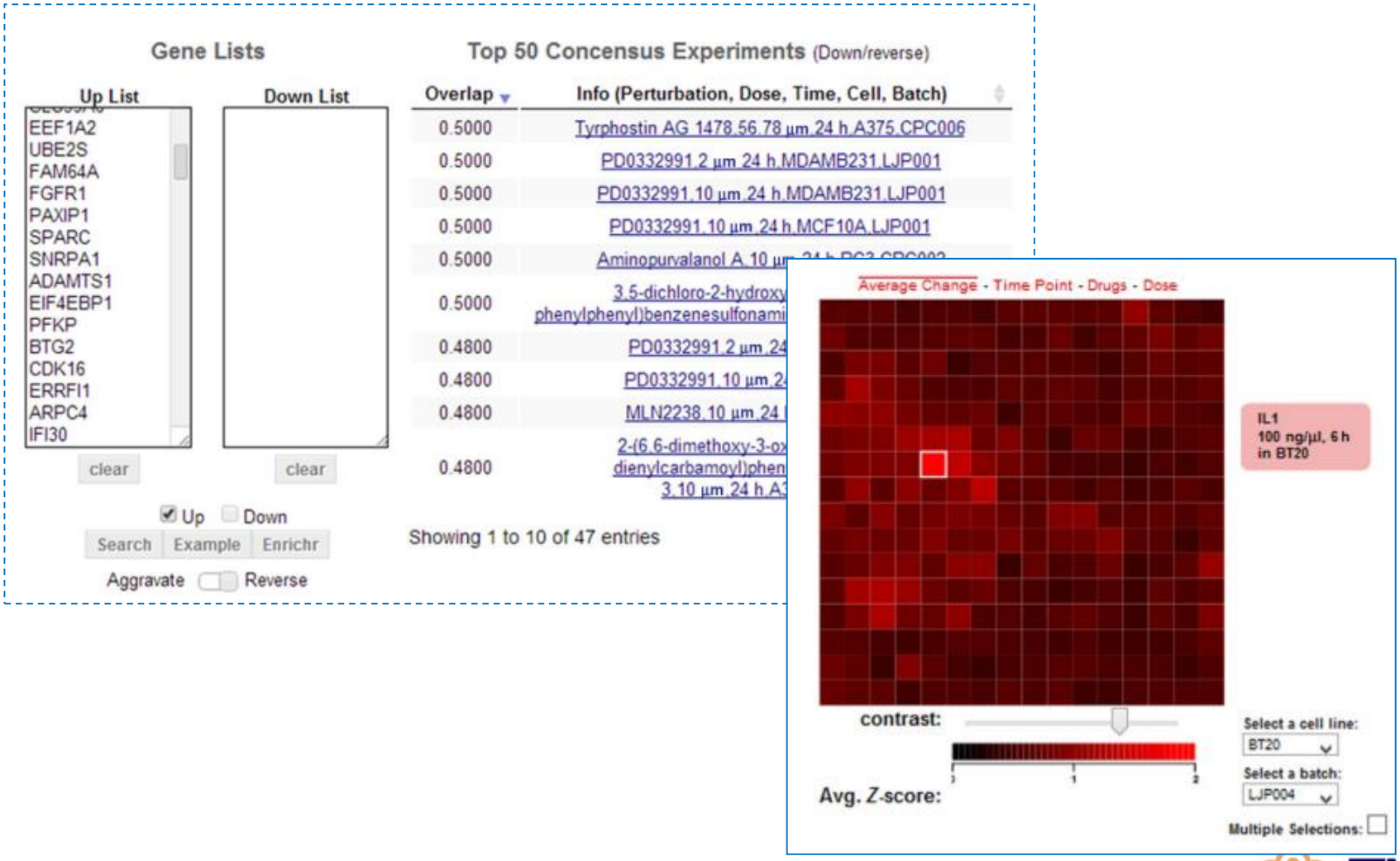
近年のビッグデータ化
LINCS

- **LINCS** (library of Integrated network-based cellular signatures)
  - GE-HTS(gene expression high throughput screening)の１つ
  - 摂動（化合物添加）を与え調節系を介して、細胞表現型を観察する
  - 遺伝子発現変化⇒差異的発現　**signature**
  - cMAP (2006, Lamb)に比べてスケール拡大
    - cMAPは、４つの細胞系列～　1300化合物　FDA認可薬剤
      Micro array (mRNA) Affymetrix U 113で遺伝子発現測定
- NIHから助成, **100万の遺伝子発現プロファイル**を **L1000 技術**で測る
  - Broad Institute cMAPと同じメンバーが考案
  - 1000遺伝子の発現しか測定しない　ゲノムワイドな遺伝子発現プロファイル（～全遺伝子22000 genesの発現）をGEOから作ったモデルで推定する
  - 相互依存性高い⇒1000遺伝子にすべて情報が含まれている
- **L1000技術**
  - 細胞溶液からリガンド媒介増幅によってmRNA増幅
  - 遺伝子特異的なProbeはcDNA（mRNA) にtaqリガーゼでアニールする
  - ProbeはPCRで増幅され、ルミネックスビーズと遺伝子特異的部分で対形成する
  - 対形成した差異染色ビーズはレーザを用いて検出され定量化される
  - ビーズの上の対形成したprobeの密度を測る　80の恒常的発現校正遺伝子
- **22412 摂動遺伝子発現**
  - 56　細胞コンテキスト（ヒト初代培養細胞、がん培養細胞）について
  - 16425　化合物、薬剤
  - 5806　遺伝子ノックアウト(RNAi, miRNA)、過剰発現
  - 総計で100万ぐらい遺伝子発現プロファイルがある
- **Genometry がL1000™ Expression Profiling技術でヤンセンと契約**
  - 25万種類の化合物

TMDU

LINCSの問合せ画面

--- LINCS Canvas Browser ---



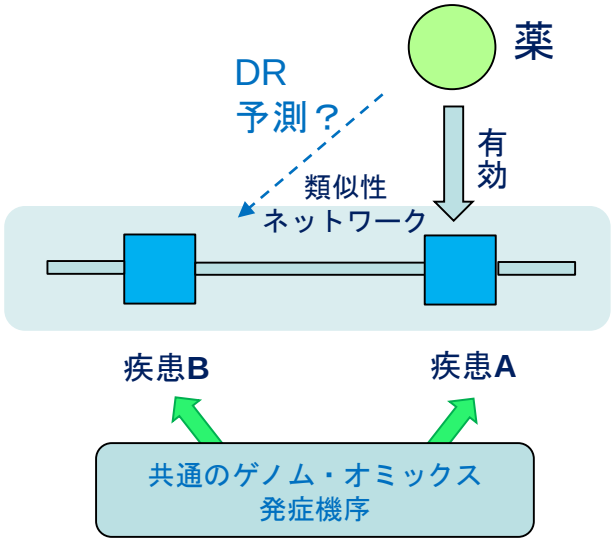
２．疾患・薬剤ネットワーク近接解析

遺伝子発現プロファイルから疾患・薬剤ネットワークへ

- 遺伝子発現プロファイルを基礎にした「シグネチャア逆位法」
  - 広くDR候補を**枚挙する**方法としては有効
  - もう少し**対象を絞り込む**方法はないか
  - 疾患の**内因的機序**(ゲノム・オミックス機序)によるDRは可能か
- 疾患のゲノム・オミックス機序による**疾患間のネットワーク化**
  - **同様に薬剤間のネットワーク化も可能**

ビッグデータ創薬/DR
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



疾患形成のゲノム・オミックス機序

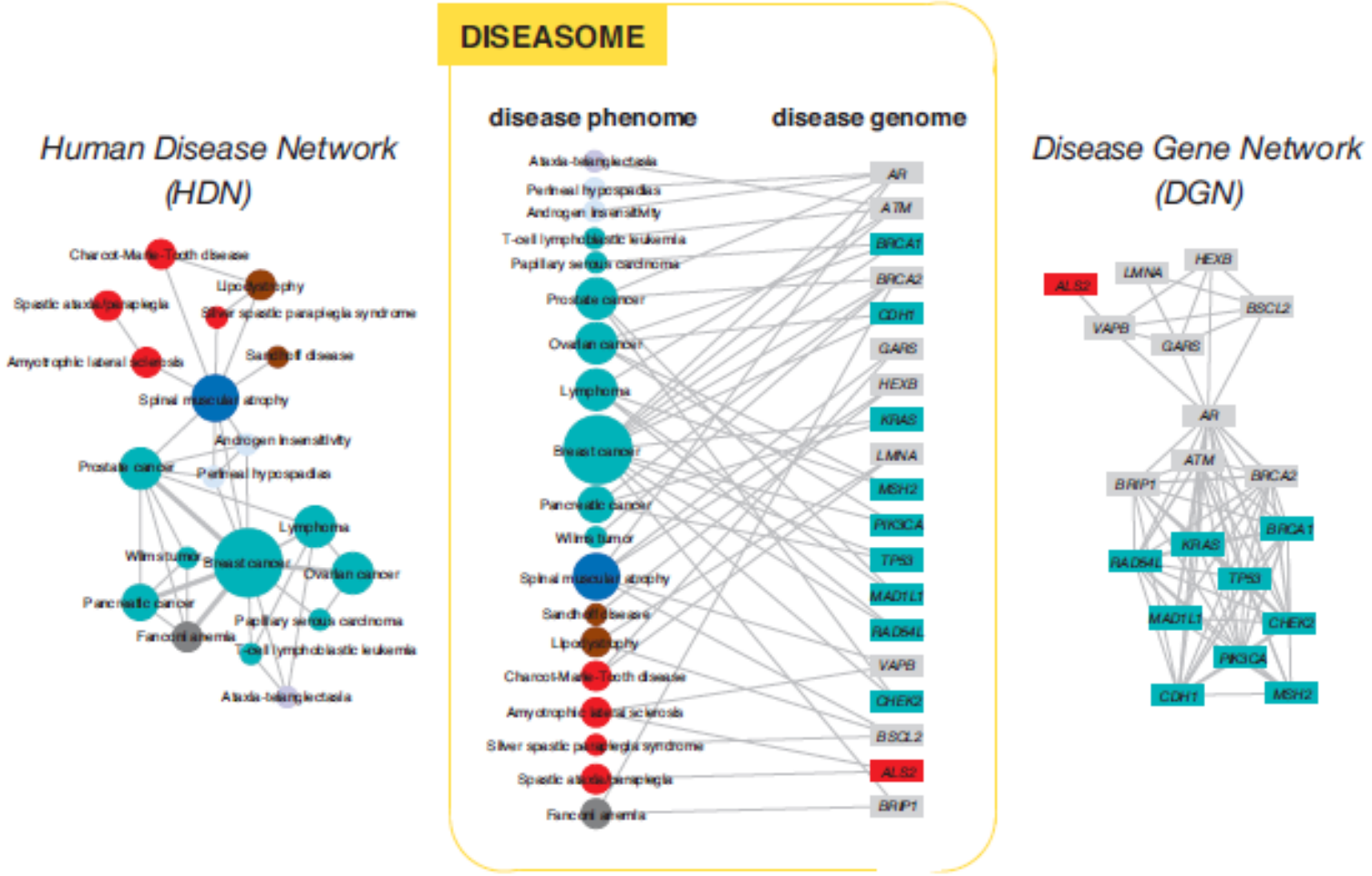
- 疾患関連遺伝子型（第１世代型）
  - 原因遺伝子、疾患感受性遺伝子の変異・多型が主要発症機序
- 疾患オミックス型（第２世代型）
  - 疾患オミックスプロファイルの変容が主要発症機序、「分子表現型 molecular phenome」
  - Trans-disease omics
- 疾患分子ネットワーク型（第３世代型）
  - 「分子ネットワークの歪み」が主要発症機序
  - 大半の疾患（先天的稀少疾患を除く）, ”common disease”はネットワークの歪み

TMDU

第１世代型
疾患原因遺伝子準拠
Diseasome

- **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

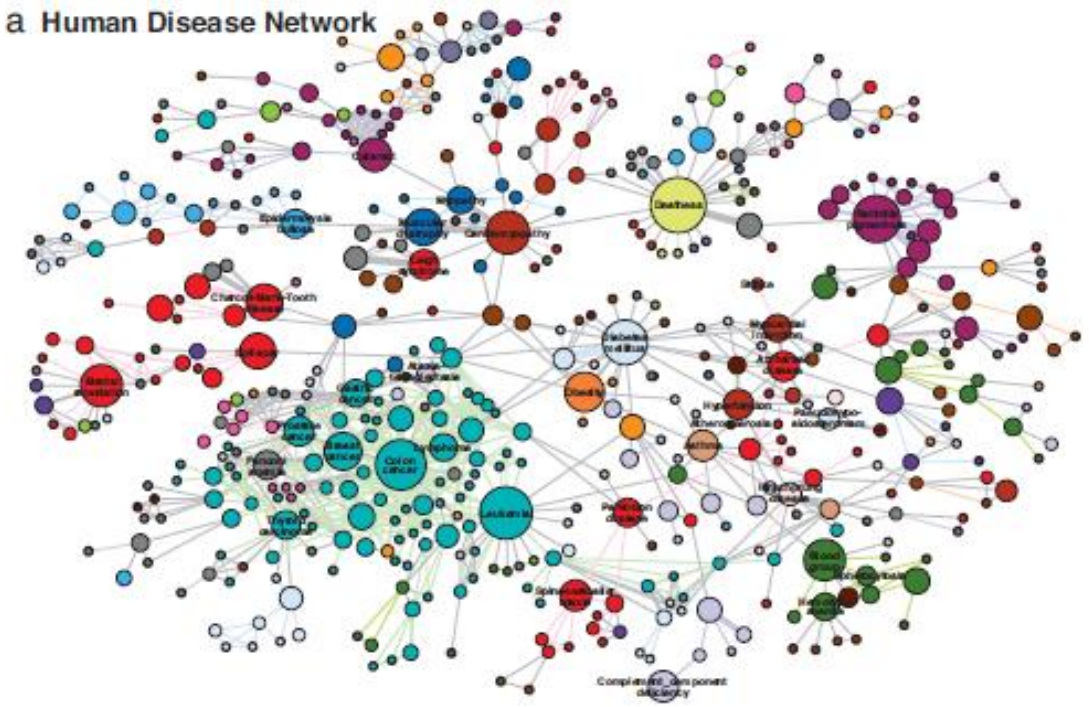
１つ以上の疾患を共有する疾患関連遺伝子

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

TMDU

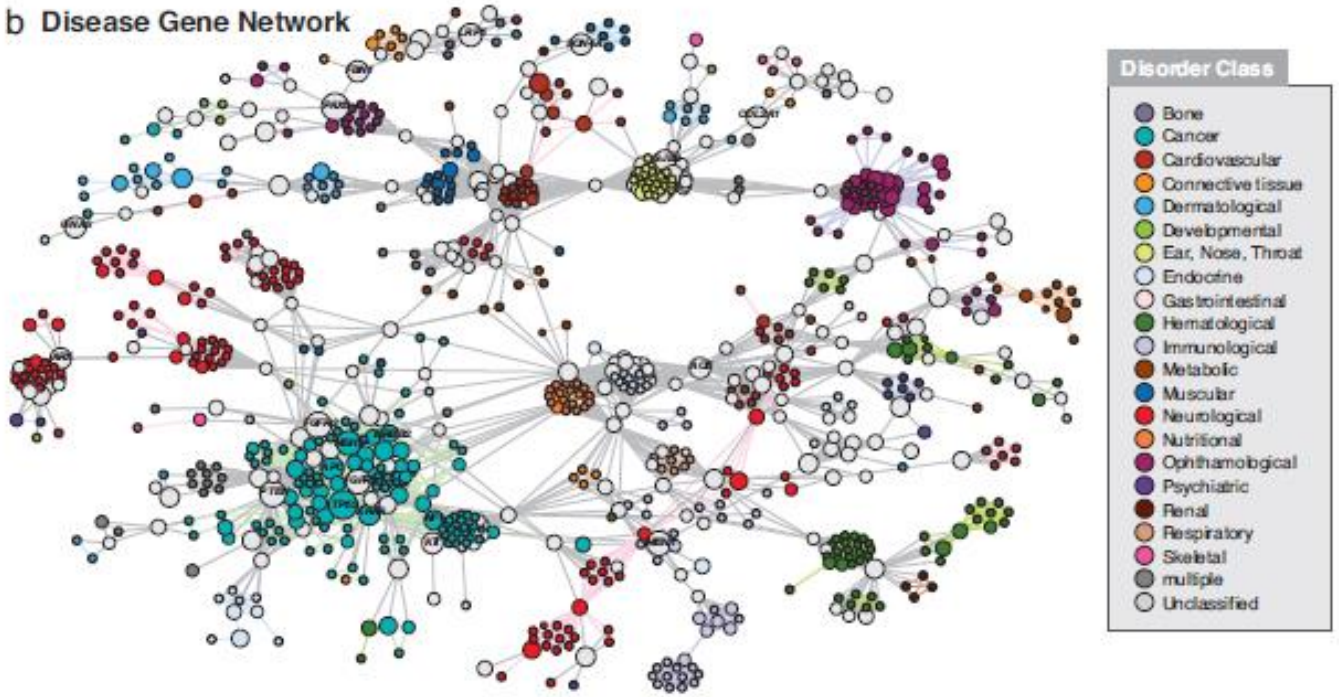
疾患
ネットワーク
(HDN)

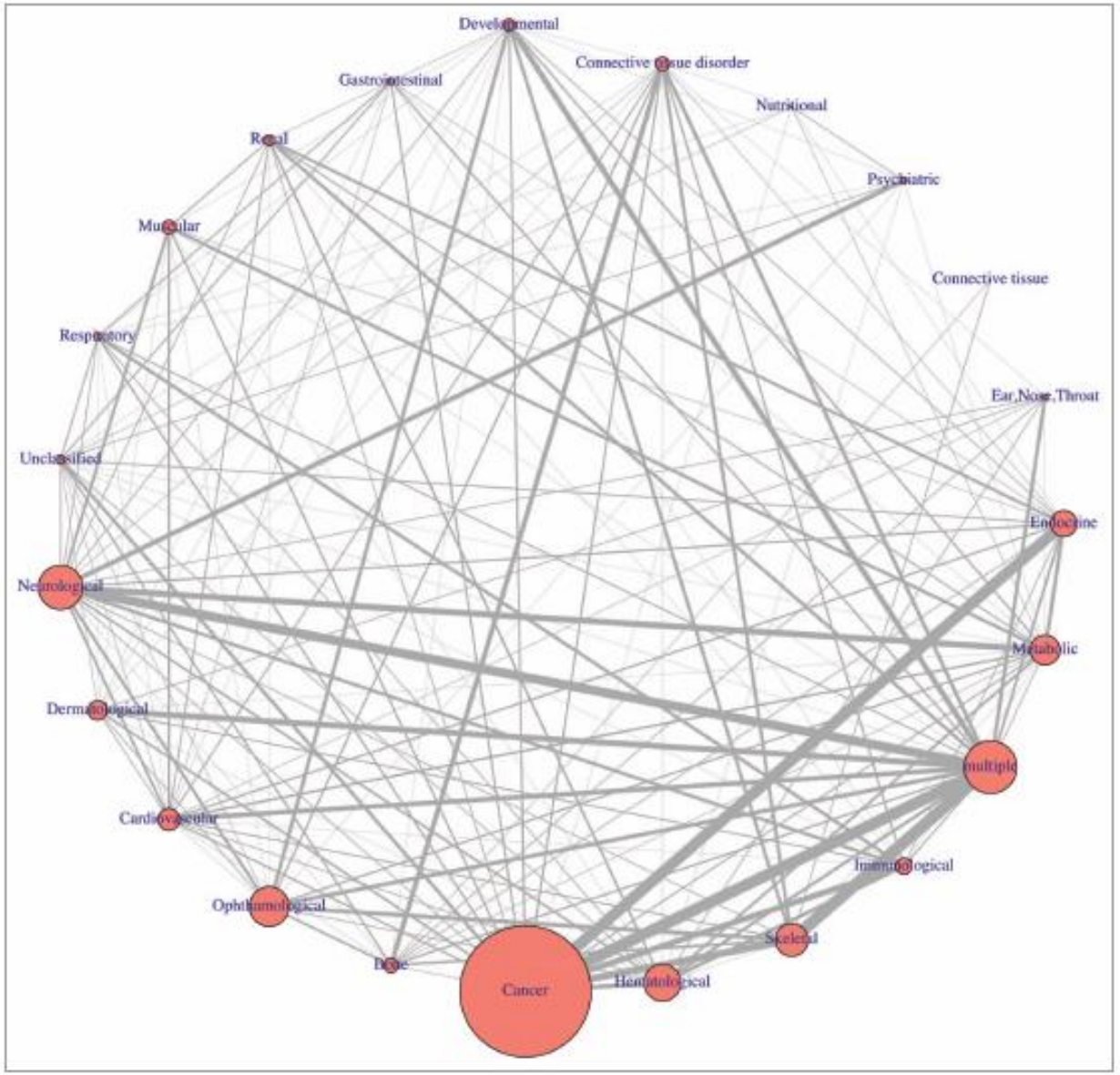
Nodeの直径
疾患に関与している原因遺伝子の数に比例
リンクの太さ
疾患間で共有している原因遺伝子の数



疾患遺伝子
ネットワーク
(DGN)

Nodeの直径
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





23グループのOMIM疾患のネットワーク

ノードの径：共通リンクの合計、リンクの太さ：共通遺伝子数

TMDU

Diseasomeを巡る状況

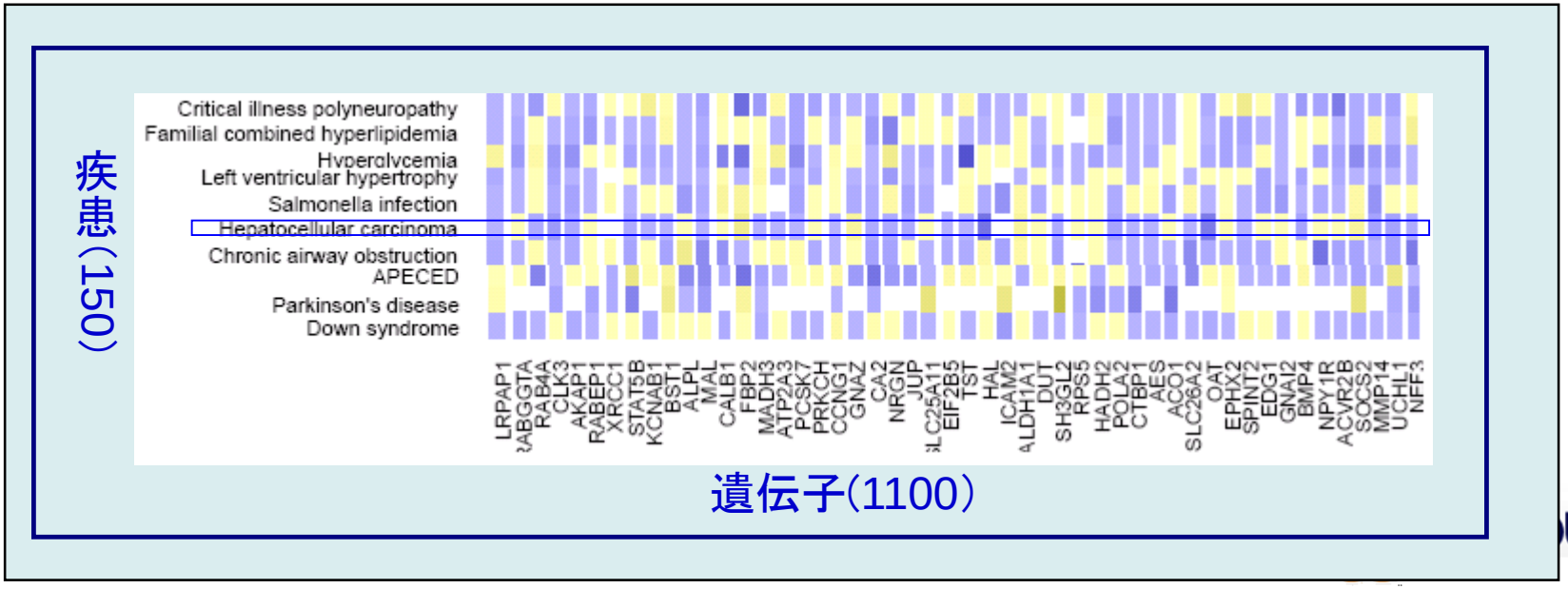
- Mendel疾患、複雑疾患、環境疾患へと発展
- **他のネットワークと融合**
  - タンパク質相互ネットワーク、代謝ネットワーク
    - PPIの近傍(Vanunu)、代謝網での酵素の基質の共有
  - GWAS（WTCCC, NIH-GAD)のSNPの共有
    - すべてがつながり偽陽性のネットワークで有効性低い
  - miRNA, 環境因子（annotation MEDLINE）
    - 電子カルテから時系列病歴収集
    - 進化的直系的表現型性（他の動物も利用）
  - パスウェイ準拠型の疾患ネットワークも
- **表現型疾患ネットワークも存在する**
    - Phenotype：MeSHの頻度をベクトルとする(van Driel)
- Diseasomeは、臨床表現型NWと分子NWを繋げる機構
    - 遺伝子を通して疾患間を移動できる
    - Systems pathobiology, nosology, personalized medicine

第２世代型

遺伝子発現プロファイ型

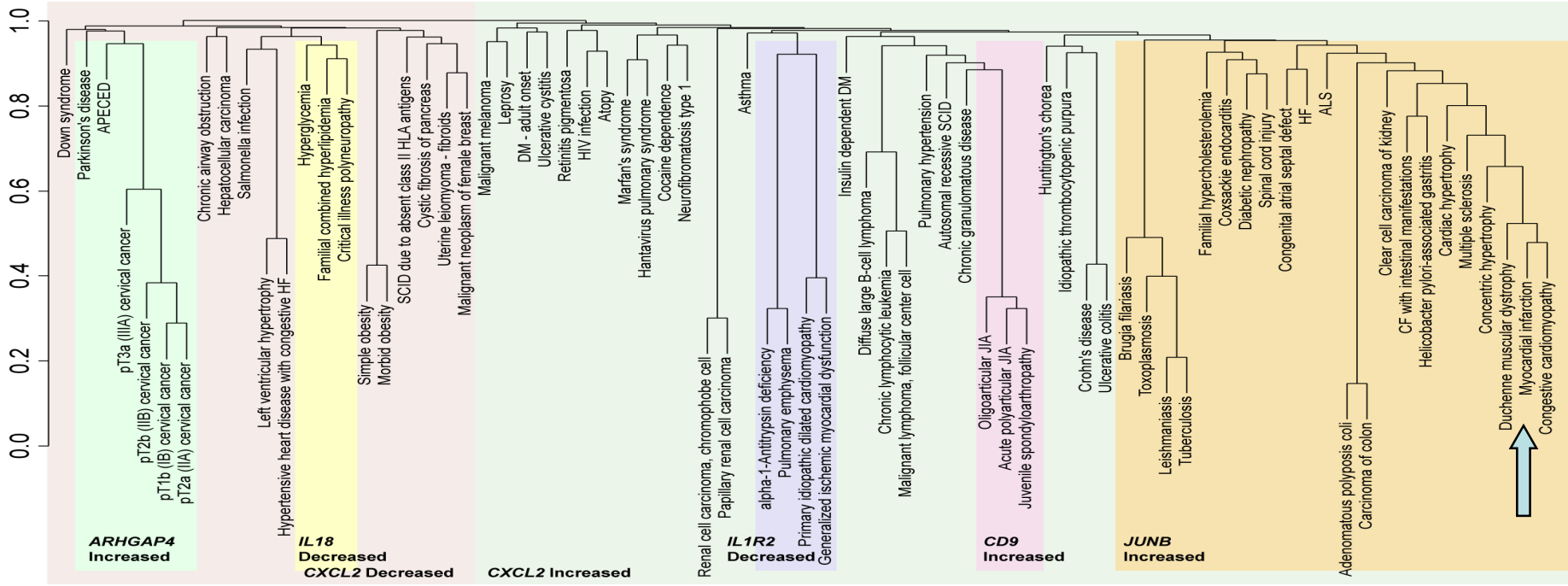
GENOMED (A.Butte et al)

- 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



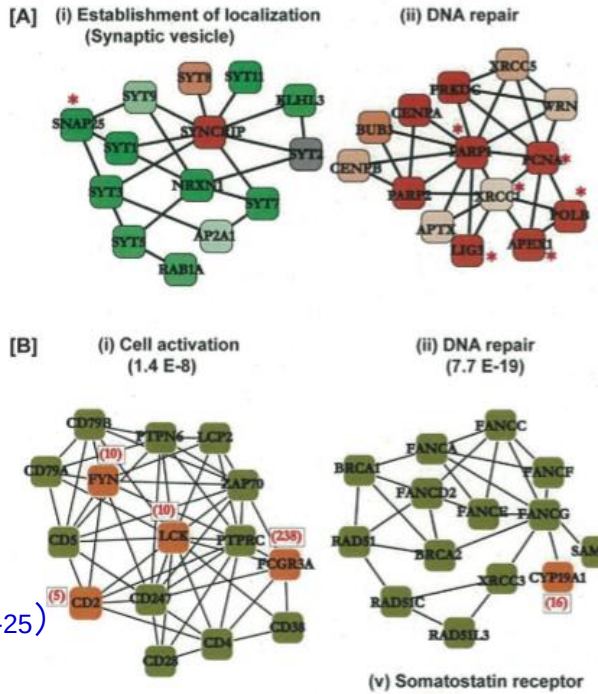
Gene-Expression Nosology of Medicine

- 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



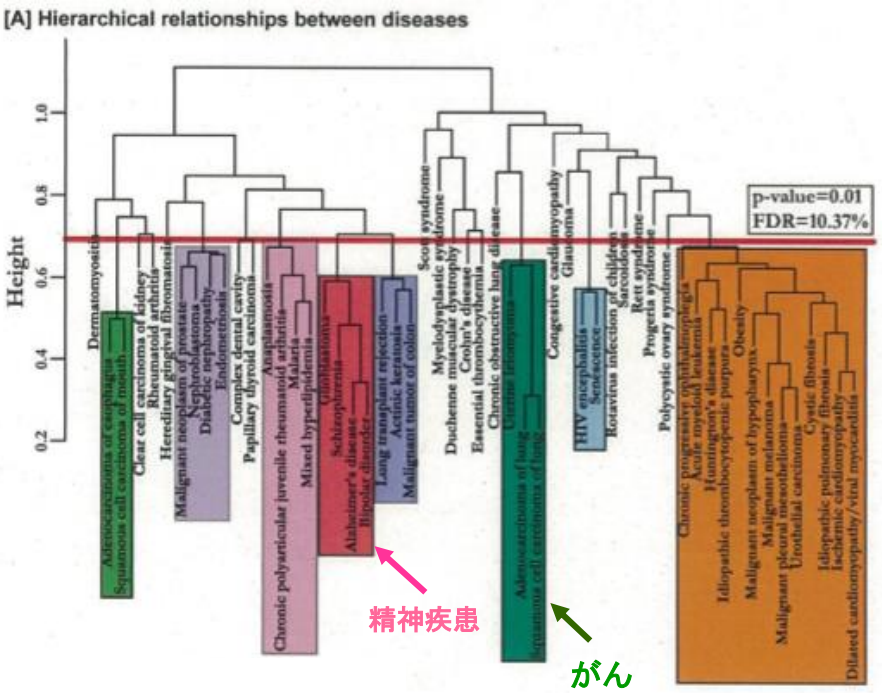
第３世代
疾患分子ネットワーク準拠型（Butte）

- ネットワークモジュール
  遺伝子発現プロファイルではなくPPIを機能4620モジュール に分解
  ＜機能moduleごとの疾病罹患時の平均発現変化＞をもとに
  疾患ネットワークを構築
- **基本方法**
  - GEOから信頼性などより54の疾患を選択
  - 各疾患について各moduleに含まれる遺伝子群の
    疾患時と健常時の発現差の*t* 統計量の平均
  - **MRS:** Molecular Response Score
    各疾患に各モジュールで定義（ベクトル量）
  - 疾患間の相関は、両疾患の健常時発現を制約とした
    MRSの偏相関係数
- **疾患ネットワークの性質**
  - **138の有意な類似性**：ランダム化ネットに対し有意
  - p<0.01,FDR=0.1
  - 疾患類似性: 肺がん群(修復pasM), 精神疾患(synapsM:SNAP-25）
- **138の有意な疾患相関**
  - 17は少なくとも1つの共通薬：14疾患は共通の薬剤に有意
  - Flourarcil（日光性角化）⇒大腸がん、ほかDoxorubicin
- 疾患の大半を占める**59モジュール**：「共通“疾患状態”モジュール」
  - 「共通疾患状態シグネチャア」薬剤標的分子に富んでいる
  - この遺伝子群を標的にする薬剤は有意に多くの他の疾患の薬剤にもなっている

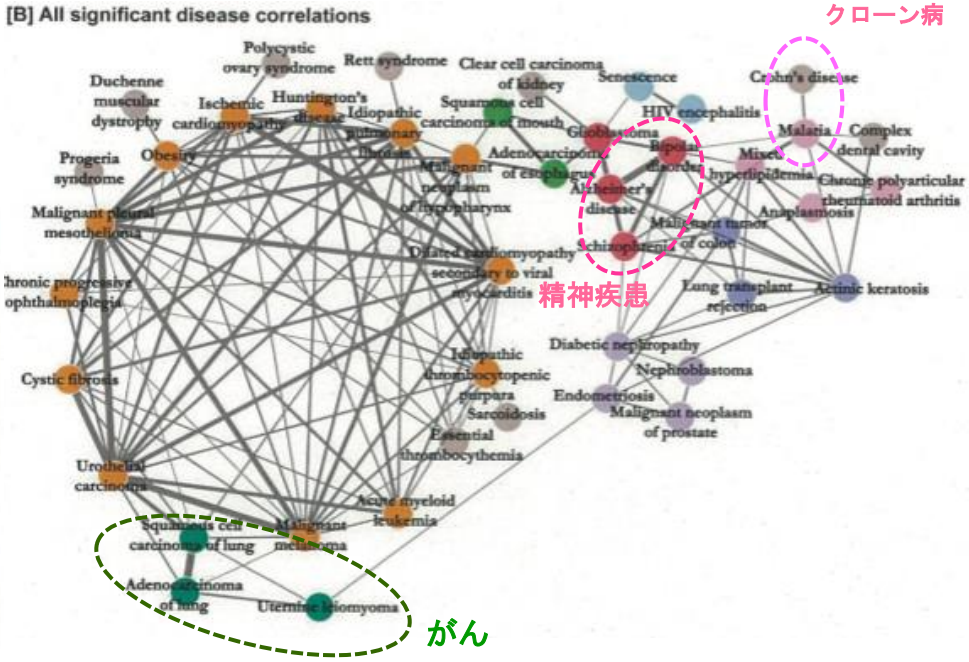


遺伝子発現の変化をPPIに投影した
疾患ネットワーク

- アルツハイマー症、統合失調症、双極性障害がグループ化
- 子宮筋腫と肺がん、マラリアとクローン病
- 17のがんが１つの群ではない。がんの異質性
- 疾患ネットワーク間の遺伝子共有は高くない（遺伝子外効果）



階層的クラスタリング



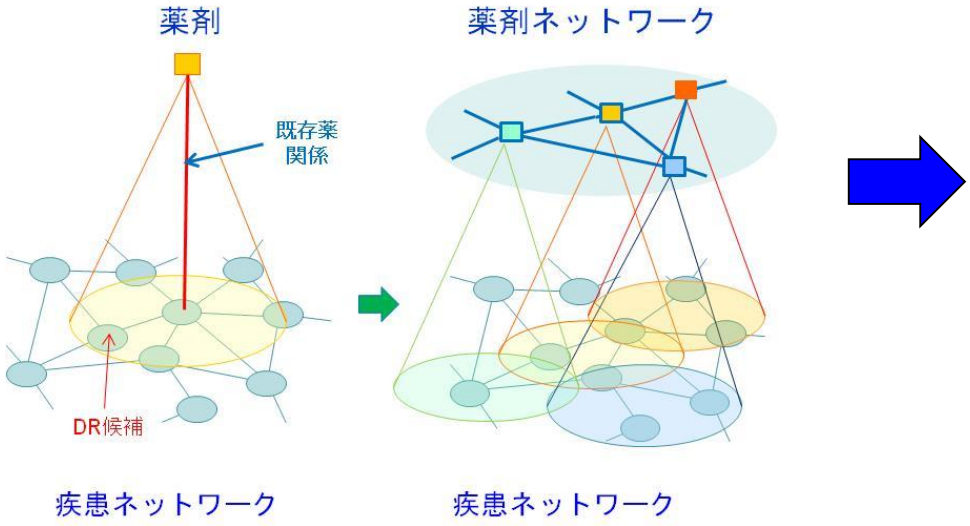
疾患ネットワーク

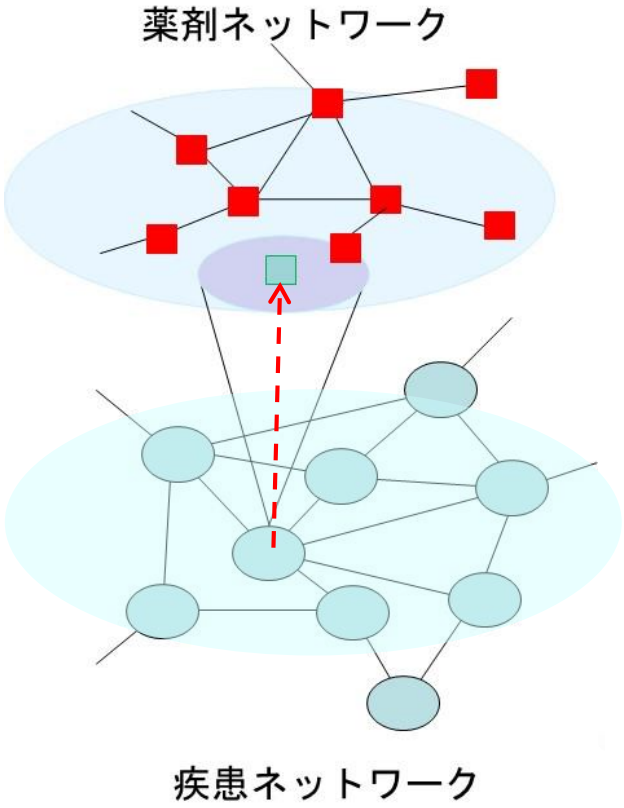
TMDU

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(finger 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**





ビッグデータ創薬・DR
（非学習型アプローチ）

生体分子ネットワーク準拠型

TMDU

疾患-薬剤ネットワーク対応から
生体分子ネットワークを基盤とする創薬/DRへ

疾患ネットワークに準拠した創薬/DR

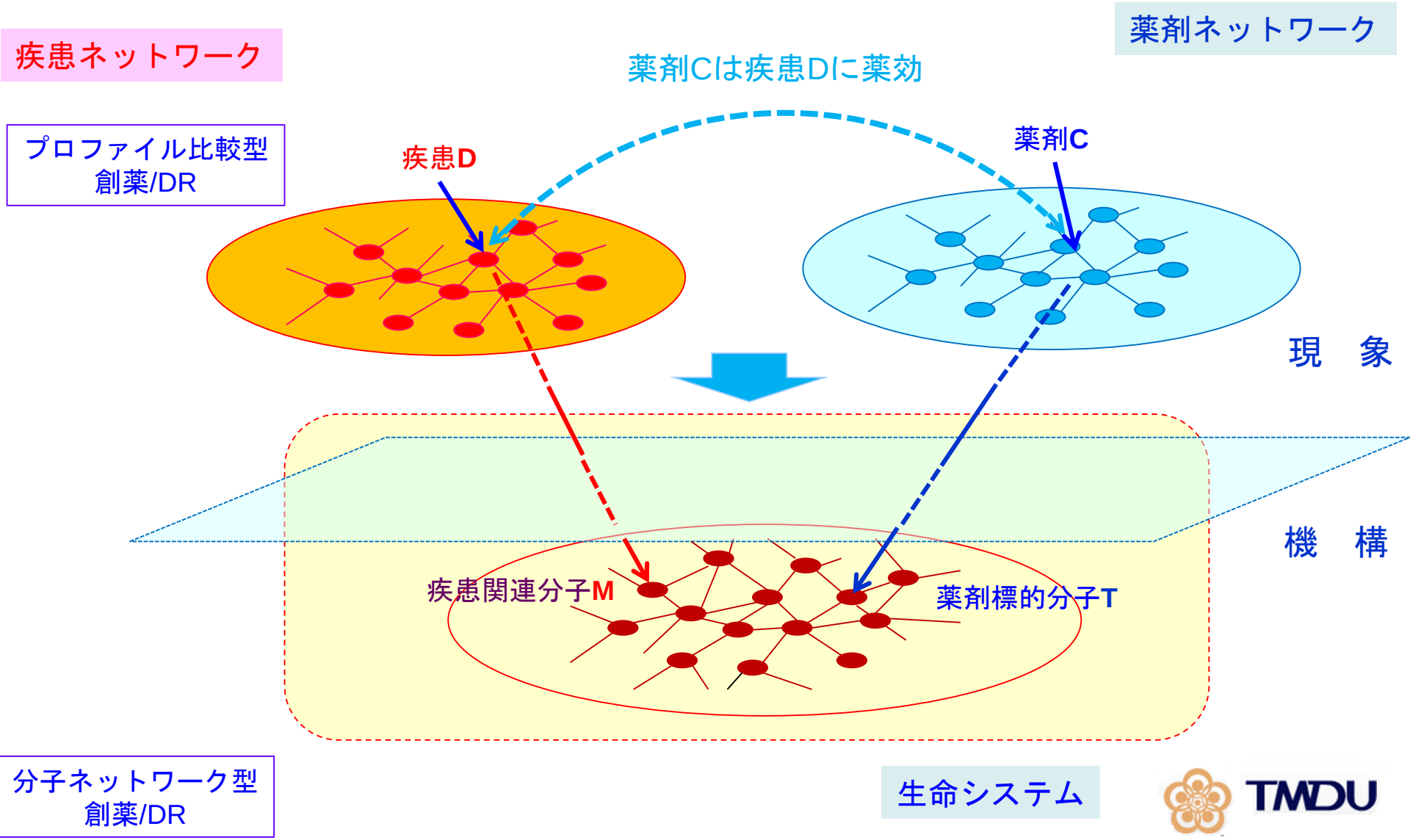
- 疾患のゲノム・オミックス機序に基づいている→内因的機序を考慮した点で評価
- しかし「疾患関連遺伝子」と「標的分子」の相互作用の関係が明示的ではない

MMOAによる創薬/DRの3層ネット枠組み

- 生命分子ネットワークを作用の＜場＞とする
- 薬剤の足場である＜標的分子＞と疾患の足場である＜疾患関連分子＞との＜相互作用＞を基礎とする枠組み

TMDU

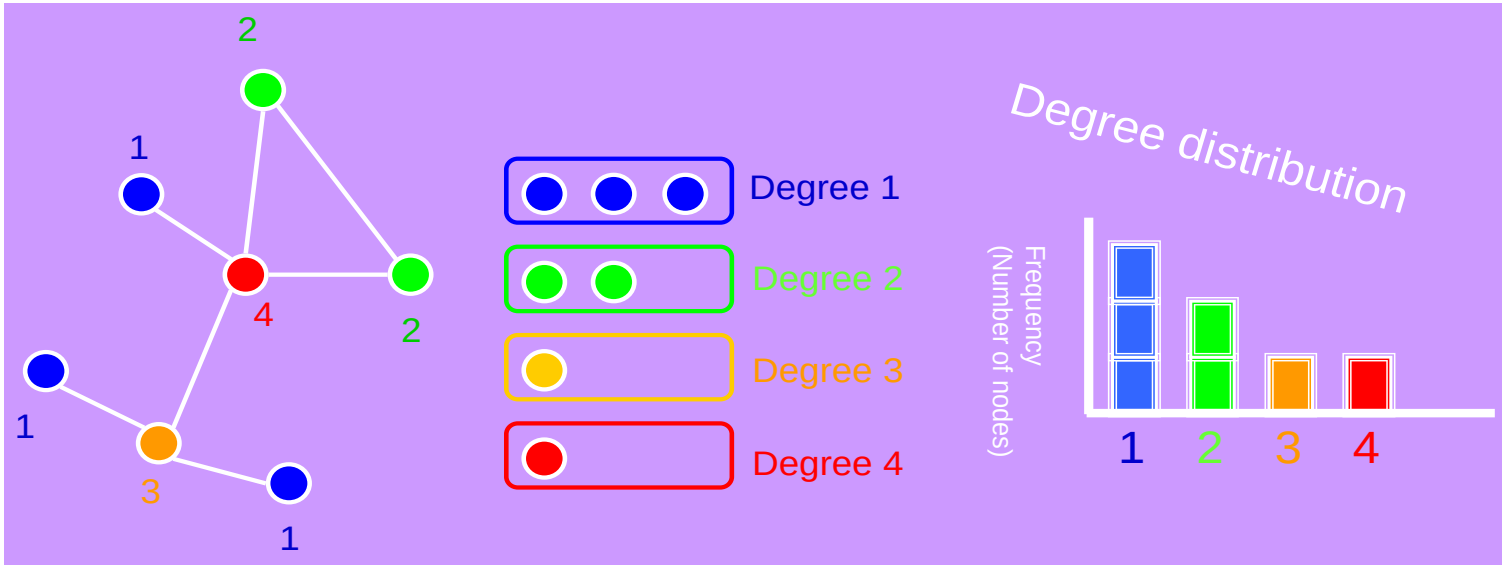
3層の生体・薬剤のネットワーク間の関係図式

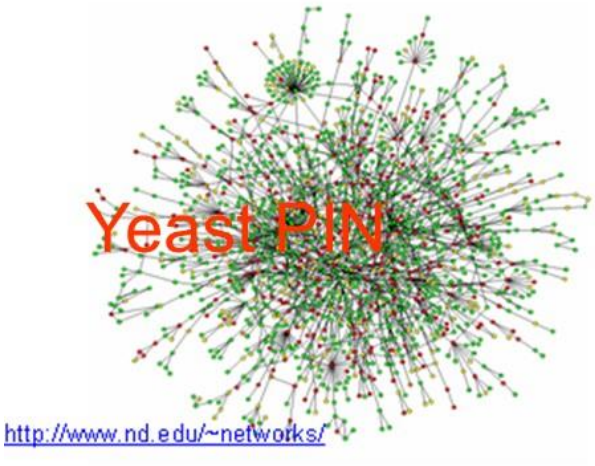


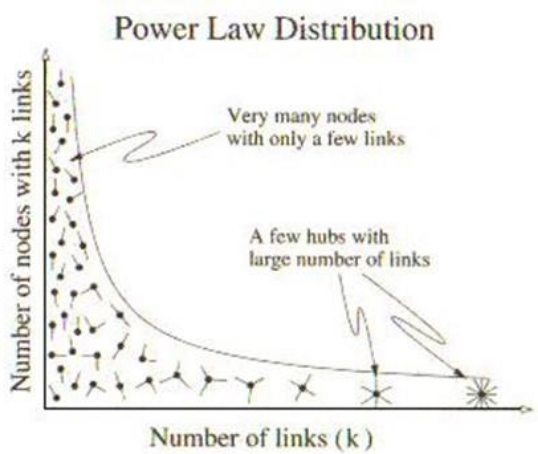
タンパク質相互作用
ネットワークの構造と薬剤標的分子

TMDU

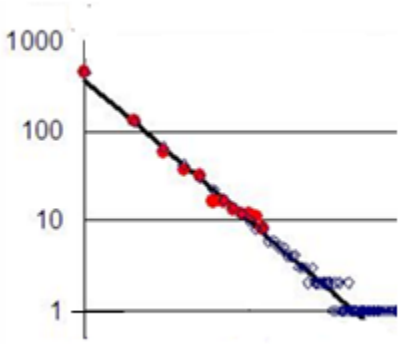
タンパク質相互作用ネットワーク(PIN)では数少ない相互作用が集中したタンパク質(hub)と相互作用が1や2の多数の末端タンパク質(branch)が存在する





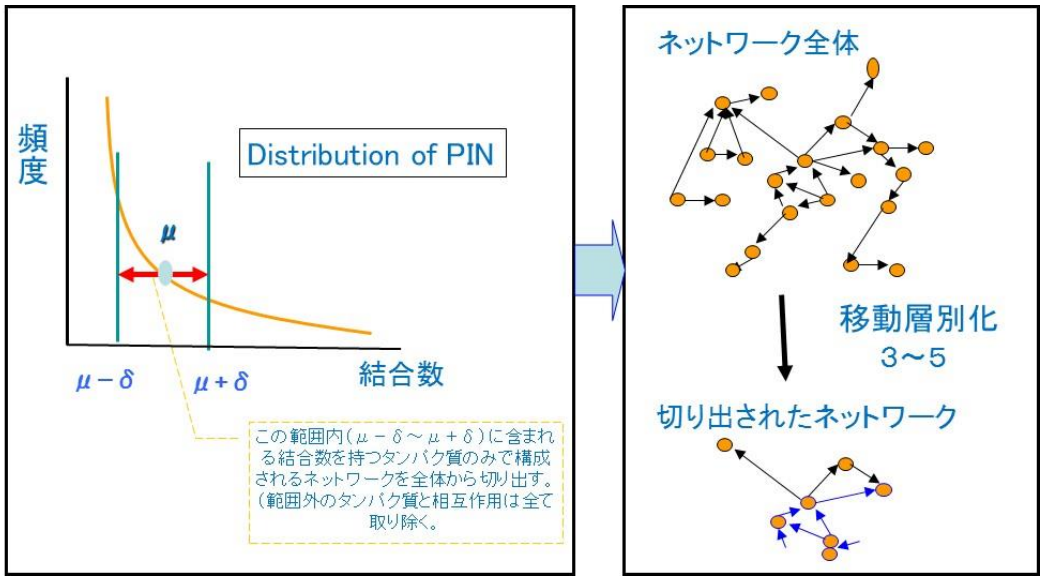


http://www.macs.hw.ac.uk/~pdw/topology/ScaleFree.html

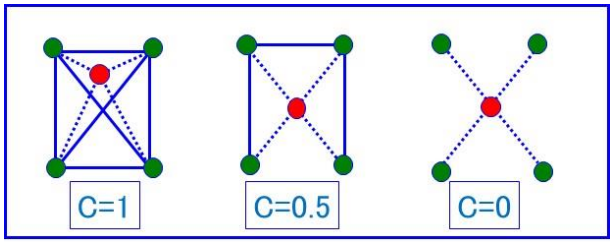


Log-log変換で直線

結合次数ごとの部分ネットワーク構造の結合密度の解析

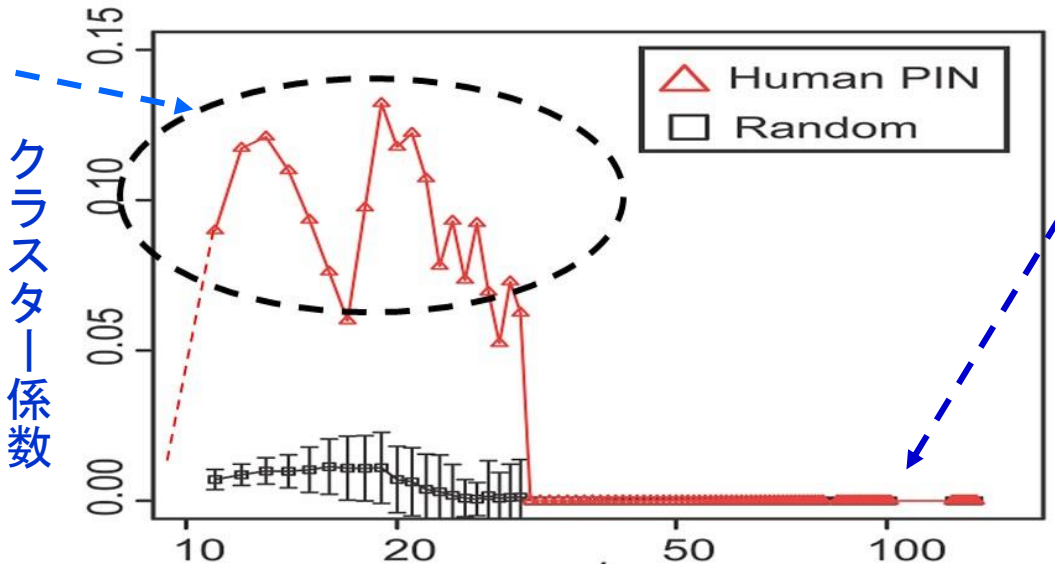


クラスター係数



Hase, T., Tanaka, H et.al (2009)
Structures of protein protein interaction network and their implications on drug design. ***PLoS Compt Biol***. 5(10):

中程度の次数（7～42）を持つタンパク質は多数の密なモジュールを構成

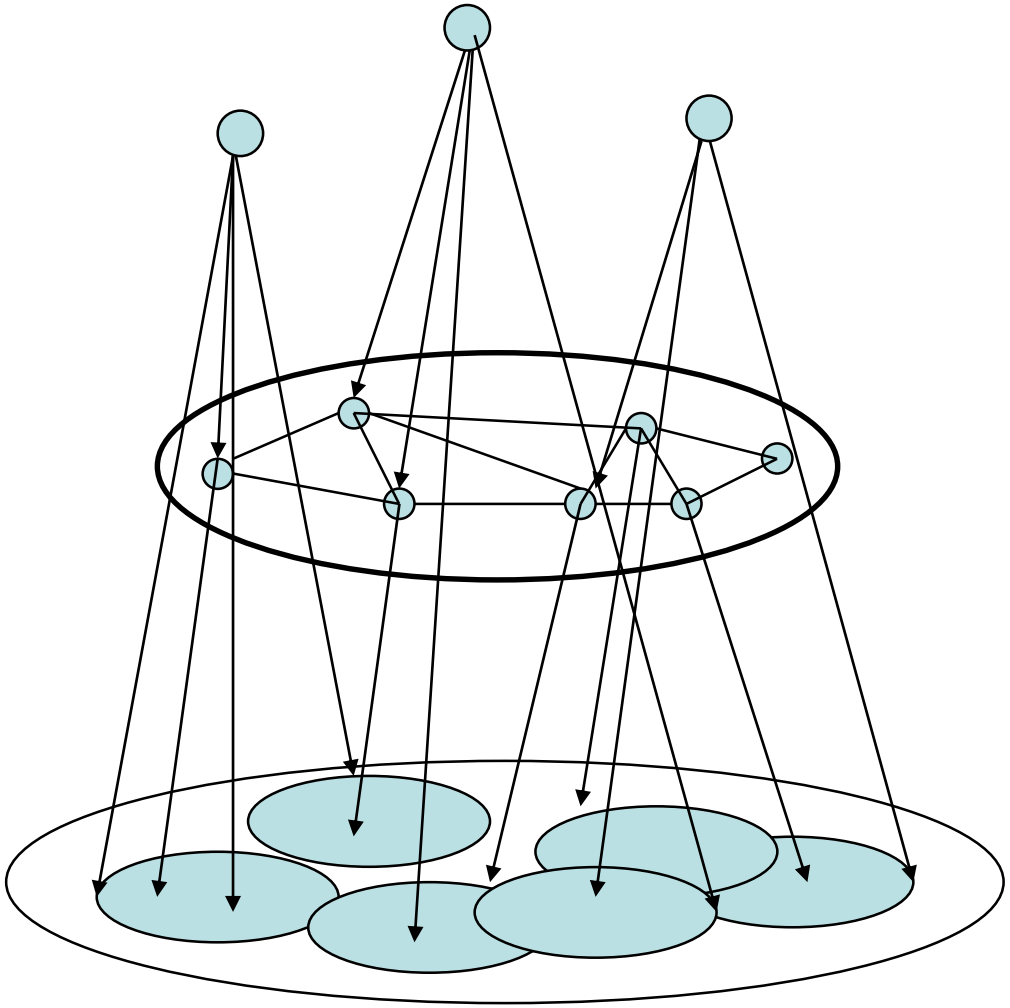


高い次数を持つノード（スーパーハブ）はお互いに密に結合しない

TMDU

タンパク質相互作用から見られる

生命情報ネットワークの構造



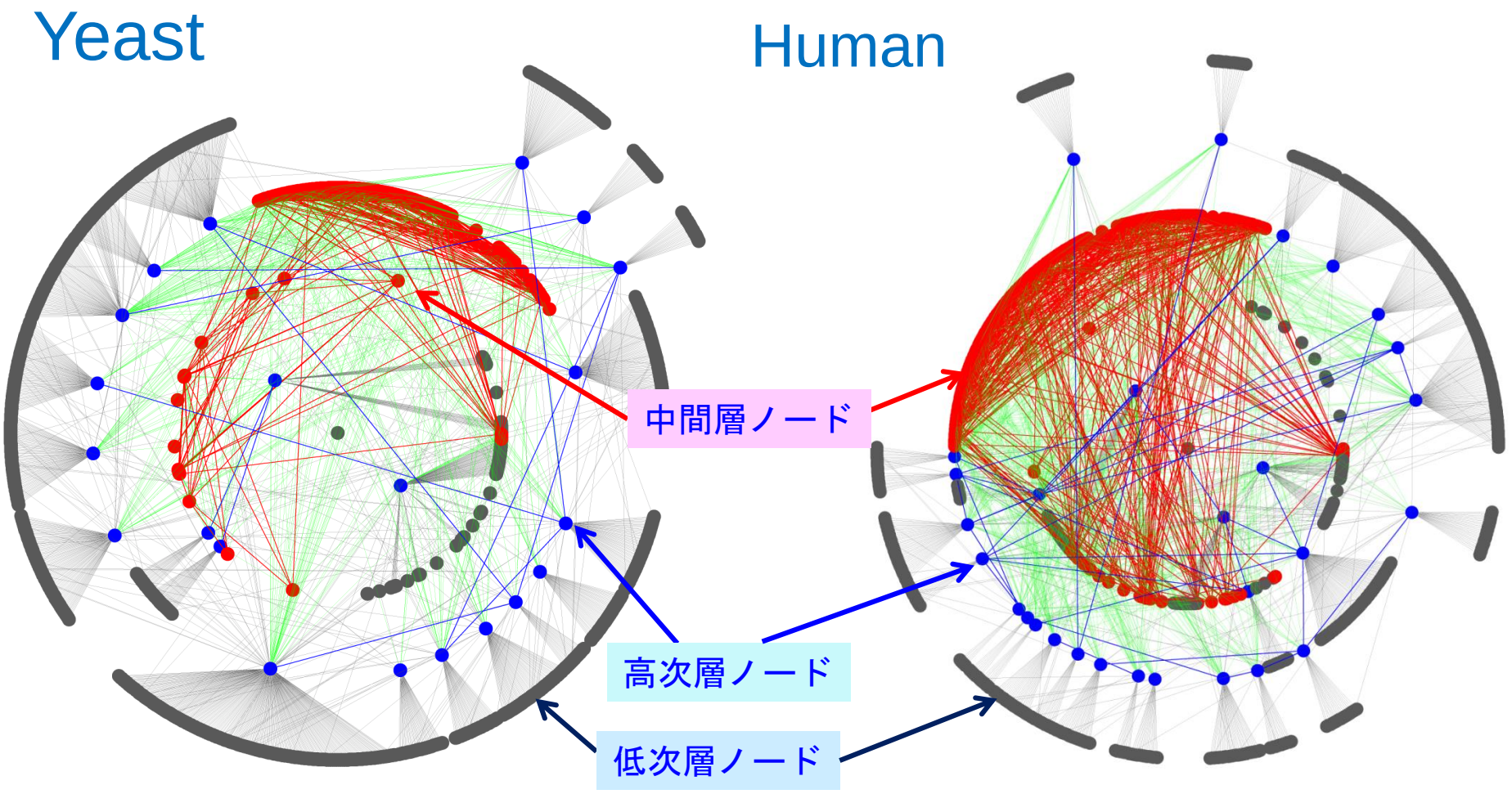
高層
高次数　ハブ
次数
＞31 ヒト
＞39 酵母

中間層
中程度次数
次数
6 ～30　ヒト
6 ～38　酵母

低層
低次数　ブランチ
次数＜6

TMDU

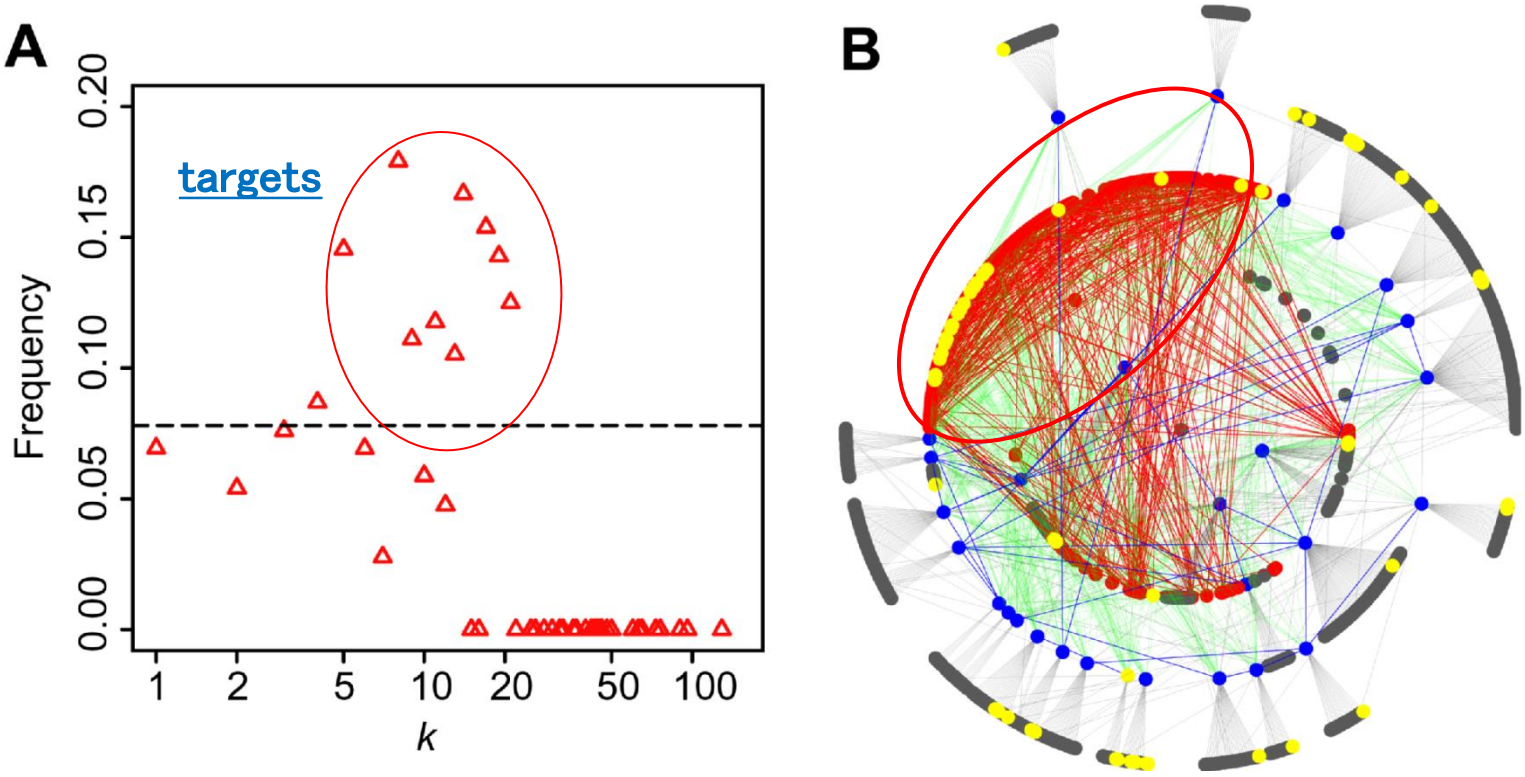
タンパク質相互作用ネットワークの Cloud Topology（3環トポロジー）



中間層の次数 ノードは PPI バックボーン を形成する

灰色, 赤, 青は、それぞれ低層、中層、高層の次数のノードをそれぞれ表す。

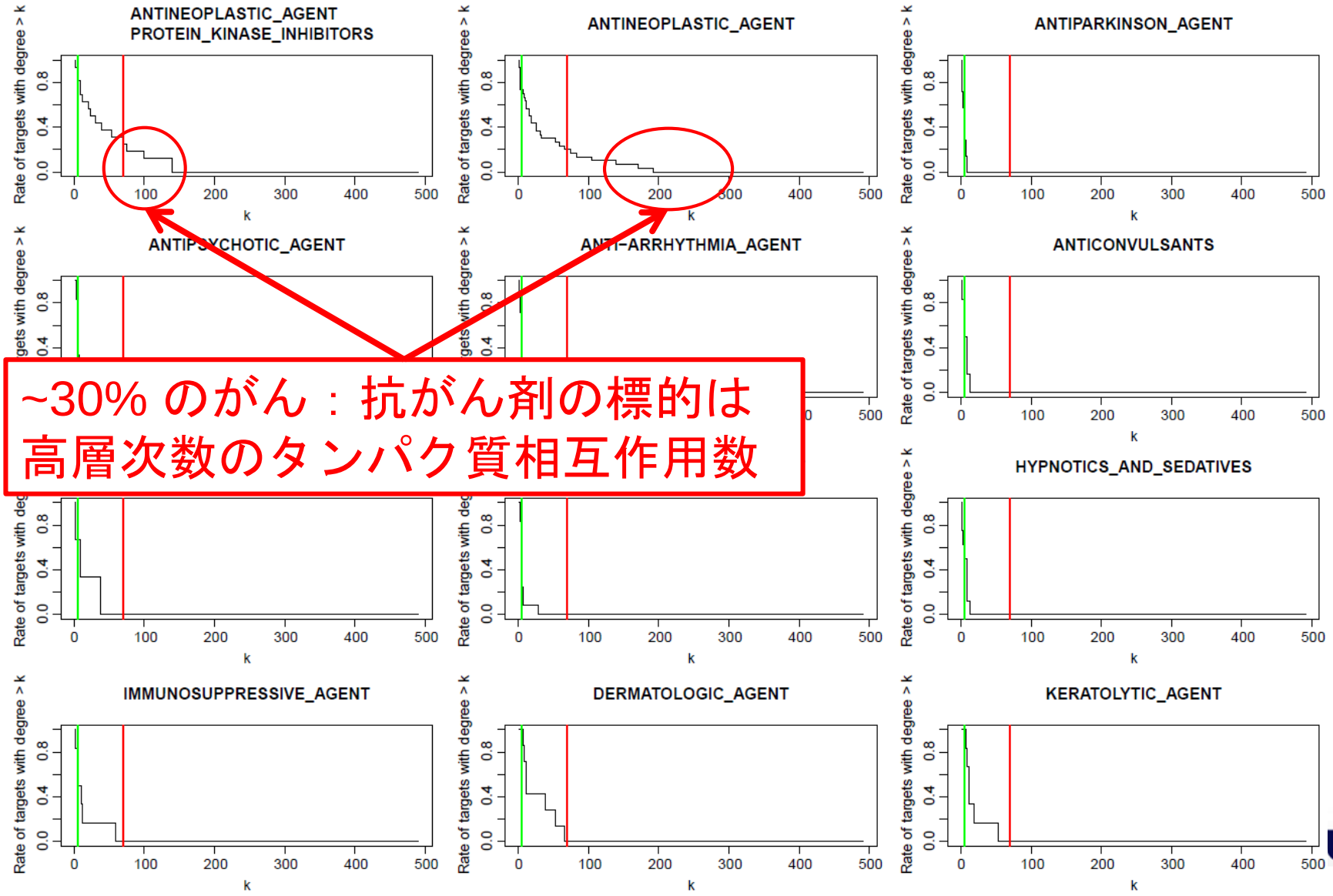
薬剤標的分子と結合度数



中層レベルのノードは治療薬として最適な標的である。それゆえ、多くの市場にある薬剤標的は、ヒトのバックボーンタンパク質に集中している

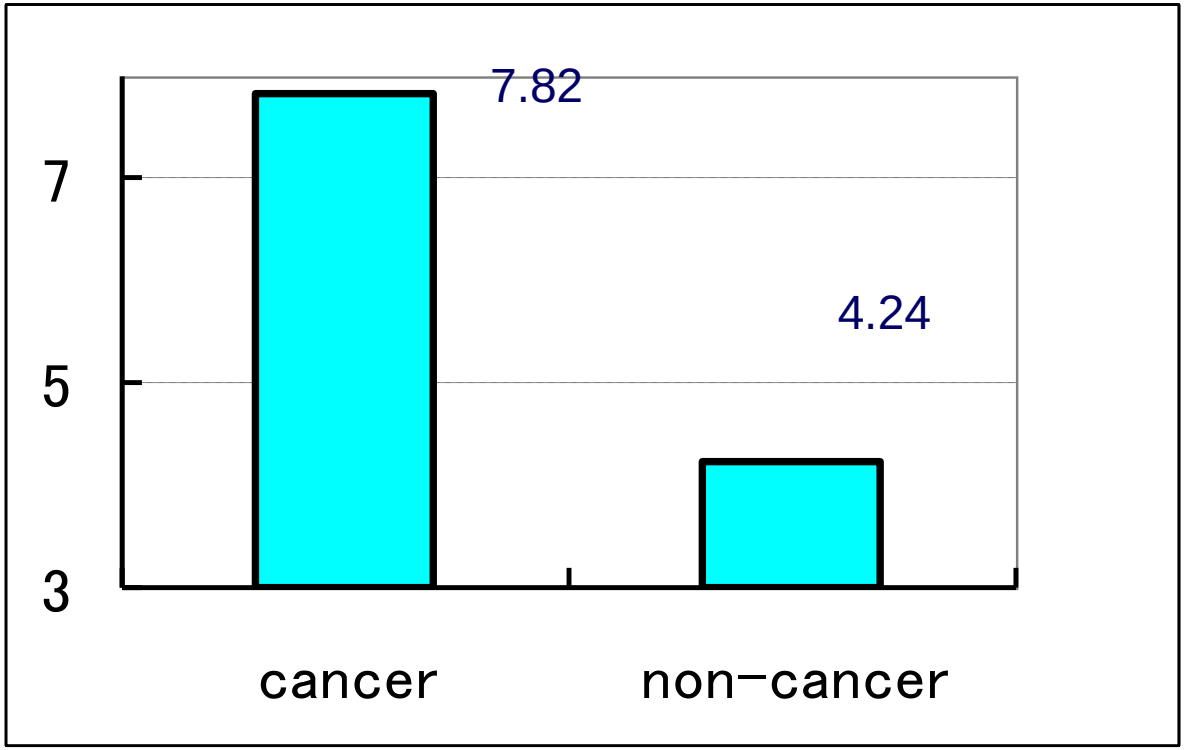
TMDU

がん疾患遺伝子は高層次数ハブのタンパク質が多い



~30% のがん：抗がん剤の標的は
高層次数のタンパク質相互作用数

薬剤標的分子と結合次数



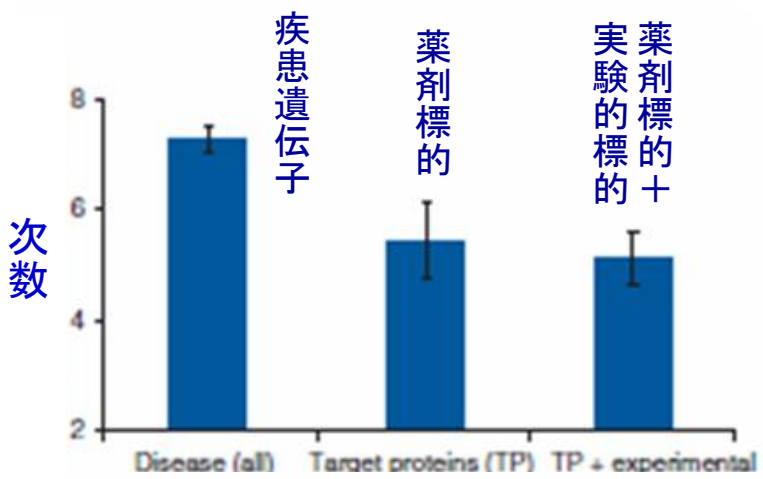
薬剤標的
全平均
4.74

タンパク質
全平均
4.06

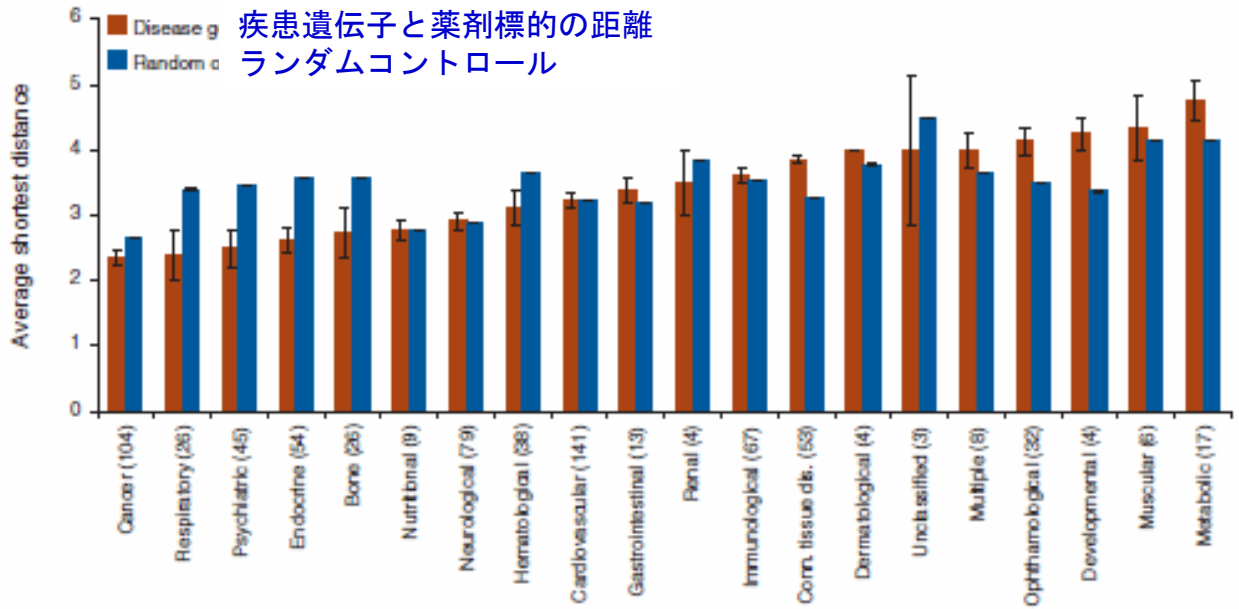
抗がん剤 (P=0.01)の標的分子は平均的に次数が高い。
抗がん剤がより厳しい副作用を起こす理由である.

標的タンパク質と疾患遺伝子の距離

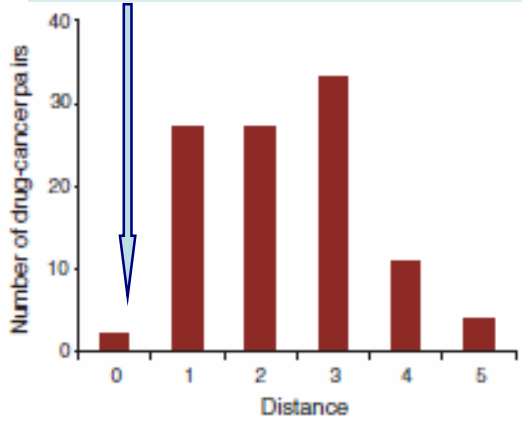
薬剤標的タンパク質と
疾患関連タンパク質の
間の距離：２～４リンク



Yıldırım　M A, et al,
NATURE Biotechnology
2009

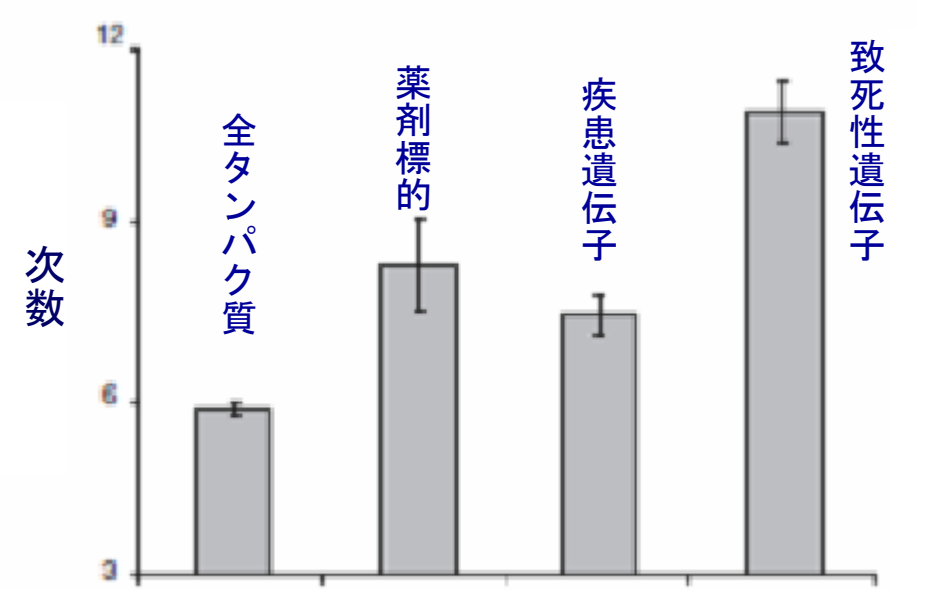


抗がん剤の場合
疾患遺伝子と距離０の標的

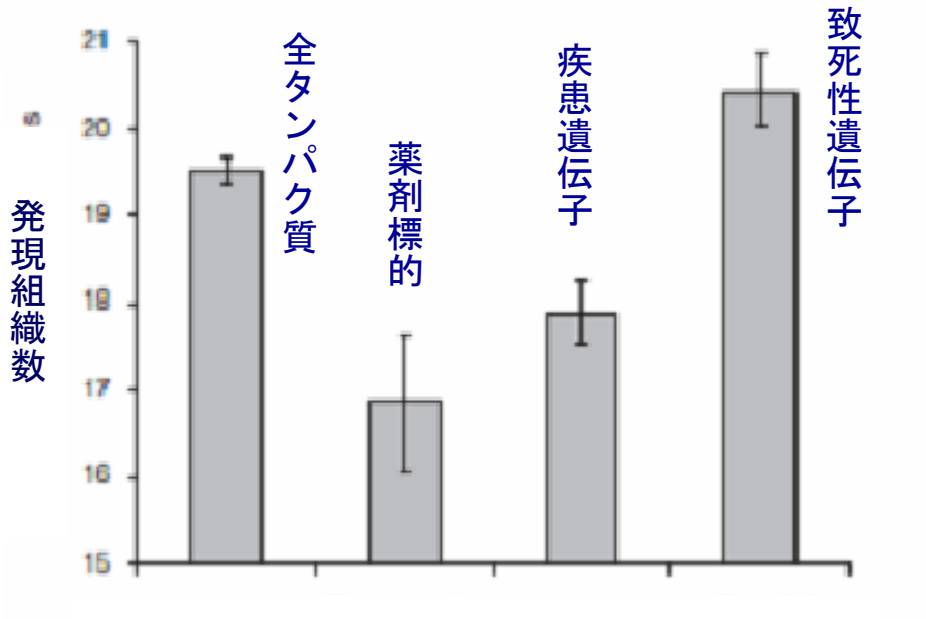


抗がん剤の標的分子と
疾患遺伝子の間に距離

薬剤標的分子と結合次数



結合次数
薬剤標的タンパク質
致死的タンパク質
疾患関連遺伝子タンパク質



発現組織数
薬剤標的タンパク質
致死的タンパク質
疾患関連遺伝子タンパク質

(YıldırımM A, et al, NATURE Biotechnology 2007)

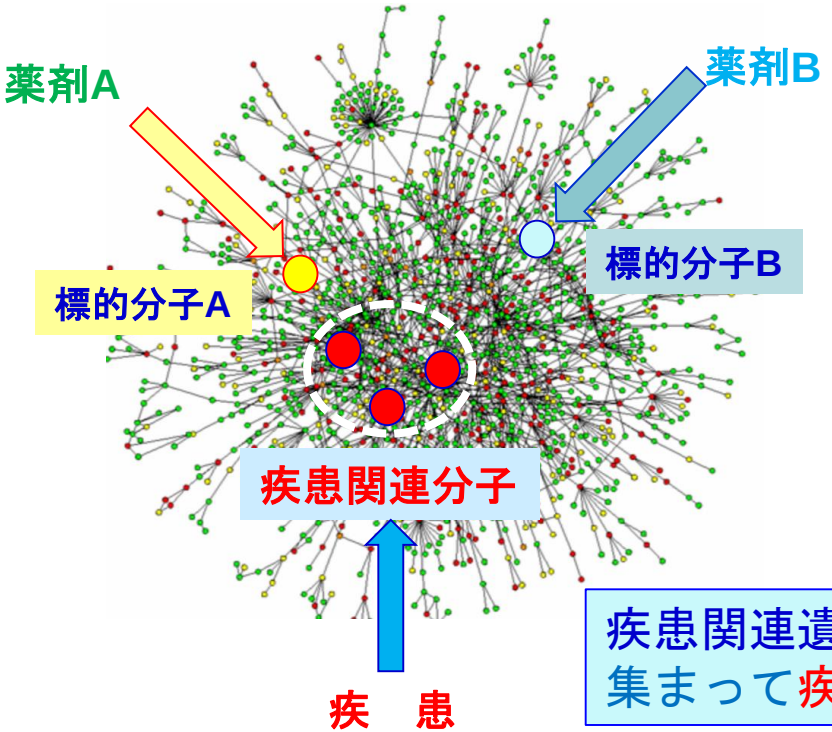
TMDU

タンパク質相互作用ネットワークを基盤にした計算創薬/DR

TMDU

標的分子や疾患関連分子の
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の基盤：生体分子ネットワーク
- 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（近接性）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**を疾患関連分子（**タンパク質**）へ翻訳、
- PPIN場内での**薬剤（標的分子）と疾患（疾患関連遺伝子）の「代理人」**の**距離・近接性**を基準に、薬理作用のインパクト力を評価

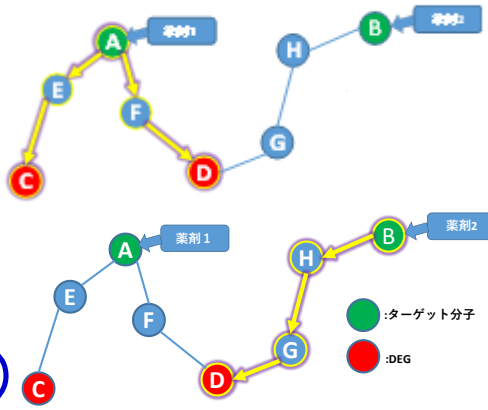


タンパク質相互作用
ネットワーク（PPIN）

疾患関連遺伝子はネットワーク上の近傍に集まって疾患モジュールを形成する

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
  - STRINGS（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

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

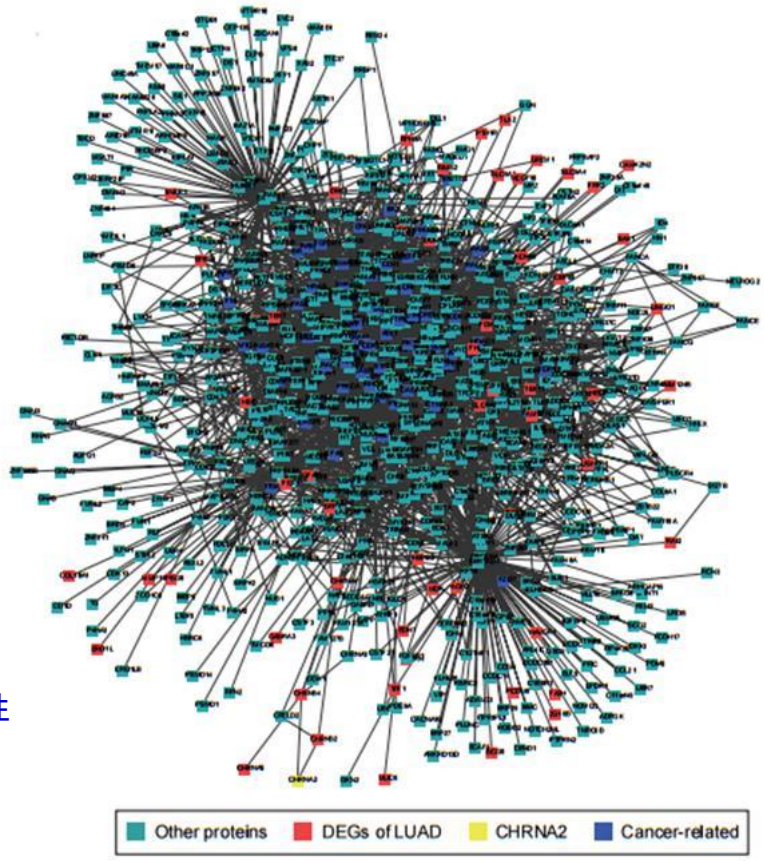
$\mathbf{P}^{s}$：時点sでの各ノードでの滞在確率　$\mathbf{M}$：各ノードへの遷移確率　γ：再出発確率

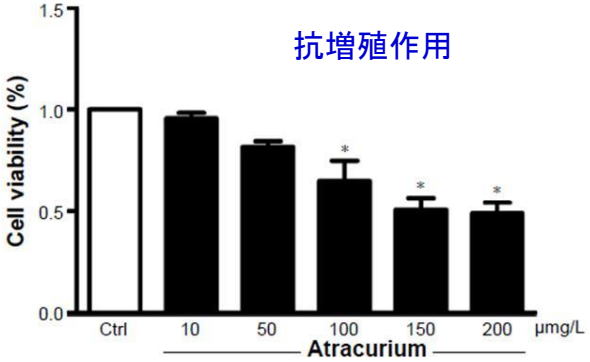
TMDU

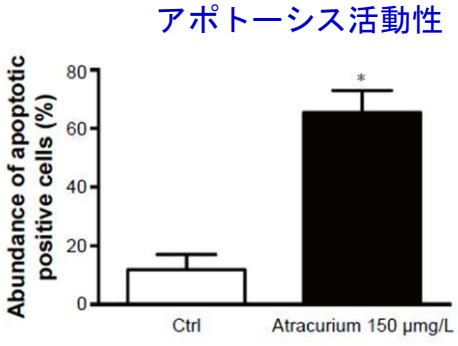
タンパク質相互作用ネットワーク DR 結果の検証

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の両方のランダム歩行で145薬剤・化合物**が共通
- 最高スコアを挙げた**Atractium**を選択
- 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



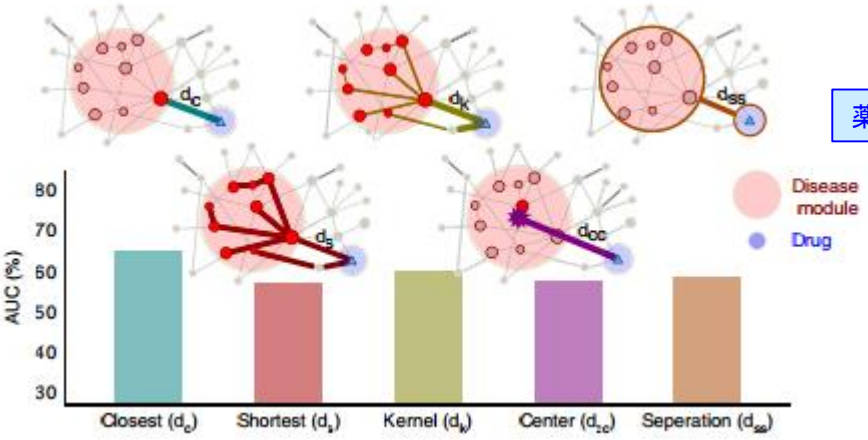




TMDU

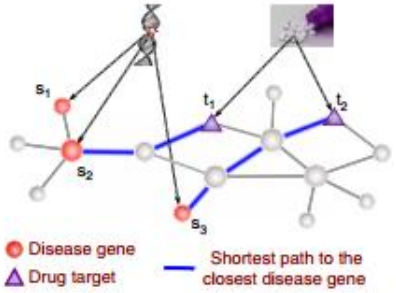
タンパク質相互作用ネットワークでの近接性によるDR

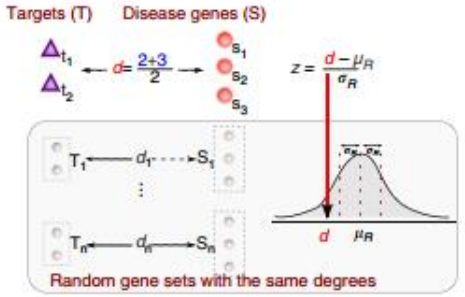
相対近接指標d_c：
①最近接の疾病関連分子との最短経路長の平均
②同じサイズで度数の分布より近接指標を計算して規格化⇒zスコア
（z < - 0.15⇒近接）
②様々な近接指標の中ではclosest measure d_c が一番薬効を予測する

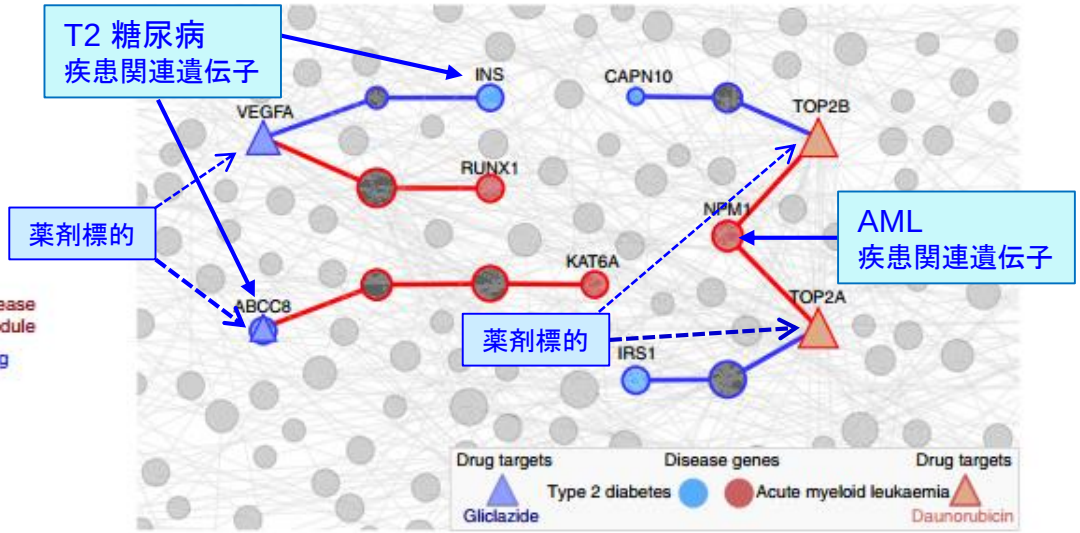


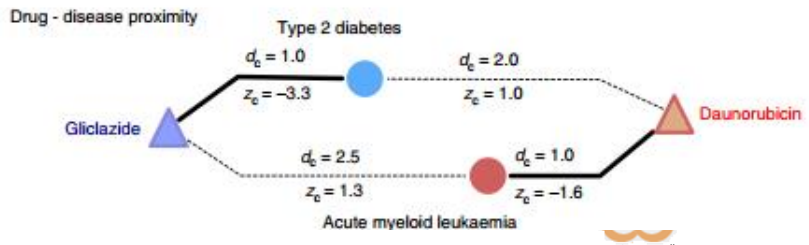
大半の薬剤は標的と疾患関連分子
２リンク離れている

(Guney, Barabasi,2016, Nat. Com)



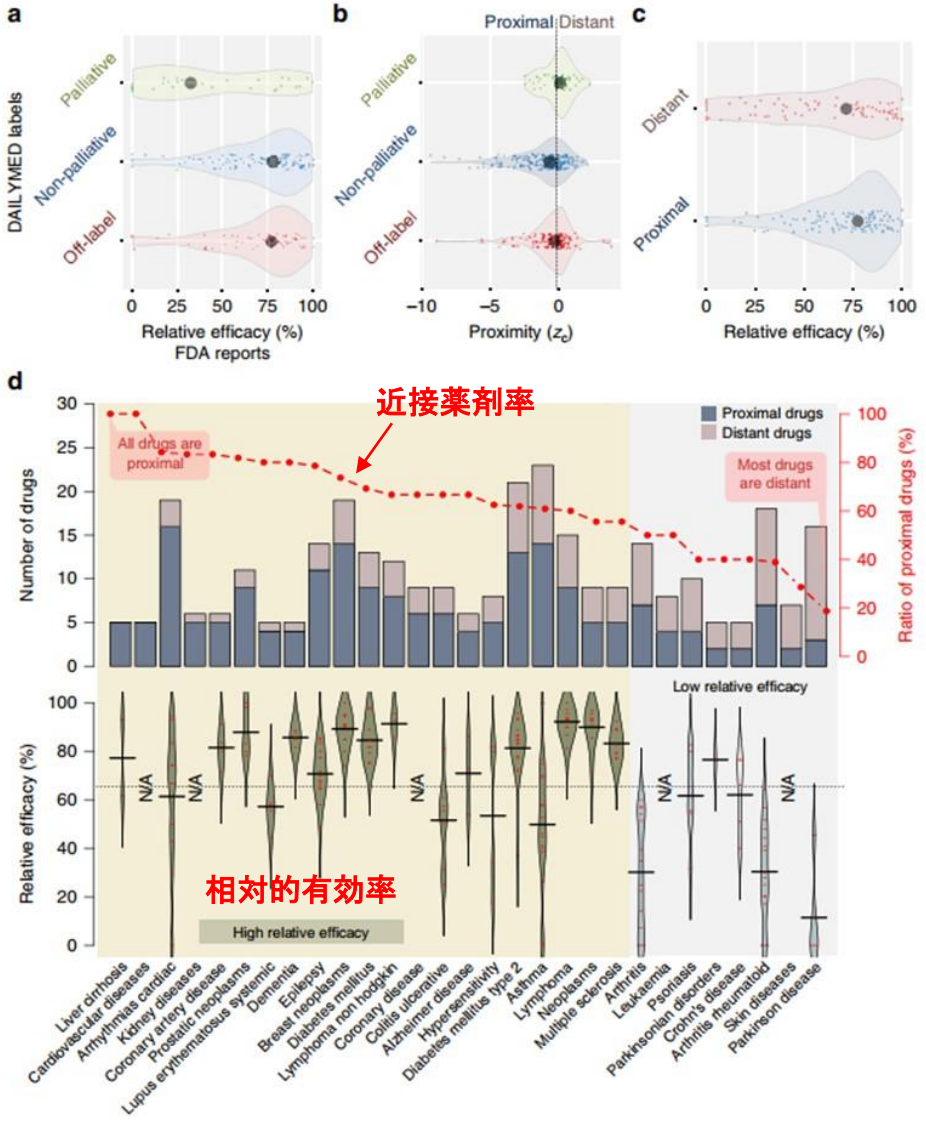






相対近接性による薬効予測

- 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



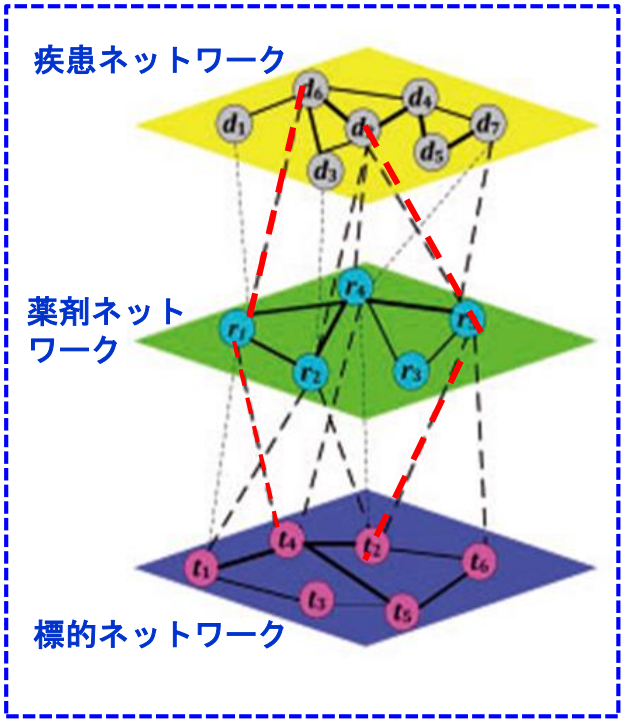
３階層生命ネットワークでの創薬/DR

- 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

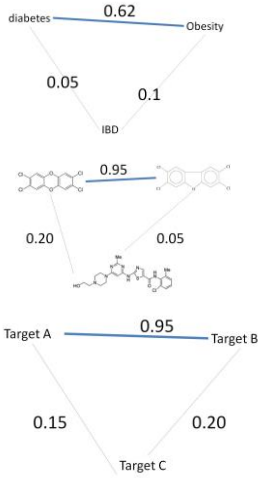
(Wang et al. 2014)

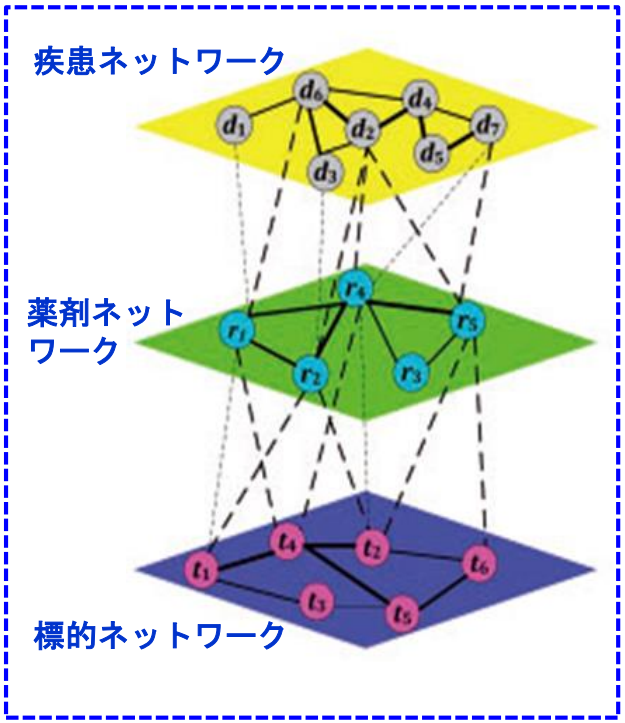


異質 ネットワーク 創薬/DR

(Wang et al. 2014)

- **３層ネットワーク構成**
  - 疾患ネットワーク(d_i)
  - 薬剤ネットワーク(r_i)
  - 標的ネットワーク(t_i)
- **各ネットワークで距離定義**
  - 疾患ネット：MeSHの共通項数
  - 薬剤ネット：Tanimotoスコア
  - 標的ネット：Smith-Waterman法
- **結合係数w(*i*,*j*)更新法**





$$w(d,r)=\sum_{d_i\in D}\sum_{r_j\in R} w(d,d_i)\times w(d_i,r_j)\times w(r,r_j)$$

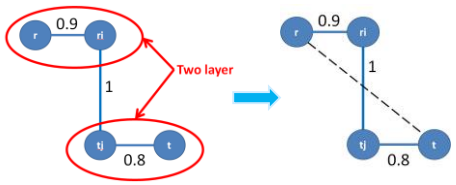
$$w(d,t)=\sum_{r_i\in R}\sum_{r_j\in R} w(d,r_i)\times w(r_i,r_j)\times w(r_j,t)$$

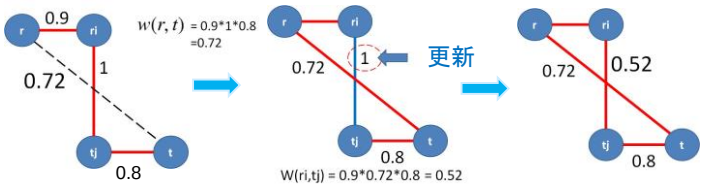
$$w(d,r)=\sum_{t_i\in T}\sum_{t_j\in T} w(d,t_i)\times w(t_i,t_j)\times w(t_j,r)$$

結合係数更新のマトリックス表示

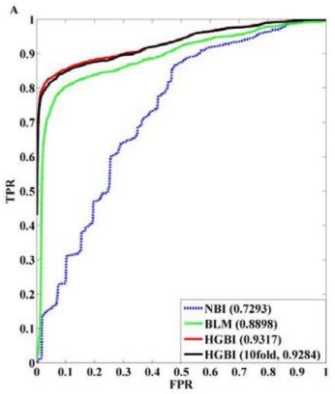
$$W_{dr}^{k+1}=\alpha W_{dr}^{k}\times(W_{rr}\times W_{rt}^{k}\times W_{tt}\times W_{rt}^{kT})+(1-\alpha)W_{dr}^{0}$$

$$W_{rt}^{k+1}=\alpha(W_{dr}^{kT}\times W_{dd}\times W_{dr}^{k}\times W_{rr})\times W_{rt}^{k}+(1-\alpha)W_{rt}^{0}$$





従来の方法より
DR推定精度高い
ROC曲線

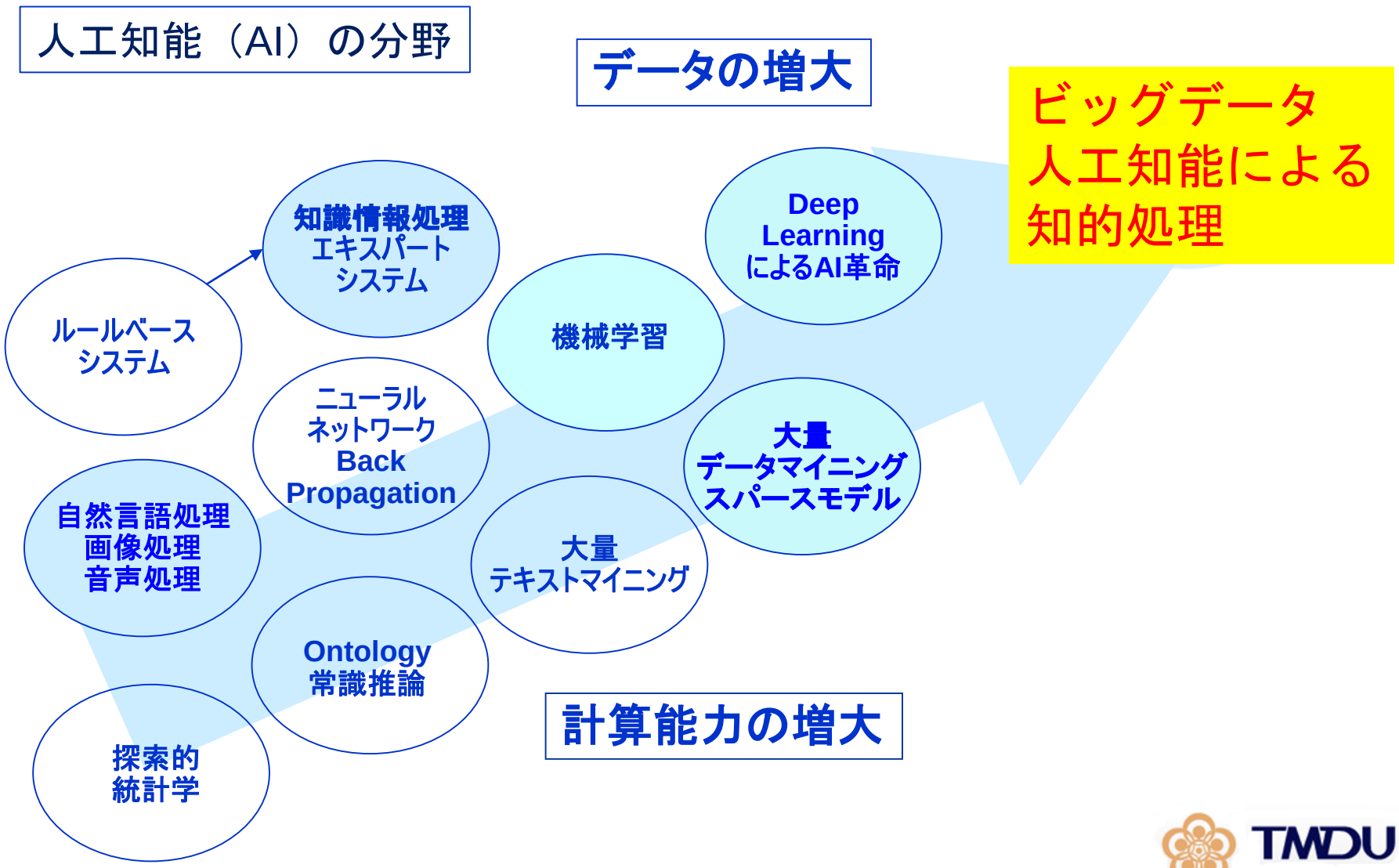


TMDU

AI創薬・DR
（学習型アプローチ）

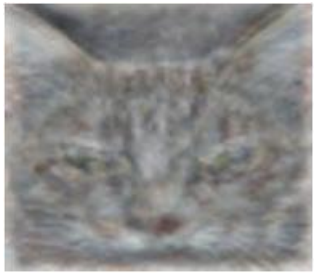
人工知能とDeep Learningの革命性

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
知識情報処理
エキスパート
システム
ルールベース
システム
Deep
Learning
によるAI革命
機械学習
ニューラル
ネットワーク
Back
Propagation
大量
データマイニング
スパースモデル
自然言語処理
画像処理
音声処理
大量
テキストマイニング
Ontology
常識推論
探索的
統計学
計算能力の増大

人工知能の最近の話題

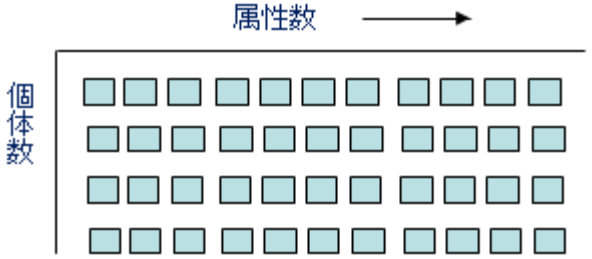
- **「アルファ碁」**（Google DeepMindによるコンピュータ囲碁プログラム）が2016年3月に数多くの世界戦優勝経験のあるプロ棋士李世乭（Lee Sedol：九段）に挑戦し、**4勝1敗と勝ち越した**
  - チェス：IBM「Deep Blue」が1997年に当時の世界champion, カスパロフ氏（ロシア）に勝利
  - 将棋：ボンクラーズ, 2012年米長永世棋聖に勝利
  - **評価経験則が人間が与えていない。強化学習を用いて自分自身と多数の対戦（3000万回）を行う**
- 人工知能が1000万枚の画像を与えて**「猫」を認識するニューロン**をできたと2012年に発表



「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)

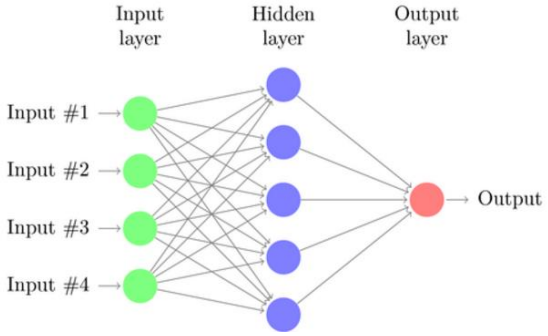
- 医療・創薬ビッグデータへの応用性は高い
- 超多次元ネットワーク情報構造の急増
  - ゲノム医療＜網羅的分子情報ー臨床表現型情報＞
  - ゲノムコホート＜遺伝素因ー環境要因(生活習慣)＞
- Deep Learning-based Network Contraction
  「DLネットワーク縮約法」

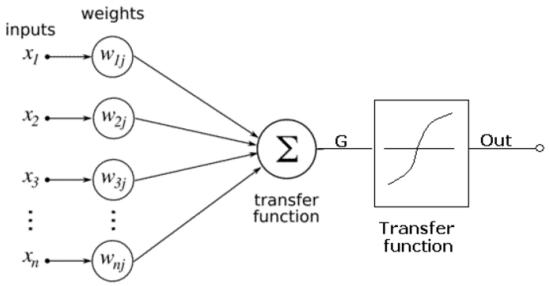
  超多次元ネットワーク情報構造⇒
  少数の特徴的ネットワーク基底に分解
- 線形分解ではない。非線形分解で基底への射影
  - 線形分解（特異値分解：SVD）との比較

TMDU

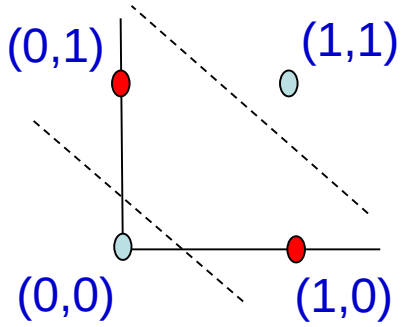
従来のニューロネットワーク

古典的Neural Network・パーセプトロン(1970年代）



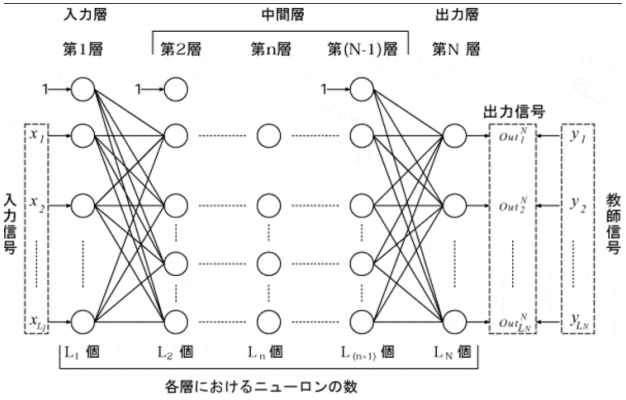


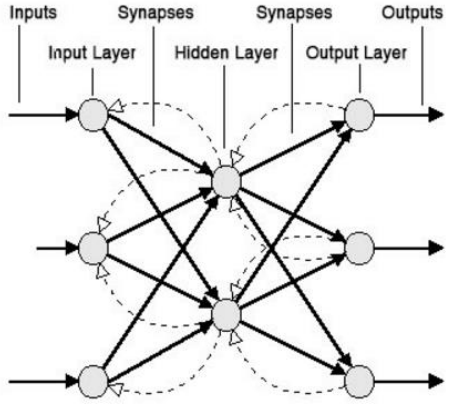
XOR



線形分離できない

多層Neural NetworkとBack projection (1980年代）





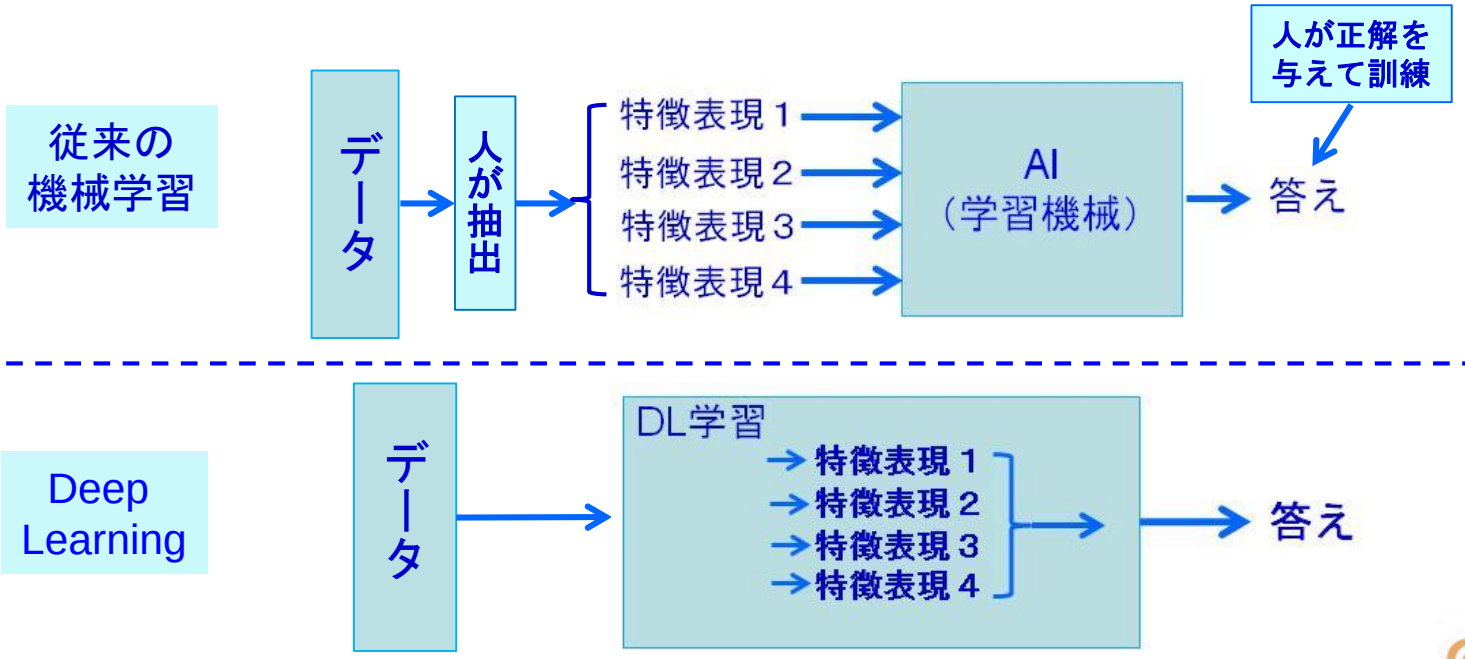
Back Propagation (1986　Rumelhart)
望ましい出力との誤差を教師信号として与える事により、次第に結合係数を変化させ、最終的に正しい出力が得られるようにする。
結合係数を変える事を学習と呼ぶ。
この学習方法には、最急降下法（勾配法）が使われる。出力層へ寄与の高いノードの重みが変更。

多層にわたる逆伝搬で修正感度減衰

TMDU

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

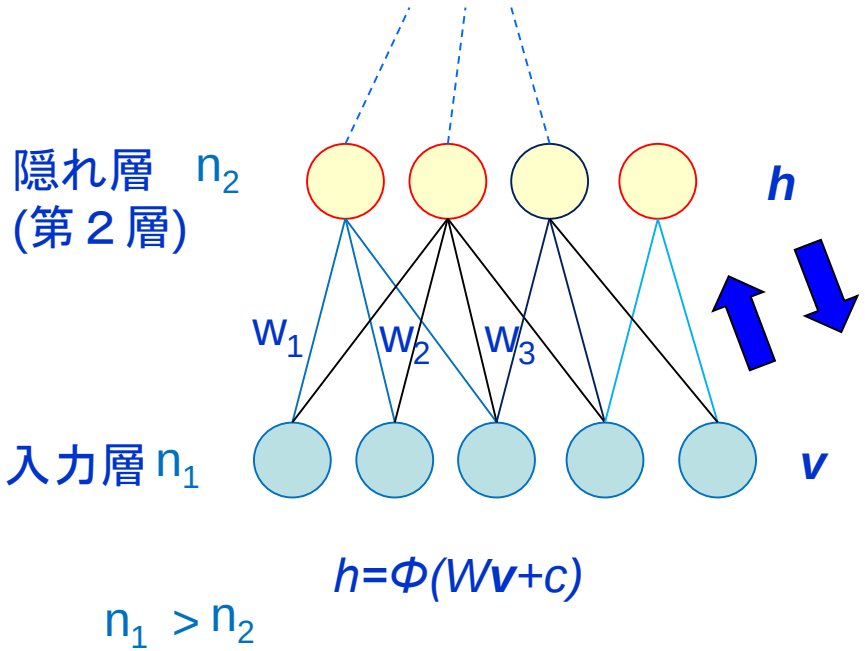


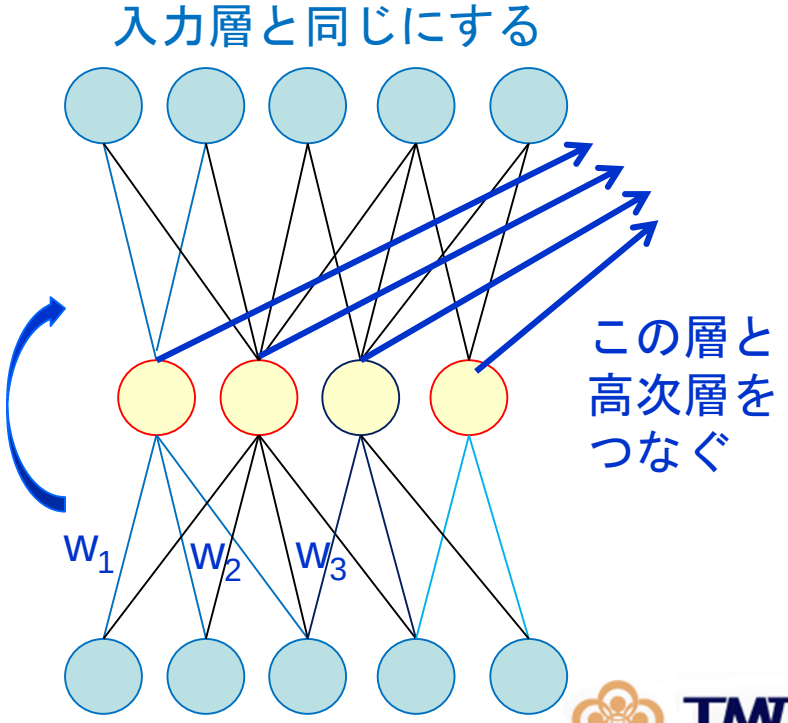
DLの革命点 Autoencode 1

- 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

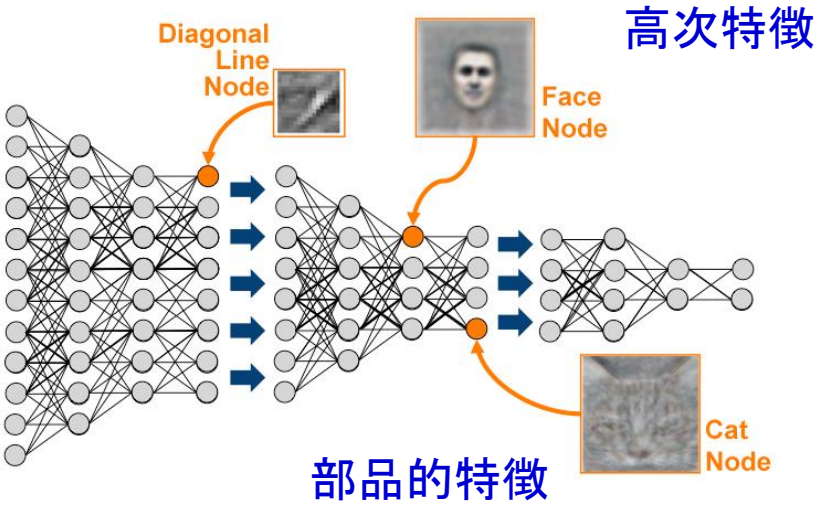
→内在的な特徴量を見出す





DLの革命点 Autoencode２

- 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**

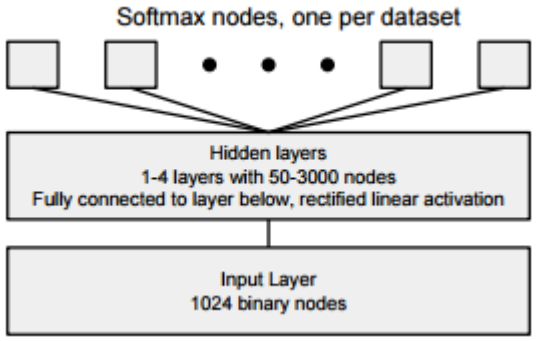


TMDU

Deep learning: 創薬からの注目

- **Kaggle** (データサイエンス競技会)に**Merck社**が**出題**
  Molecular Activity Challenge (2012).
  - 15データセットから構造活性相関（QSAR)のデータを学習
  - 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：Merkの自社モデルより15%予測精度向上
  - **deep learningによるマルチタスクニューラルネット**モデル
- **マルチタスクディープラーニング（MT-DNN)**：
  - 各課題のNNに共通な層（shared layers)を用意
    - 関連性のある他の課題のデータを利用できる
    - データ数の増大により雑音を除去できる
- Google が Stanford 大と共同研究(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"

1つのネットで多タスクを行う



Massively Multitask Networks

「AI創薬」の現在の研究

- **標的分子探索に人工知能を用いた方法**
  - 疾患に有効性のある標的分子の探索
    - Hase－Tanakaの多層Deep AutoEncoderを用いた標的分子探索法
- **Virtual Screening（QSAR)への人工知能・機械学習の応用**
  - **標的分子に相互作用（ヒット）する化合物の探索**
  - Ligand-based AIバーチャルスクリーニング
    - Kaggle, Unterthinerなど
  - Structure-based AIバーチャルスクリーニング
    - Wallachなど
- **その他**
  - 人工知能を用いた化合物自動設計
    - Gomez-Bombarelliなど
  - 合成経路自動探索
  - AI毒性学

TMDU

Artificial Intelligenceと創薬

- 標的分子選択と妥当性検証
  - 疾患に対する適切な分子標的の選択
- Virtual screening と判定
  - 適切な化合物に対する相互作用の有無の判定
  - 研究例：ChEMBLに対するdeep learning
    - 13 M 化合物特徴量（ECFP12), 1.3M 化合物, 5k 薬剤標的
    - Ligand-based 標的予測,7種の予測法とAUC比較
    - Deep learning: SVM, k-nearest nb, logistic回帰より有効
    - DLで構造活性相関を学習する
    - 特徴量の抽出、薬理機序への理解
  - リード最適化
- システム薬理学
  - ネットワーク病態学よりの創薬戦略
  - 他のシステムへの影響(毒性, 副作用)

Pharmacophoreの摘出

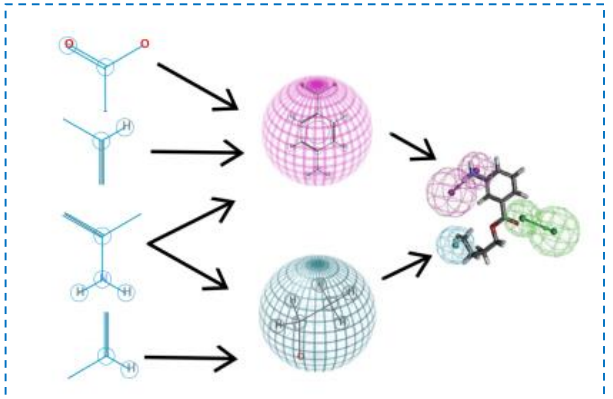


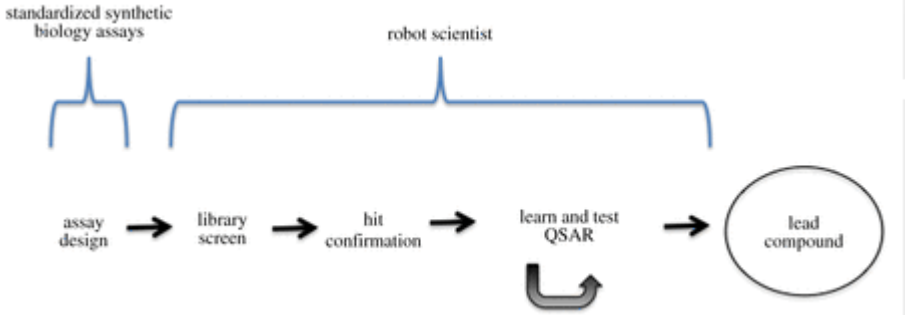
Figure Hierarchical nature of fingerprint features: by combining the ECFP features we can build reactive centers. By pooling specific reactive centers together we obtain a pharmacophore that encodes a specific pharmacological effect.

ECFP(chemical substructure: Enhanced cyan fluorescent)

AI創薬のDL以外の方法

- Berg社のAI創薬
  - AIを方法として膵臓がんの抗がん剤を開発中
  - 膵臓がんと非患者の14兆のゲノム・オミックス情報を比較。
  - 調節不全パスウェイのシステム推定
  - システム薬理学的AIによる創薬
- マンチェスター大学（ Cambridgeとも共同）
  Artificially-intelligent Robot Scientist for new drugs
  - ライブラリースクリーニング, ヒット化合物の確証, リード化合物などの自動化
  - 構造活性相関（Quantitative Structure Activity Relationship）(QSAR) を反復学習する
  - 熱帯病、寄生体のDHFR（ジヒドロ葉酸還元酵素：薬剤耐性）を標的にして学習、細胞を合成生物学操作
    血管新生阻害因子（抗がん剤）をDR候補を探索
  - 最上位にコンセプト木（“root: assay triple screen”など)

Robot scientist Eve at work

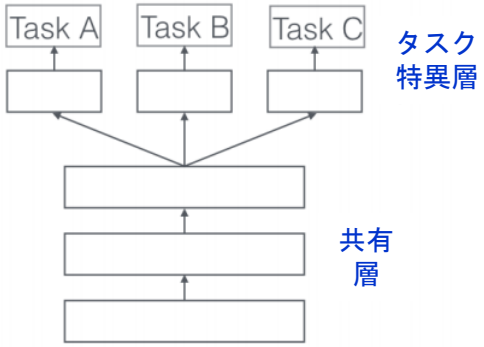


TMDU

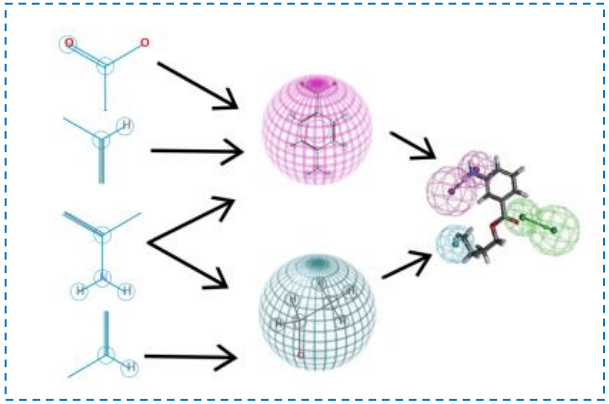
Virtual Screening（QSAR)へのAIの応用

バーチャルスクリーニングへのAIの応用 (ligand-based)

マルチタスクDNN



- Kaggleでの課題：QSARモデル
  - マルチタスク・ディープラーニング
  - 15種類のassayのデータに１つのMT-DNN
  - 各課題に共通な層（shared layers)を用意
    - 関連性のある他の課題のデータを利用できる
    - データ数の増大により雑音を除去できる
    - 個々の課題に単一NNを用意するより高精度（Dahl2014)
- Unterthinerの大規模な構造活性相関（QSAR)研究
  - ChEMBLに対するdeep learning
  - 13 M 化合物特徴量（ECFP12),
  - 1.3M 化合物, 5k 薬剤標的
  - Ligand-based 標的予測,
  - 7種の予測法とAUC比較
  - Deep learningがSVM, k-最近隣法, logistic回帰より有効
  - 特徴量の抽出、薬理機序への理解

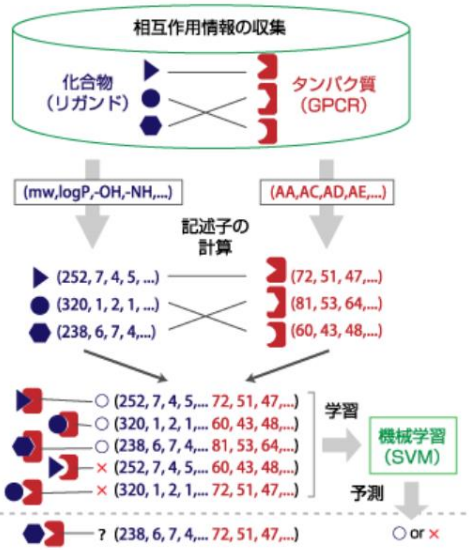


ECFP(chemical substructure: Enhanced cyan fluorescent)

Figure Hierarchical nature of fingerprint features: by combining the ECFP features we can build reactive centers. By pooling specific reactive centers together we obtain a pharmacophore that encodes a specific pharmacological effect.

奥野らのCGBVS法

膨大な化合物候補と多数の標的タンパク質候補との組合せの相互作用評価.
既知のタンパク質（標的）と化合物の相互作用を機械学習,相互作用の有・無を判定



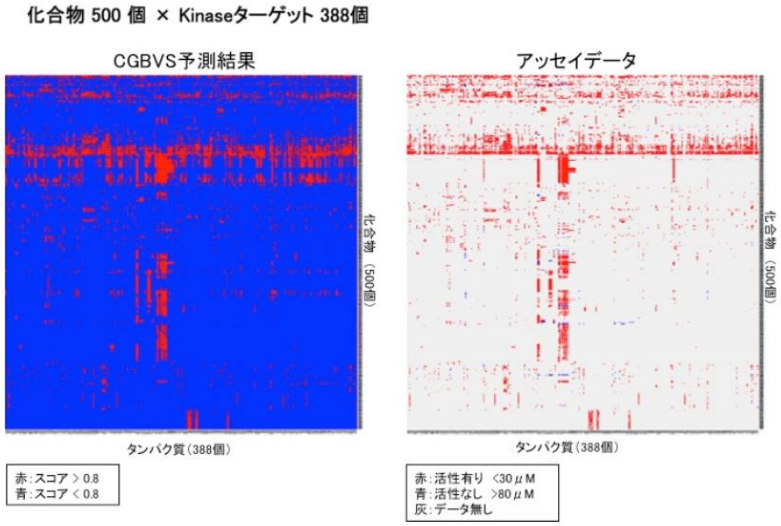
タンパク質と薬剤候補化合物学習の記述子

(1) 標的タンパク質：２アミノ酸や３アミノ酸の出現頻度、構成アミノ酸の特性など

(2) 薬剤候補化合物：分子量、炭素などの構成原子数、部分構造の有無、疎水性度など、化合物の通常2次構造の特徴と物性

(3) 両者の間の相互作用情報

これまでの相互作用の有無が既知である標的タンパク質と化合物の組を選び、相互作用がある場合を「正例」として、相互作用がない場合を「負例」としてする



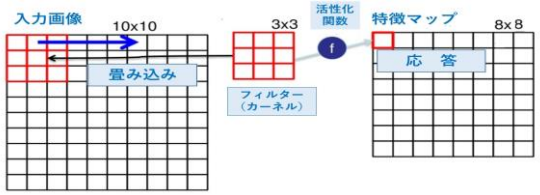
バーチャルスクリーニングへのAIの応用（structure-based)

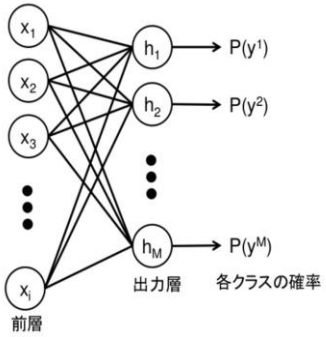
AtomNet (Wallch, 2015)

- トポロジー(近接)情報を学習できるDeep Learning法
  - コンボリューション型Deep Learning (convolutional deep neural network)
- 入力データ
  - 薬剤標的分子：１Å３Dグリッドでの原子座標
  - 標的の結合サイト内の低分子
- ネットワーク構造
  - ３D　コンボリューション層構成
  - ３D　convolutional filter
- 学習法：確率的勾配降下法、逆伝播法
- 結果：活性あるいは非活性クラスの確率を推定
- 既存法と比較評価
  - ドッキング法（Smina, Autodock Vina）を上回る予測精度（AUC）

TMDU

バーチャルスクリーニングへのAIの応用（structure-based)

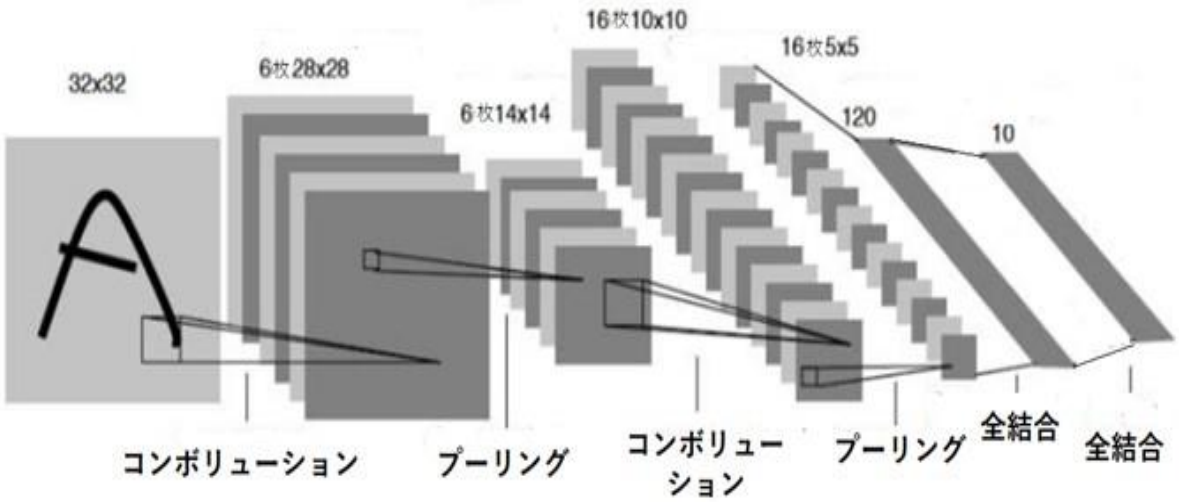




Softmax

$$P(y^i) = \frac{\exp(h_i)}{\sum_{j=1}^{M} \exp(h_j)}$$

各クラスの確率を算出して
最大値を認識クラスとする



コンボリューション型ＤＮＮの全体のアーキテクチュア

TMDU

Deep Learningに基づく
標的分子探索

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

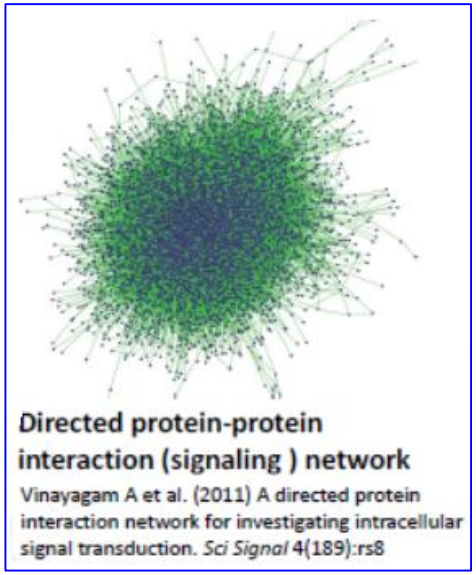
TMDU

特徴的ネットワーク基底への分解

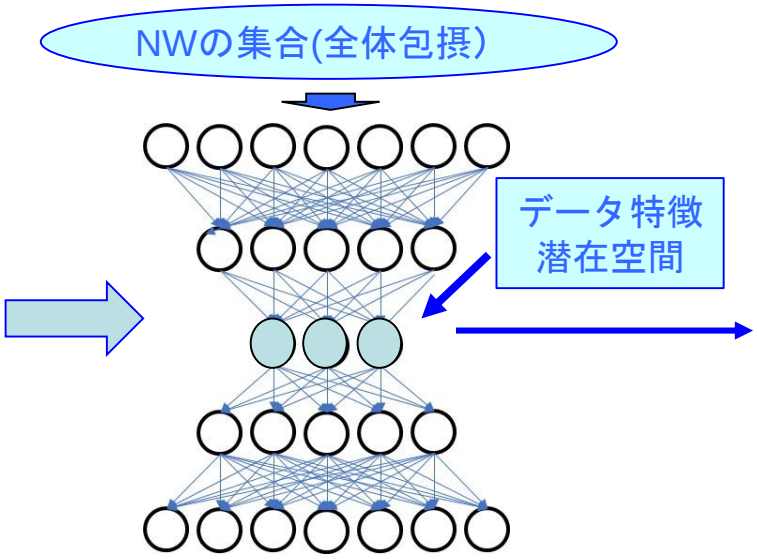
特徴的ネットワーク基底の和に縮約

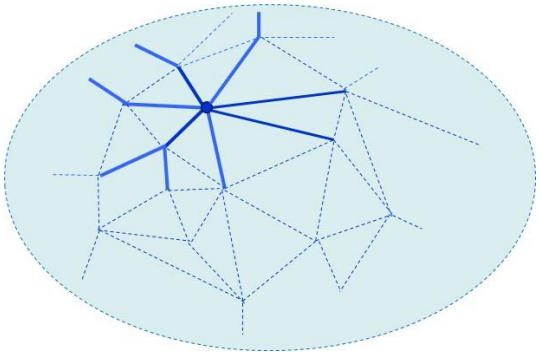
特定のノードを起点とした素NW（部分NW）の集合
全体NWを包摂する集合にDL反復自己学習

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

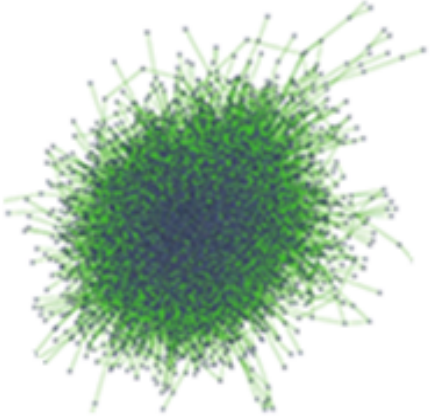
ネットワーク・タンパク質の標的性の判定

TMDU

Our computational workflow

Step1: Input data

Genome-wide PIN

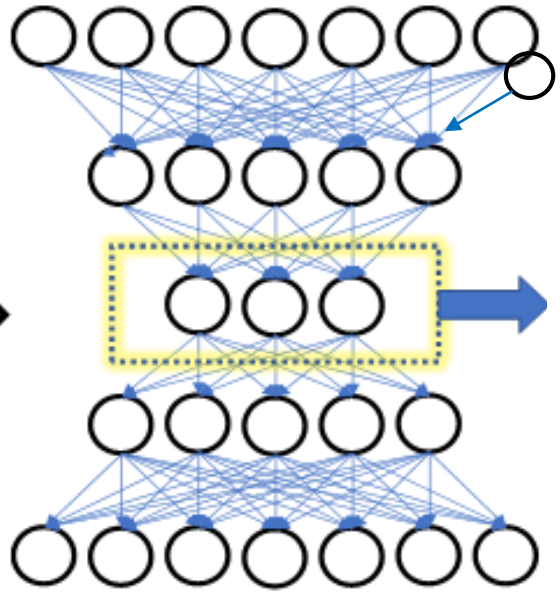


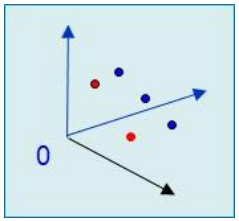
Drugs and their targets information

Step 2: Feature Engineering

Feature engineering by "deep autoencoder" and a state-of-the-art feature selection algorithm

Dimensional reduction by "deep autoencoder"



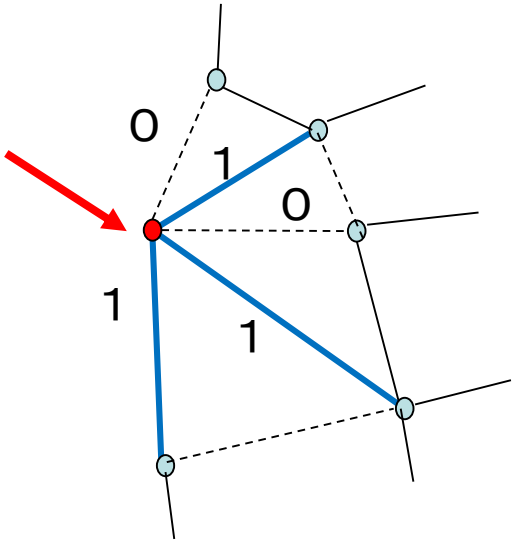


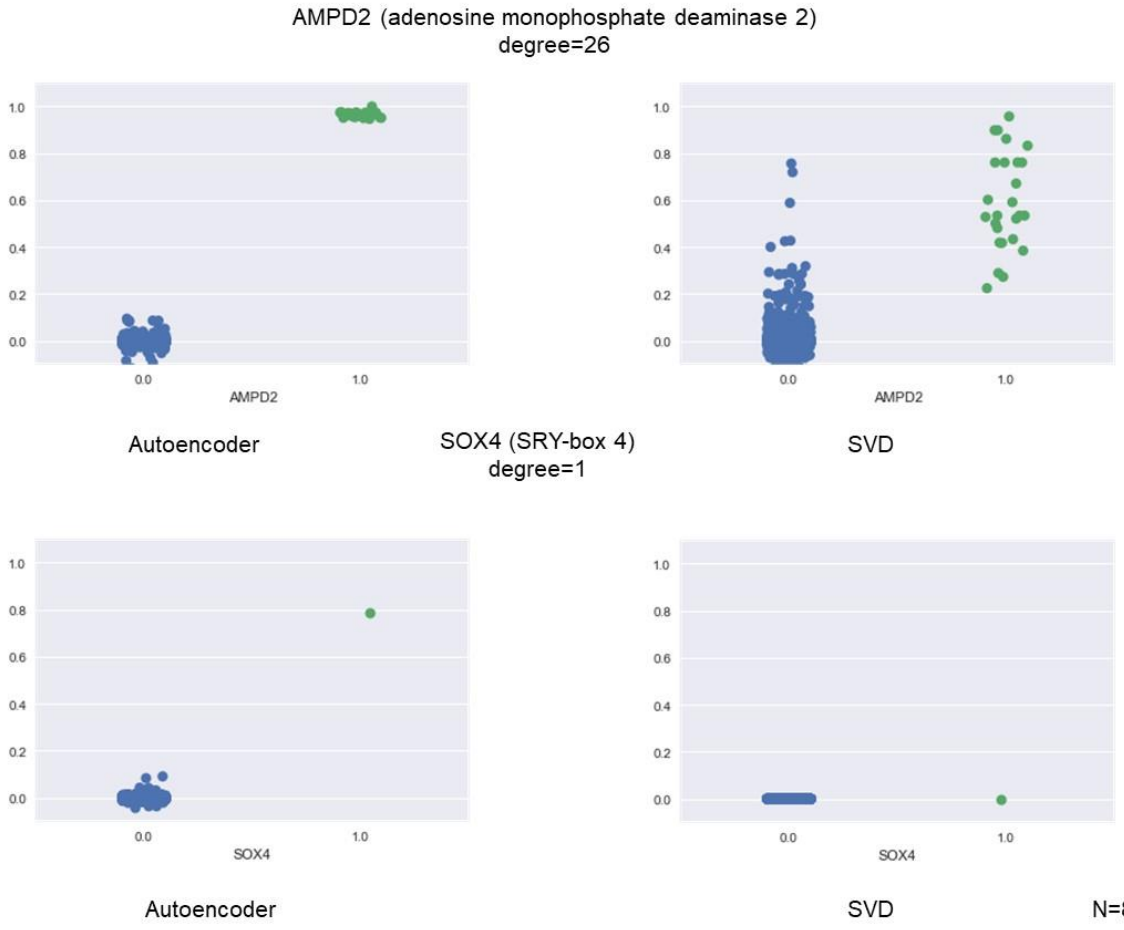
Latent space

Deep Learning と SVD (singular value decomposition)の精度の違い

あるタンパク質ノードに注目する
　周りのノードで
結合しているノードは　1
結合していないノードは0
とすると0, 1の近接ベクトルで結合を表現できる。

v_i= (0,0,0,1,0,1,0,……..)



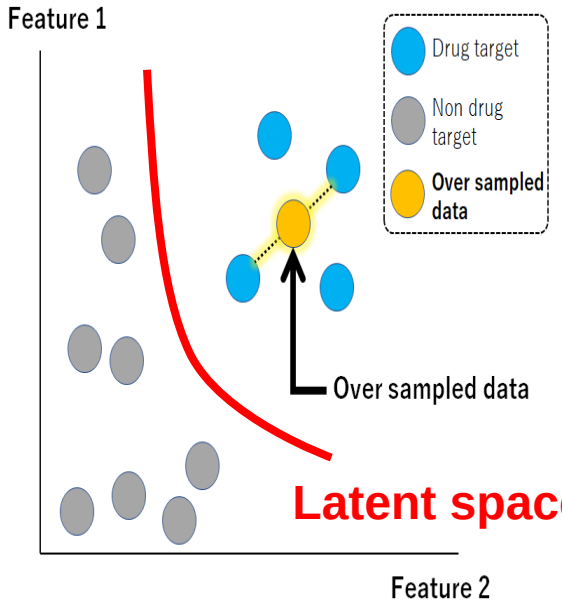


Step 3: Classifier model

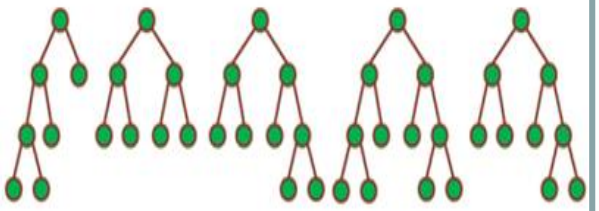
Step 4: Target prioritization

A binary classifier model to target prioritization by state-of-the-art machine learning algorithms

SMOTE algorithm to build a training data



Xgboost algorithm to build a binary classfier



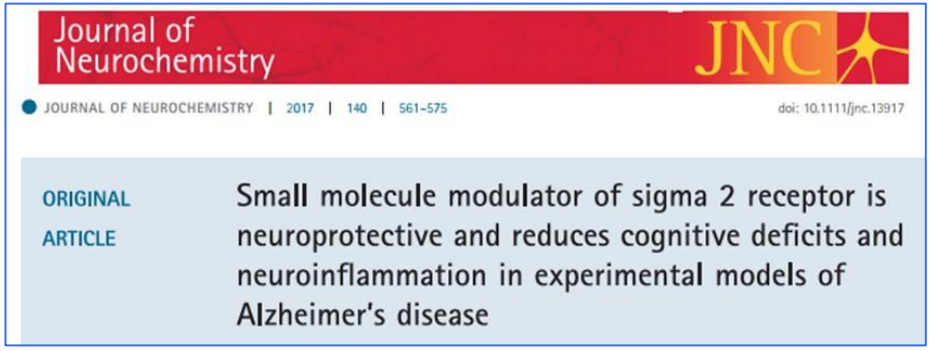
Scores for potential targets

Gene	Score (mean probability)
GRASP	0.982971499
PGRMC1	0.98234516
GPM6A	0.98234516
NRP2	0.975193546
PFKM	0.972127568
DLGAP2	0.953659343
CD81	0.941095327
IQGAP1	0.926867425
TROVE2	0.916886333
TOP3B	0.915745595
TJP1	0.914564961
PDGFB	0.914082375
SETD2	0.905462331
CFLAR	0.900456515
PROS1	0.883435477
SIT1	0.879989294
SIGLEC7	0.879989294
SHC2	0.879989294

TMDU

実験的研究との付合 1

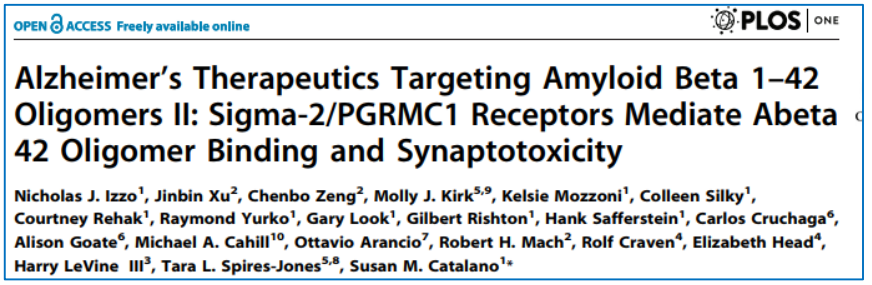
PGCM1：progesterone receptor membrane 1


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

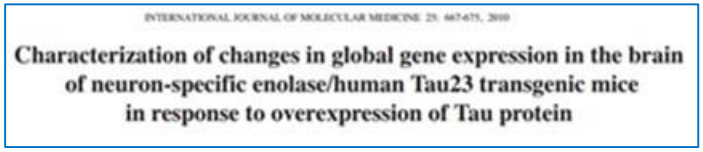

OPEN ACCESS Freely available online PLOS ONE

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]*

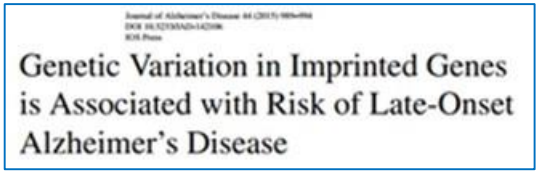
神経保護的効果（neuroprotective)認知不全・炎症に治療効果

GPM6A：Glycoprotein M6A

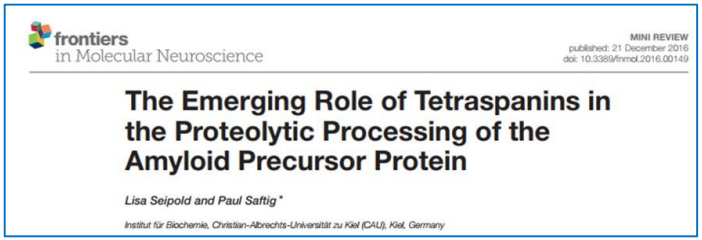

INTERNATIONAL JOURNAL OF MOLECULAR MEDICINE 25: 667-675, 2010

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

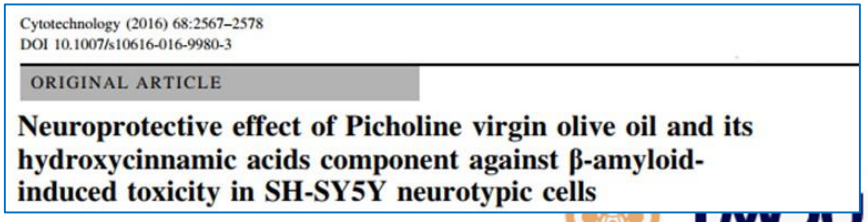
MINI REVIEW published: 21 December 2016 doi: 10.3389/fnmol.2016.00149

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

GRASP	PIK3C2B	PKIA
PGRMC1	NEU3	PFKP
GPM6A	SLC25A38	PAN2
NRP2	TNFSF12	GLUD1
PFKM	ADRA1B	DNM3
DLGAP2	DPM2	ITGA5
CD81	NLRP12	RILPL2
IQGAP1	NLRC4	MAEA
TROVE2	UIMC1	NCDN
TOP3B	IL8	DGCR14
TJP1	VAV1	PACSIN3
PDGFB	ARHGEF1	CD46
SETD2	WISP2	NIT1
CFLAR	PRKCE	ICAM4
PROS1	TBXA2R	GNA13
SIT1	TSPAN4	STK40
SIGLEC7	EPHB4	ROGDI
SHC2	LOC63920	CDH10
SH2D1A	PSEN1	WSB2
	SPOCK3	PHPT1
	TSPO	
	SLC4A1	

アルツハイマー症に対する有効な薬剤標的分子の候補を100以上見出した。

SLC25A38 (APPOPTOSIN)

SLC25A38は アルツハイマー症・脳梗塞患者の脳において増加。 さらに、SLC25A38 の発現低下はBax/BH3IやAβ/glutamateによって誘導されるニューロンの死亡によるアポトーシスを抑制する

Previous Next

Featured Article | Articles, Cellular/Molecular

Appoptosin is a Novel Pro-Apoptotic Protein and Mediates Cell Death in Neurodegeneration

Han Zhang, Yun-wu Zhang, Yaomin Chen, Xiumei Huang, Fangfang Zhou, Weiwei Wang, Bo Xian, Xian Zhang, Eliezer Masliah, Quan Chen, Jing-Dong J. Han, Guojun Bu, John C. Reed, Francesca-Fang Liao, Ye-Guang Chen, and Huaxi Xu

Journal of Neuroscience 31 October 2012, 32 (44) 15565-15576; DOI: https://doi.org/10.1523/JNEUROSCI.3668-12.2012

TMDU

AI準拠DRの例

FDA承認薬剤 **adalimumab** と **etanercept** はDR候補薬剤として期待できる。これらの薬剤はTNF-α （免疫応答を肝要なサイトカイン) の抑制分子で、TNF-α の過剰発現は、特に中枢神経系に炎症を起こす。

MedGenMed
Medscape General Medicine

MedGenMed. 2006; 8(2): 25.
Published online 2006 Apr 26.
PMCID: PMC1785182

TNF-alpha Modulation for Treatment of Alzheimer's Disease: A 6-Month Pilot Study

Edward Tobinick, MD, Assistant Clinical Professor of Medicine, Hyman Gross, MD, Clinical Professor of Neurology, Alan Weinberger, MD, Associate Clinical Professor of Medicine/Rheumatology, and Hart Cohen, MD, FRCPC, Associate Clinical Professor of Medicine/Neurology

CNS Drugs
November 2016, Volume 30, Issue 11, pp 1111–1120

Treatment for Rheumatoid Arthritis and Risk of Alzheimer's Disease: A Nested Case-Control Analysis

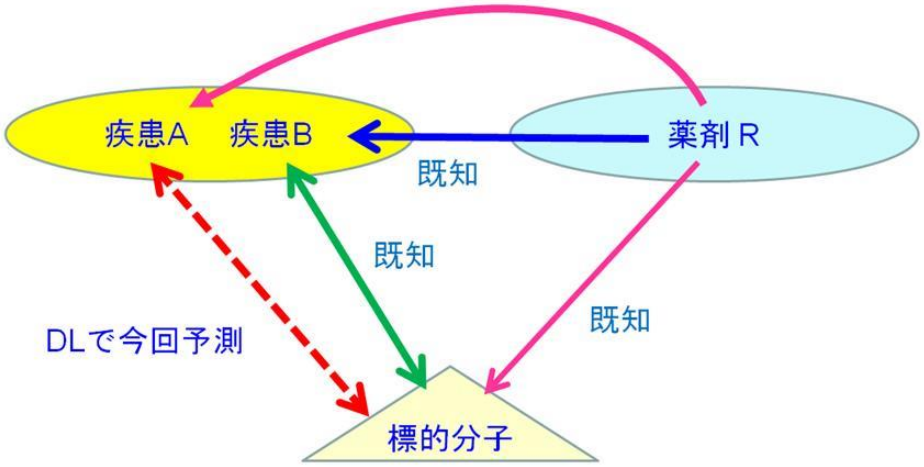
Authors　Authors and affiliations

Richard C. Chou, Michael Kane, Sanjay Ghimire, Shiva Gautam, Jiang Gui

TMDU

アルツハイマー症のDR薬剤候補

DR候補薬剤：
アルツハイマー症の候補標的分子が、ある既承認薬剤Rの標的分子と同一であれば薬剤Rは、アルツハイマー症にも有効と期待。

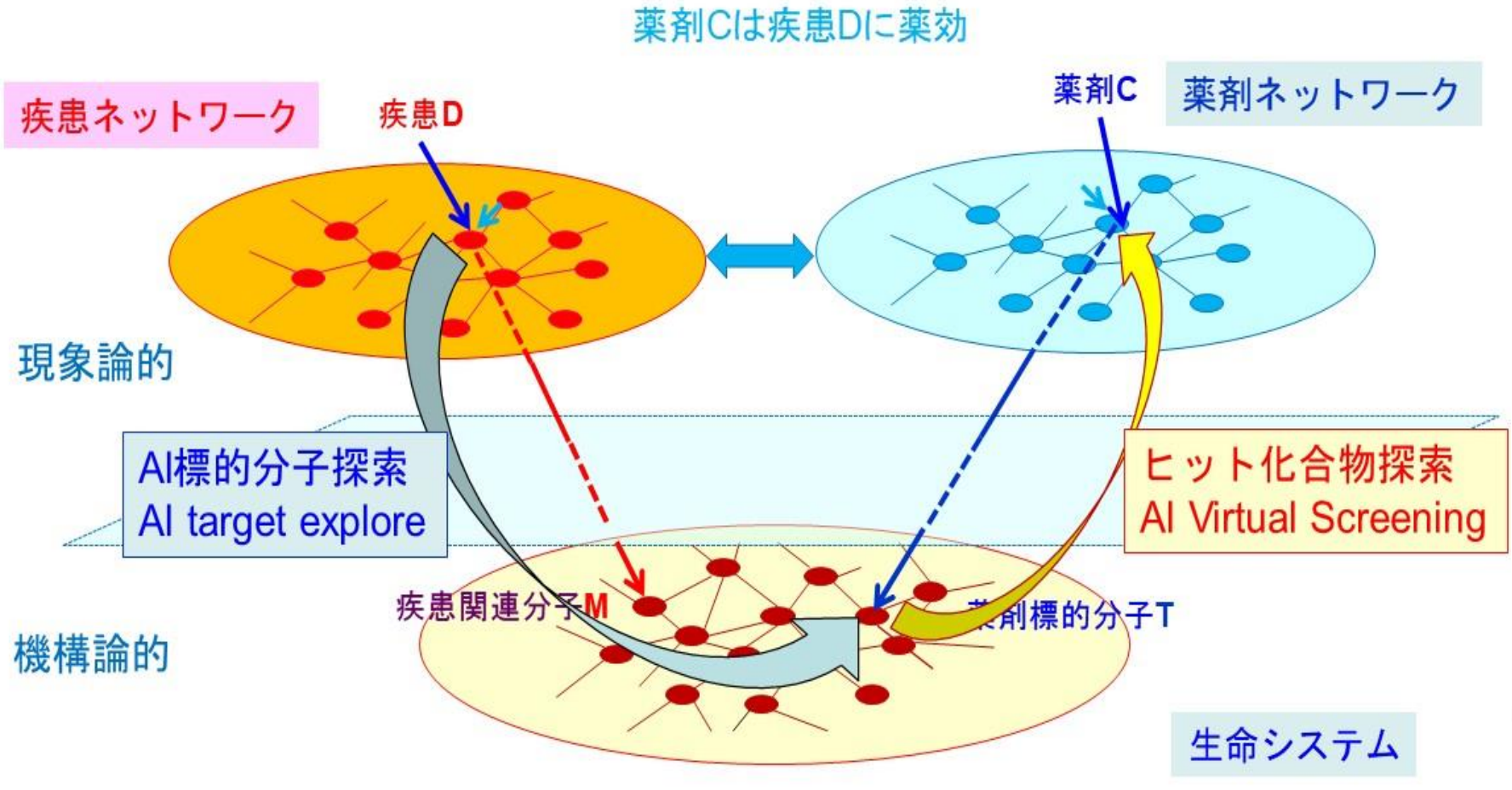


repositonable drug	taregt	# of target	category
Tamoxifen	PRKCB PRKCE PRKCG ESRRG	4	Anti-Estrogens; Antineoplastic Agents; Antineoplasti
Mianserin	SLC6A4 DRD3 OPRK1 ADRA1B	4	Adrenergic Agents; Adrenergic alpha-Antagonists; A
Amitriptyline	SLC6A4 OPRK1 ADRA1B OPRM1	4	
Dextromethorphan	SLC6A4 PGRMC1 OPRM1 OPRK1	4	Alkaloids; Antitussive Agents; Central Nervous Syste
Mirtazapine	OPRK1 ADRA1B DRD3 SLC6A4	4	Adrenergic Agents; Adrenergic alpha-Antagonists; A
Tramadol	OPRM1 OPRK1 SLC6A4	3	Alcohols; Amines; Analgesics; Analgesics, Opioid; C
Zinc	MPG SERPINA1 SERPIND1	3	Acetates; Acetic Acid; Acids; Acids, Acyclic; Acids, N
Amoxapine	SLC6A4 DRD3 ADRA1B	3	Adrenergic Agents; Adrenergic Uptake Inhibitors; Al
Etorphine	OPRM1 OPRK1 OPRL1	3	Alkaloids; Analgesics; Analgesics, Opioid; Central N
Tapentadol	OPRM1 OPRK1 SLC6A4	3	Analgesics; Analgesics, Opioid; Benzene Derivatives
Loxapine	ADRA1B DRD3 SLC6A4	3	Antipsychotic Agents; Antipsychotic Agents (First Ge
Pethidine	OPRK1 OPRM1 SLC6A4	3	Acids, Heterocyclic; Adjuvants; Adjuvants, Anesthesi
Talampanel	GRIA1	1	Benzazepines; Heterocyclic Compounds; Heterocycli
Etanercept	FCGR3B	1	Amino Acids, Peptides, and Proteins; Analgesics; A
Vitamin E	PRKCB	1	Antioxidants; Benzopyrans; Chemical Actions and Us
N-[(2R)-2-benzyl-4-(hydroxyamino)-4-	LTA4H	1	
Adalimumab	FCGR3B	1	Amino Acids, Peptides, and Proteins; Anti-Inflamm
ALPHA-HYDROXYFARNESYLPHOSPH	FNTB	1	Alcohols; Fatty Alcohols; Hydrocarbons; Lipids; Orga

TMDU

AI創薬の実現

3層生体・薬剤ネットワークによるAI創薬の過程



今後の戦略・方向

- ビッグデータ医療・創薬時代：次元縮約
- 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

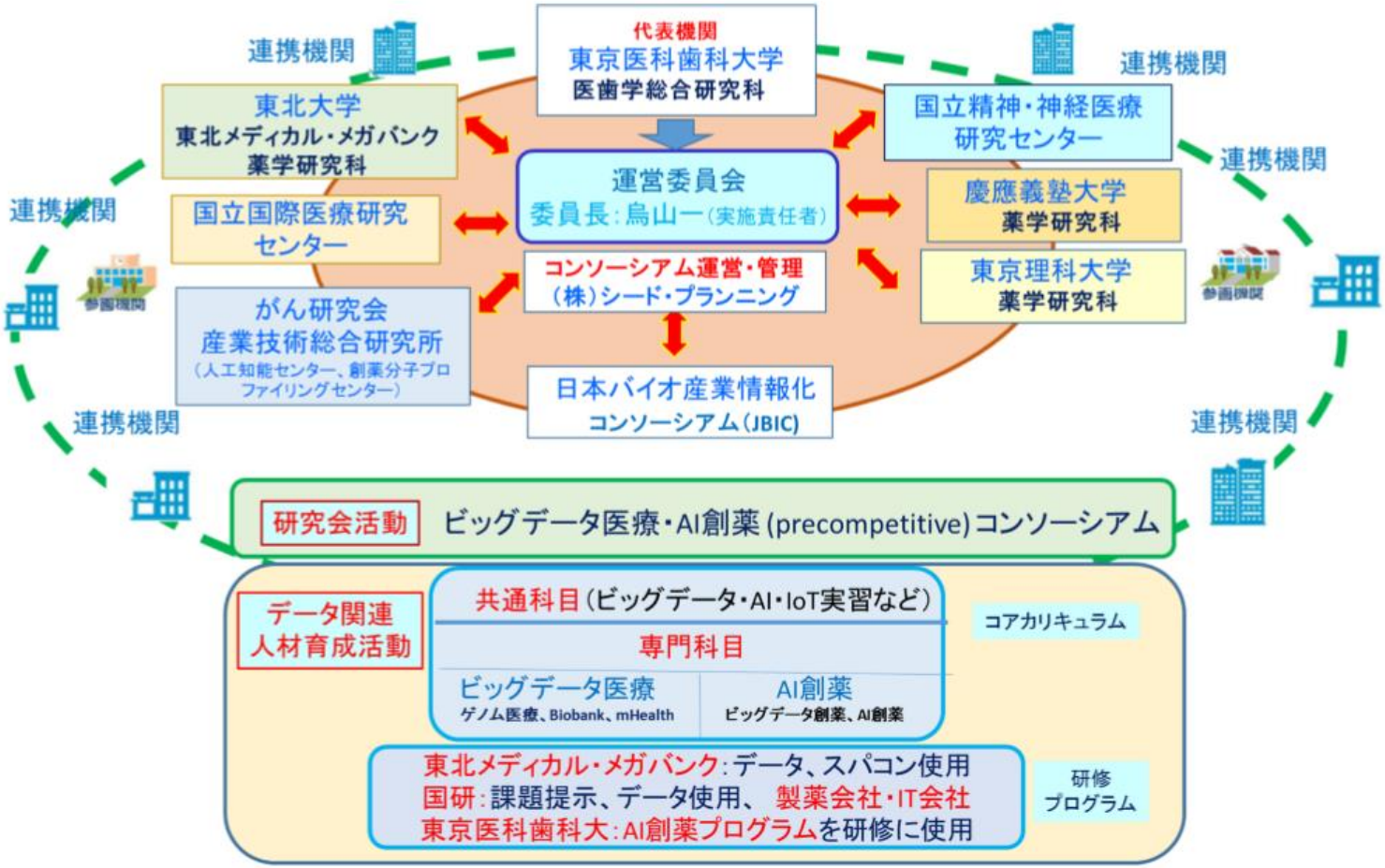
TMDU

田中　博　著

「AI創薬・ビッグデータ創薬」

薬事日報社　6月23日刊行

ビッグデータ医療・ＡＩ創薬コンソーシアム



2018.2.23-25, Harvard/MIT/TMDU - Datathon

TMDU

ご清聴有難うございます