

医療ビッグデータを利用した 新しい製薬産業の姿

東京医科歯科大学医学部 臨床腫瘍学分野 生命情報学
東北大学 東北メディカル・メガバンク機構
田中 博



医学・ヘルスケア分野への ビッグデータ時代の到来

医療ビッグデータ時代

- (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・個別化予防
医療/ヘルスケアの適確性の飛躍的な増大

医療の「ビッグデータ革命」

～何が新しいのか～

1) 臨床診療情報

- 従来型の医療情報 電子化による蓄積
 - 臨床検査、医用画像、処方、レセプトなど

2) 社会医学情報

- 従来型の社会医学情報 電子化による蓄積
 - 疫学情報・集団単位での疾患罹患情報

3) 新しい種類の医療ビッグデータ

- 網羅的分子情報・個別化医療
 - **ゲノム・オミックス医療の網羅的分子情報**
 - **Biobank,ゲノムコホートによる分子/環境情報**
- 生涯型モバイル健康管理 (mHealth)
 - ウェアラブル・生体センシング

旧来タイプの
医療データの
大容量化

新しいタイプの
医療ビッグデータ

医療の「新しいビッグデータの革命性」

～ゲノム・オミックスデータの基軸的な特徴～

＜目的もデータ特性も従来型と違う＞

従来の医療情報の「ビッグデータ」

Big “Small Data” ($n \gg p$)

医療情報・疫学調査では属性数：数十項目程度

— 目的：Population MedicineのBig Data

⇒個別を集めて「集合的法則」を見る

網羅的分子情報などのビッグデータ

Small “Big Data” ($p \gg n$)

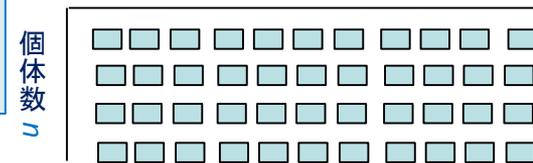
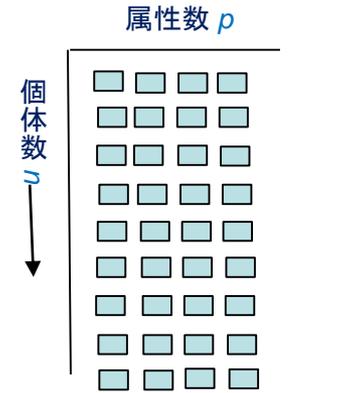
1 個体に関するデータ属性種類数が膨大

属性(p)に比べて個体数(n)少数:従来の統計学が無効

「新 np 問題」：GWASは単変量解析の羅列

— 目的：例えば医療の場合 個別化医療 Personalized Medicine

⇒大量データを集めて「個別化パターン」の多様性を抽出



新しいデータ科学の必要性

医療の「ビッグデータ」革命は どんな既存のパラダイムに挑戦しているか

- Population medicineのパラダイム転換
 - <One size fits for all>のPopulation医療はもはや成り立たない
 - 個別化医療 “Personalized (Precision) medicine”
 - 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
 - 臨床研究を科学にする従来の範型RCTは、個別化概念に破綻した
 - <statistical evidence based>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・ビッグデータからの知識生成（BD2K）
- 創薬の戦略パラダイムの転換
 - <ビッグデータ創薬>へ:分子特性準拠からビッグデータ準拠へ
 - 網羅的分子データ/ネットワーク情報からの計算論的創薬・DR
 - Disease, Drug, Targetの各ネットワーク間の相互写像関係

ビッグデータ医学・医療

次世代シーケンサのインパクト

次世代シーケンサを始めとするhigh-throughput分子情報収集の急激な発展

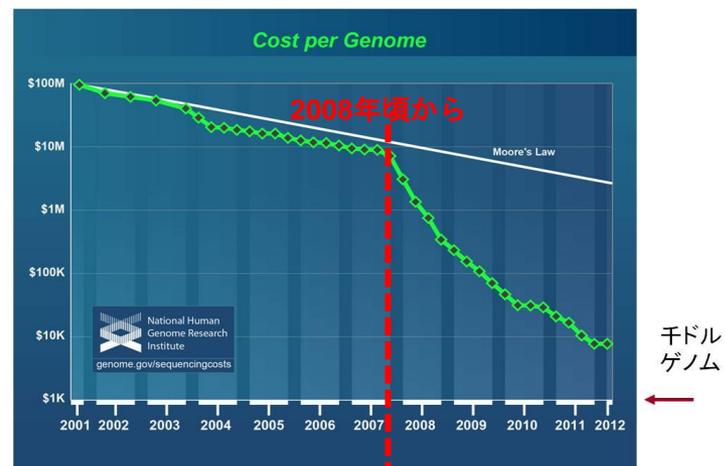
急速な高速化と廉価化 ヒトゲノム解読計画13年,3500億円⇒1日,10万円

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 |
| リード長 (bp) | 2 x 100 | 2 x 150 |
| データ産出量 (Gb) | 約600 | 約120 |
| 試薬コスト (ヒト1人全ゲノム) | 数十万円 | 不可 エクソームのみ |

HiSeq X システム 10台構成 (経費1/5)



DNA Sequencing Cost: the National Human Genome Research Institute

シーケンス革命 2007/8

ゲノム(配列決定)機器の進歩は、計算機のムーアの法則を越えている!

ゲノム医療の第1の流れ

ゲノム医療の臨床実装

Clinical Sequencing

Nic Volker



- Wisconsin 小児病院（全米4位）2009年、3才の男子。
- 2歳から原因不明の腸疾患で、腸のいたるところに潰瘍が発生。
- クローン病かと疑うが、クローン病の既報の遺伝子変異なし
- 2年間で130回の外科的切除手術を行うが再発を繰り返す。これ以上行う治療がなくなった。
- Nicの全エキソンの配列を次世代シーケンサ決定
- WESで見出された16000個のDNA配列異常を慎重に分析



XIAP (X連鎖アポトーシス阻害タンパク質遺伝
変異 TGT(cysteine)→TAT(tyrosine) (203番目)

アポトーシスの阻害因子 免疫系が腸を攻撃する自己免疫
を阻害 これまでのヒトゲノム配列で見出されていない
ショウジョウバエからチンパンジー見いだせず

臍帯血移植（造血幹細胞移植）を実施（2010年6月）
2010年7月半ば（42日後）には、食事が取れるまでに回復した。
現在は普通の男子と変わらぬ健康な生活を送っている。
2010年の12月に3回連載で全米に記事・記者にピューリツア賞
全米のゲノム医療に関する様相が一変。ゲノム医療が始まった



Medical College of
Wisconsin, Human &
Molecular Genetics
Center
Howard Jacob
(a major mover of
the whole field, Topol)



Wisconsin医科大学小児病院および Froedtert 病院のゲノム医療

- Wisconsin医科大学 Genome sequencing program

- Nic君に続いて（翌年3月まで6例）

- 候補選択（nomination）

- 従来の検査・診察で診断困難な症例

- Multidisciplinary 患者選択委員会でレビュー

- 6-8時間のアセスメントとカウンセリング

- 32 全ゲノム, 550 全エクソーム（2015年4月まで）

- アメリカ病理学会（CAP）およびClinical Laboratory

Improvement Amendments (CLIA:CMS) 基準：最初外注 Froedtert 病院

- データ解析：in-houseのBIで

- Baylor医科大学病院 2番手（すでに準備?）

- Wisconsinに続いて臨床ゲノム配列解析

- 病院内にWhole genome laboratory 設立(2011.Oct)

- In-houseでシーケンシング/変異分析

- CAP/CLIA認証の検査室を病院内に立ち上げる。

- 臨床分子遺伝学者によって解析・結果報告

- そのほかにWashington大学、Partnerなど多数つづく



Wisconsin
小児病院

Wisconsin 医科大学 (MCW)

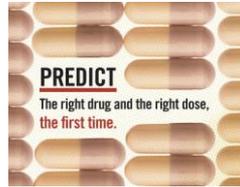


Baylor医科大学



第2の流れ

薬剤代謝酵素多型性のゲノム医療 バンダービルト大学病院



■ PREDICTプロジェクト

34項目の薬剤代謝酵素CYP多型性判定Chip
医師の処方オーダー時に警告提示（2010から）

Pharmacogenomic Resource for
Enhanced
Decisions in Care and Treatment



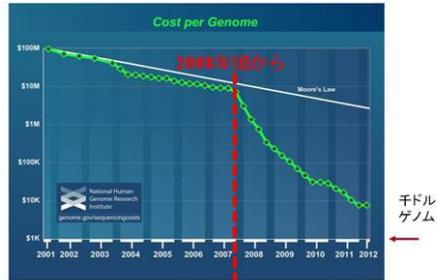
クロピドグレル処方
電子カルテの警告画面
商品名プラビックス：抗血栓剤
ステント留置手術の後に処方

CYP2C19の多型性で*2/*2の場合は
代謝機能が低いので(poor metabolizer)
血栓が凝固する
薬剤投与の応答は不十分である

この患者の場合(*2/*2)プラスゲレル
(商品名エフィエント)に替えるか

分量を2倍にしると警告している

第1の流れ ゲノム・オミックス医療の発展



DNA Sequencing Cost: the National Human Genome Research Institute

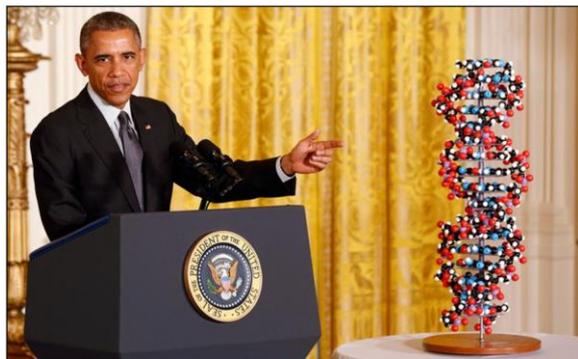
シーケンス革命 2007/8

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

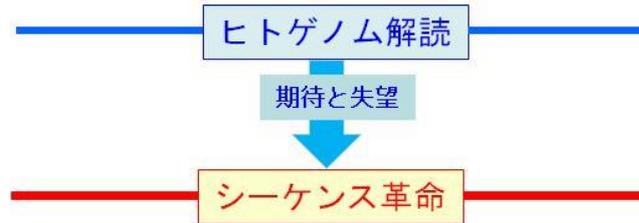


| | HiSeq2500 | Ion Proton |
|------------------|-----------|---------------|
| 本体価格 | 約1億円 | 約3500万円 |
| モード / チップ | ハイアウトプット | ラビッドラン |
| 解析時間 | 11日 | 27時間 |
| リード長 (bp) | 2 x 100 | 2 x 150 |
| データ産出量 (Gb) | 約600 | 約120 |
| 試薬コスト (ヒト1人全ゲノム) | 数十万円 | 不可 エクソームのみ |

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



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



2003年

2007年

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

ゲノム多型性の認識
.Hapmap2002開始
GWAS研究の興隆

TCGA (2006),国際
がんコンソーシアム
ICCG(2008)の
成果2011から出現

2009年

2010年

Undiagnosed
Disease原因遺
伝子のPOC同定
MCW小児病院

薬剤代謝酵素多型性
電子カルテで警告
Preemptive PGx
Vanderbilt大病院

Cancer Driver
Geneの同定と
抗がん剤治験
Mayo Clinic

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

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

第一期

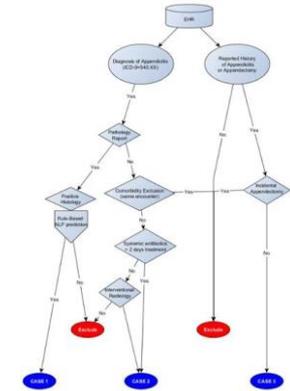
第二期

臨床表現型との統合 eMERGE 計画

electronic **M**edical **R**ecords + **G**enomics (NHRI-funded)

- **phase I (2007-2011) EMR-basedゲノム研究の探求**

- EMR(臨床phenotyping)とbiorepositoryに基づく
- GWAS等 (EMR-based GWAS) が可能か。
- 開始時は**GWAS全盛時代**。まだ**ゲノム医療の臨床実装**は始まらず
- 電子カルテより臨床表現型情報 phenotypingルール
- 計画開始時参加施設 : Mayo, Vanderbilt Univ., Marshfield, Univ. Washington, Northwestern Univ.など5施設,



PheKB: phenotyping ルール

- **phase II (2011-2015) ゲノム医療の臨床実装へ舵を切る**

- **MCWの臨床実装のインパクト**、Vanderbiltの**先制ゲノム薬理PG x**
- 電子カルテと遺伝情報の統合
 - 電子カルテへのゲノム情報の統合
 - **PheKB (Phenotype Knowledge Base)**
 - **ゲノム医療の実装、PGxの臨床応用**
 - **結果回付 Return of Result, ELSI等**
- 4つのサイトが新しく加わる
 - 小児病院グループとMount Sinai, Geisinger

- **phase III : 2015より始まる**

- **NHGRI予算化のコンソーシアムと連携**

- **CSE consortium等と連携**
 - “Clinical Sequencing Exploratory Research”



個別化医療から Precision Medicine

個人の遺伝素因・環境要因に合わせた (tailored) 医療
One size fits for all の Population 医療とは異なる

趣旨：基本は、個別化医療 Personalized Medicine の概念と変わらないが、目的は診断/治療の個人化ではなく層別化を明確化

概念の拡張：Personalized Medicineが標榜された時から10数年経っている

医療ビッグデータ時代の到来による個別化医療の拡張

(1) 遺伝素因 X 環境(生活習慣)要因のスキーマ重視

遺伝要因(SNPや変異：Genome)だけでなく環境・生活習慣要因(Exposome)の重視
疾患発症は2つの要因の相互作用と明快に強調。電子カルテの臨床表現型
(Clinical Phenome)情報の重要性認識。

(2) ゲノムコホート・Biobankの重視

Precision Medicineを実現する「情報基盤」として、ゲノムコホート/Biobankが必要であることを認識。Real world dataの重視

(3) 日常生理モニタリング情報の包摂

モバイルヘルス(mHealth)・wearableセンサーによる大量継続情報収集の重視

医療ビッグデータ時代の到来（米国）

ゲノム医療の実践

第1段階 ゲノム医療の発展

次世代シーケンシングの臨床普及 (2010~)

全ゲノム (X30 : 100Gb)・エキソーム解析 (X100 : 6Gb)

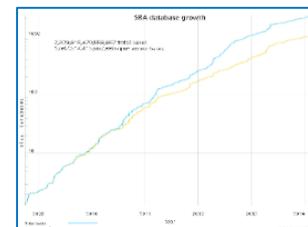
米国では数十の著名病院で実施

ゲノム・オミックス情報の蓄積



DNA Sequencing Cost: the National Human Genome Research Institute

2000兆塩基 (2 Pb)
が登録(NCBI:SRA)



第2段階 医療ビッグデータ時代

医療情報との統合

電子カルテからの
臨床フェノタイプ

医療ビッグデータ

学習アルゴリズム

ゲノム医療知識

人工知能AI



MayoClinicでは
10万人患者WGS

医療ビッグデータ

ビッグデータ医学/医療の第2の流れ

Biobankとゲノムコホートの世界的興隆

バイオバンクの目的・機能の変化

- 従来は再生医療のための生体標本や臨床研究の資料保存、近年はゲノム医療の基盤としての役割が認識され、世界的に普及
- ゲノム/オミックス個別化医療、創薬の情報基盤：**疾患型BioBank（疾患ゲノムコホート）**
 - 大規模調査にて、疾患罹患患者の網羅的分子情報（ゲノムなど）とそれに対応する臨床表現型情報（臨床検査、医用画像、処方歴、手術歴、病態経過、転帰など）の収集。
 - 疾病の分子機序や治療戦略、予後予測、創薬科学への貢献
- 個別化予防の情報基盤：**Population型BioBank（健常者ゲノムコホート）**
 - 「健常者」前向きコホート。調査開始時の網羅的分子情報（ゲノム）と臨床環境情報（exposome）を集めて、長期間（生涯）を追跡するゲノム・コホート
 - 主に遺伝子要因情報も含めた疾患の発症リスク予測、重症化予測

欧米のBiobank

- 英国 **UK biobank**
 - 50万人の健常者。40~69歳（2006-2010, 62Mポンド）、追加調査（2011-16, 25Mポンド）
 - 健診データ（血液・尿・唾液サンプル、生活情報）とゲノム情報（SNPアレイを集め、健康医療状況を追跡する。）
- 英国 **Genomics England**,
 - 2013開始、2017年までに10万人のゲノム配列収集。全ゲノム次世代シーケンス
 - 最初の対象は稀少疾患（患者・家族）、がん患者、最初はEnglandのみ。企業とのコンソーシアム
- 欧州 **BBMRI** (Biobank/Biomole. Res. Infra.)
 - 250以上の欧州各国のBioBankを統合
- オランダ **Lifeline**
 - 165000人北部オランダ 2006年開始 30年間の追跡、3世代コホート（世界初）
- **Precision Medicine Initiative Genome Cohort**
 - これまでのBiobank（例えばBioVUなど）を集めて100万人のゲノムを集める

ビッグデータ医学/医療の第3の流れ 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, 1000genome, ICGC, TCGA等
 - ゲノム変異・多様体
 - dbSNP, HGMD, **Clinvar**, Clingen, OMIM, GWAS catalog, Matchmaker Exchange
 - 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transcriptome: **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**, **STRINGS**, **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



医療ビッグデータ：まとめ

- 臨床ゲノム・オミックス医療の進展
 - Clinical Sequenceのインパクト
 - 網羅的分子情報も含めた臨床症例データ
 - 個別化医療、Precision Medicine
- Biobank, GWAS, 疾患レジストリの拡充
 - ゲノム患者対照分析、ゲノムコホート
 - Population型は個別化予防
- 網羅的分子情報DBの大規模化と利用
 - Clinvar, LINCS, HPRD, STRINGS

わが国での現状「ゲノム医療元年 (2016)」

■「ゲノム医学実現推進協議会」(中間報告) 2015.7

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

●例：がんの網羅的分子診断と個別化治療

- 国立がん研究センター東病院
 - ドライバー遺伝子の診断。分子標的薬の治験グループに割当て
- 静岡県立がんセンター 上記と同様の内容のプロジェクト
- 京大腫瘍内科 (OncoPrime), 岡大, 北大, 千葉大 診療施設併設型BB

■AMED (日本医療研究開発法人) がゲノム医療を推進

●IRUD (Initiative on Rare and Undiagnosed Disease)

未診断疾患の原因遺伝子をIRUD拠点病院が審査して解析センターがシーケンシング。その後、DB化する。

●ゲノム医療実現推進プラットフォーム事業

●臨床ゲノム情報統合DB事業

ゲノム医療では、米国と水を空けられている。しかし、Biobank Genomic Cohortでは我が国の状況はそれほど遅れてはいない。Biobank準拠のゲノム医療/創薬推進を行うべきである。また、第2世代のゲノム医療：多因子疾患に着手すべきである

ビッグデータ創薬/DR

創薬を巡る状況

- 医薬品の開発費の増大
 - 1 医薬品を上市するのに約700億円
- 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相損耗)
- 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
- 臨床予測性の向上
 - 罹患者のiPS細胞を使う
 - **ヒトのビッグデータを使う**



Drug Repositioning

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

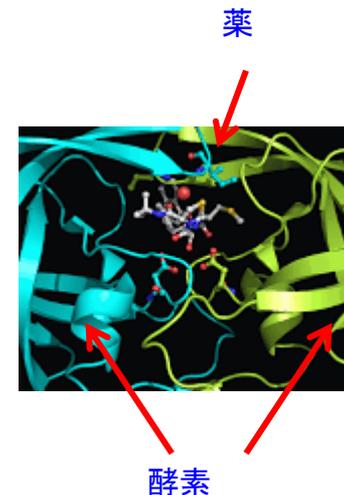
- (1) 既承認薬なので、ヒトでの安全性や体内動態などが既知で臨床試験で予想外の副作用や体内動態の問題により開発が失敗するリスクが少なく**開発の成功確率が高い**
- (2) 既にあるデータや技術（動物での安全性データや製剤のGMP製造技術など）を再利用することで、**開発にかかる時間とコストを大幅に削減できる**

ビッグデータ創薬/DR

計算論的(*in silico*)創薬の新しい方向

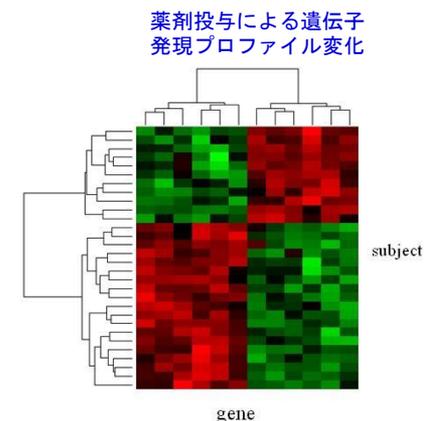
これまでの計算論的創薬

- 分子(構造)中心 (molecular-structure oriented)
- 分子構造解析・分子設計 (*in silico* drug design)
- Structure-based rational drug design
- 標的分子の分子構造解析
 - 薬剤 (リガンド) との結合構造 (結合ポケット)
- リガンドの分子設計—分子力学・量子化学
- リード化合物の構造最適化
- 定量的構造活性相関(QSAR)の利用



新しい計算論的創薬のアプローチ (*Mol. profile* drug design)

- 「オミックス創薬/DR, システム創薬/DR」
- 網羅的分子プロファイル・分子ネットワーク変化
- 薬剤—標的分子の結合を取り巻く genome-wide な分子環境, 「疾患システムの変化」という把握
- 化合物, 標的分子, 疾患間の関係の「ビッグデータ」



ビッグデータ創薬/DR方法

ビッグ網羅的分子KB/DBの利用

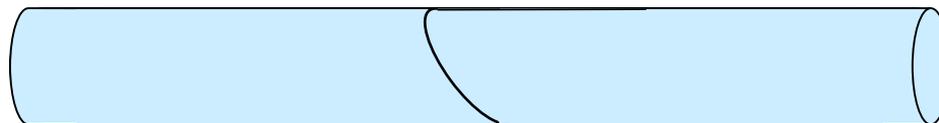
1. Transcriptome 創薬/DR
2. 疾患ネットワーク創薬/DR
3. Interactome 創薬/DR
4. 階層的ネットワーク創薬/DR

ビッグデータ・バイオリソースの利用

1. 疾患レジストリー準拠臨床無作為化治験
(registry-based clinical randomized trial)

KB/DB

Bioresource



創薬・DR・治験

ビッグデータ・BioBank



ビッグ網羅的分子KB/DB

Transcriptome創薬

遺伝子発現プロファイル空間での
ビッグデータ創薬/DR

ビッグデータ創薬/DRの基本原則 1

Transcriptome創薬原理

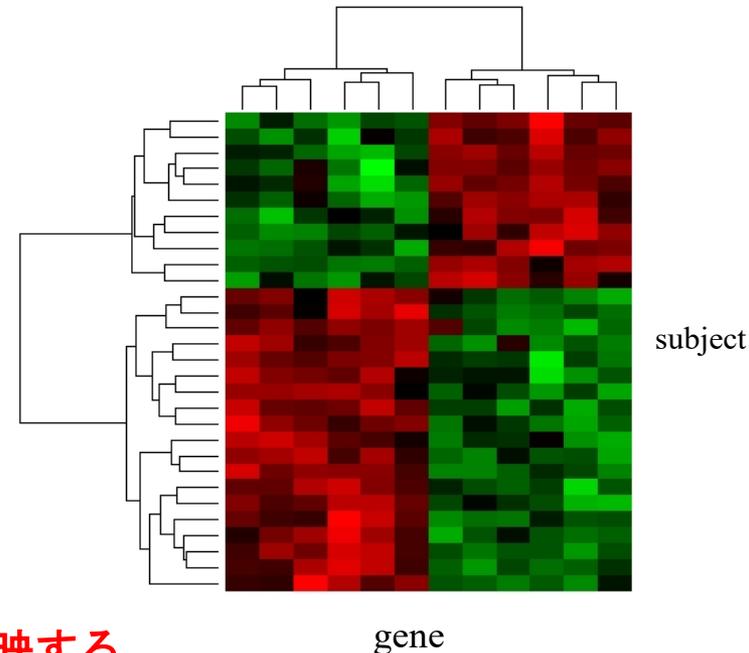
- 薬剤特異的遺伝子発現 (Drug-induced SDE)
 - CMAP(Connectivity Map)
 - 薬剤(化合物) 投与による遺伝子発現プロファイルの変化
 - 米国 Broad Institute, 1309化合物, MCF7, PC5など5がんセルライン, 7000 遺伝子発現プロファイル
 - Signature (遺伝子発現刻印 : 差別的発現遺伝子の代表的遺伝子群)
Signature of Differential gene Expression
 - DB利用 : SDEをquery, 順位尺度で類似性の高い順に化合物を提示
 - LINCS 100万サンプルへ拡張

- 疾病特異的遺伝子発現 (Disease-associated SDE)

- 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
 - 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

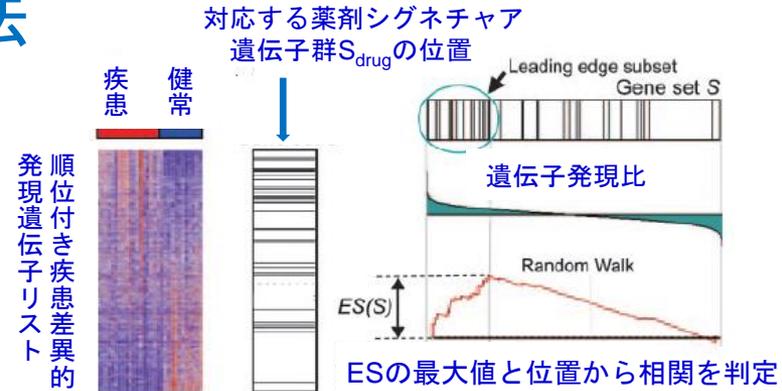


遺伝子発現プロファイルによる有効性予測

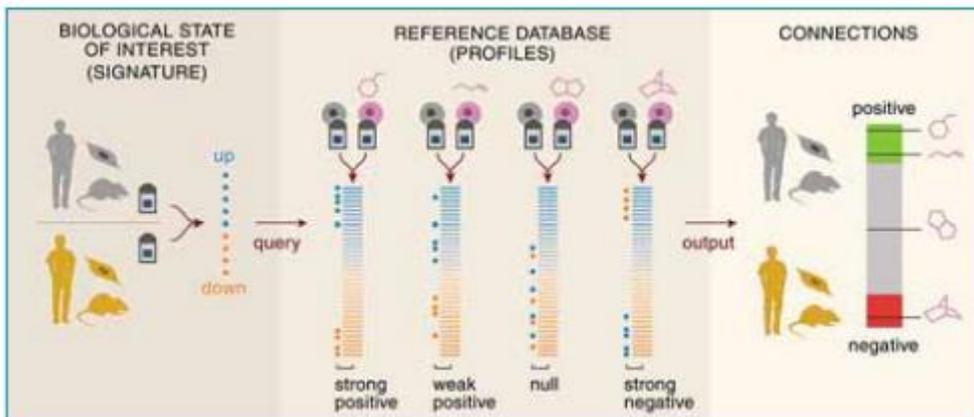
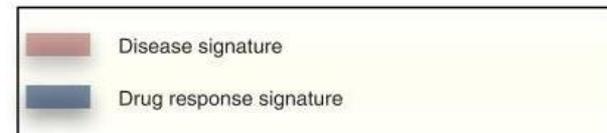
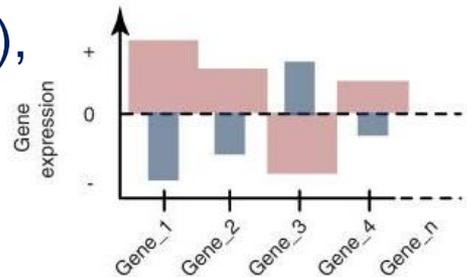
遺伝子発現シグネチャ逆位法 (signature reversion)

- 薬剤特異的遺伝子発現シグネチャ
- 疾患特異的遺伝子発現シグネチャ
- **有効性予測**：両者が負に相関する
- 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
- 例：炎症性腸疾患IBDに 抗痙攣剤(topiramate), 骨格筋委縮にウルソール酸

GSEA



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_{drug} 中に該当する遺伝子が存在したらESを加算、無ければ減算



遺伝子発現プロファイルによる**毒性予測**

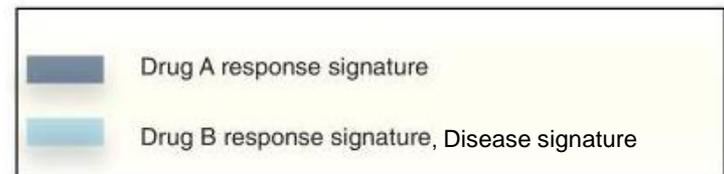
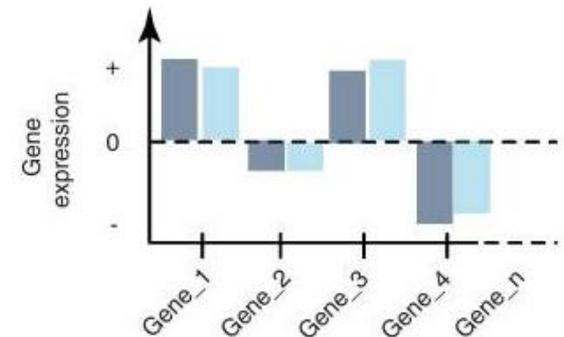
- 連座法 **guilt-by-association** :

- **薬剤－疾患間 副作用予測**

- 薬剤特異的シグネチャと
- 疾患特異的シグネチャが
- ノンパラメトリック相関 正
- **毒性・副作用の予測**

- **薬剤－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類似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



Transcriptome原理による 疾患－薬剤関係に基づく計算DR

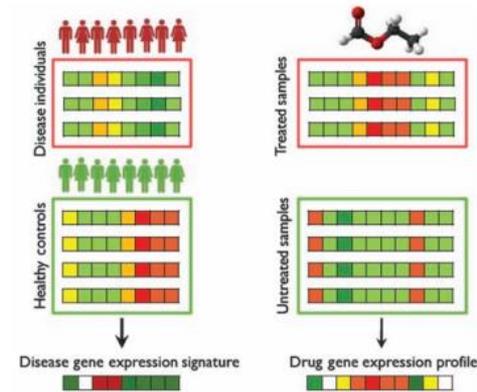
アプローチ (Sirota, Butte 2011)

- NCBI・GEOから
100疾患のシグネチャを取得
- c-Mapより得た164の薬剤・化合物
の薬剤特異的遺伝子発現profile
疾患-薬剤間で類似性スコアを計算
- 16000組の疾患-薬剤間の2664組が
有意、半数以上が治療的関連性あり
- 100疾患の内、53疾患が有意に164薬剤と関連

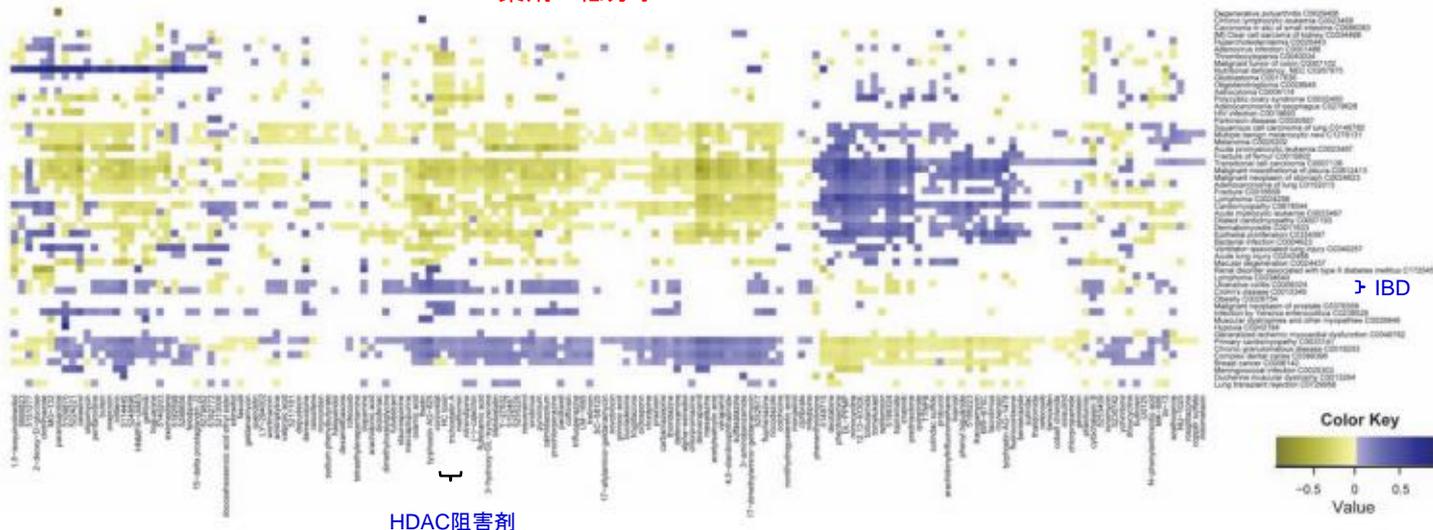
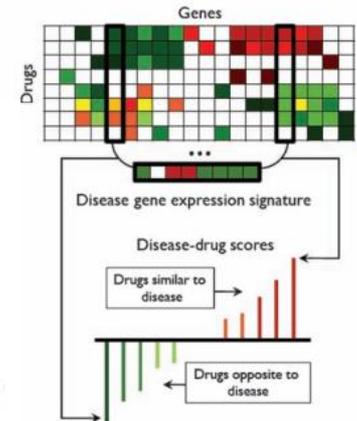
薬剤・低分子

Table 1. Drugs and diseases with the most indications.

| Drugs with most indications | | Diseases with most indications | |
|---------------------------------|---------|--------------------------------|----------------------------------|
| Vorinostat | HDAC阻害剤 | 21 | Transitional cell carcinoma |
| Gefitinib | | 18 | Melanoma |
| HC toxin | | 18 | Cardiomyopathy |
| Colforsin | | 17 | Adenocarcinoma of lung |
| 17-Dimethylamino-geldanamycin | | 16 | Multiple benign melanocytic nevi |
| Trichostatin A | | 16 | Squamous cell carcinoma of lung |
| 3-Hydroxy- α -kynurenine | | 15 | Malignant neoplasm of stomach |
| 5114445 | | 15 | Dermatomyositis |
| Dexverapamil | | 15 | Malignant mesothelioma of pleura |
| Prochlorperazine | | 15 | Primary cardiomyopathy |



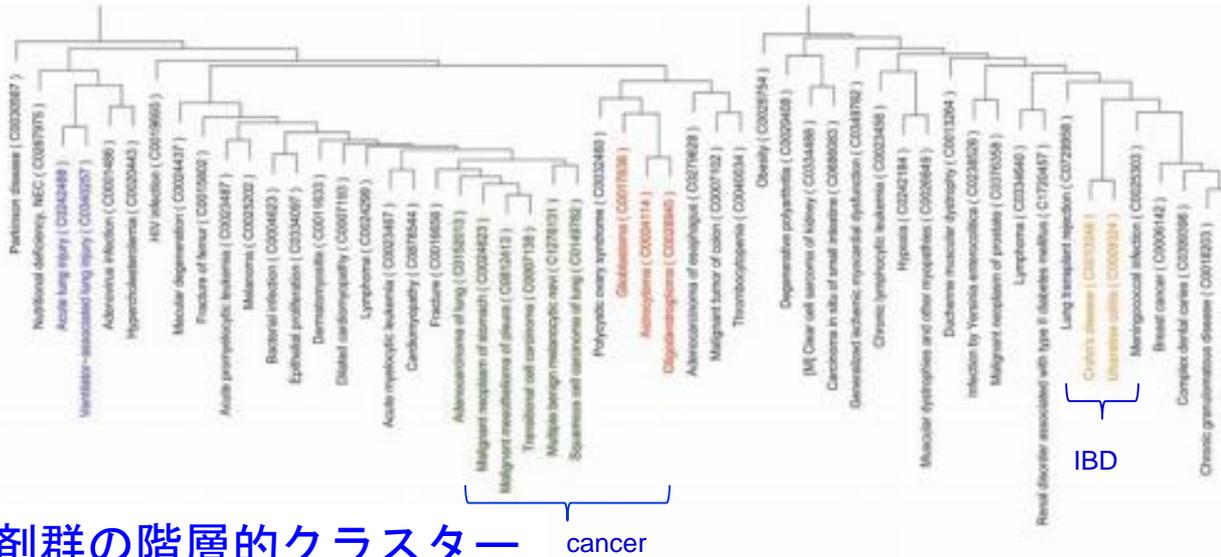
Reference database of drug gene expression



IBD

Transcriptome原理による 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

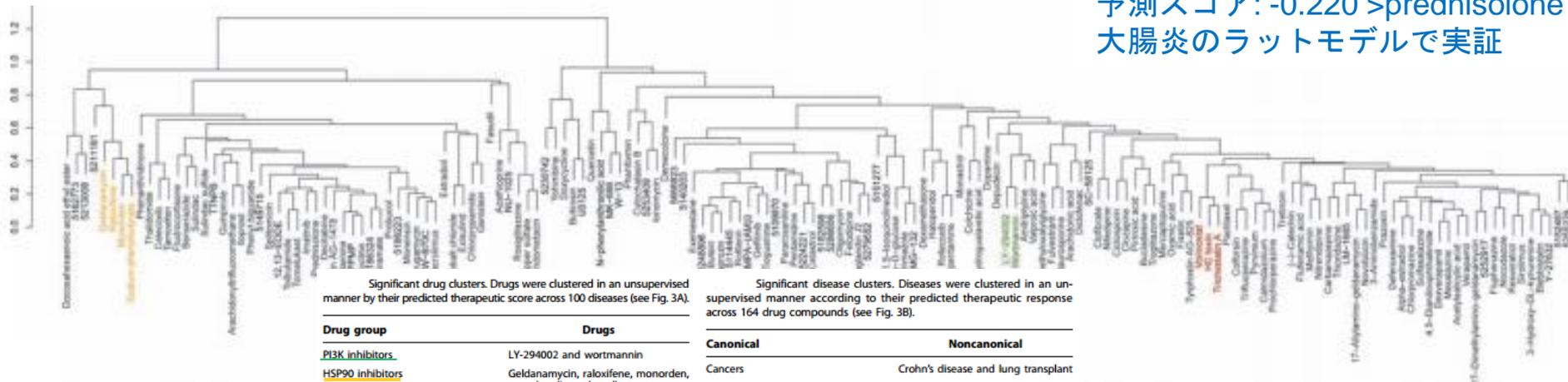
薬剤のクラスター解析

- ・HDAC阻害剤
Hctoxin, 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).

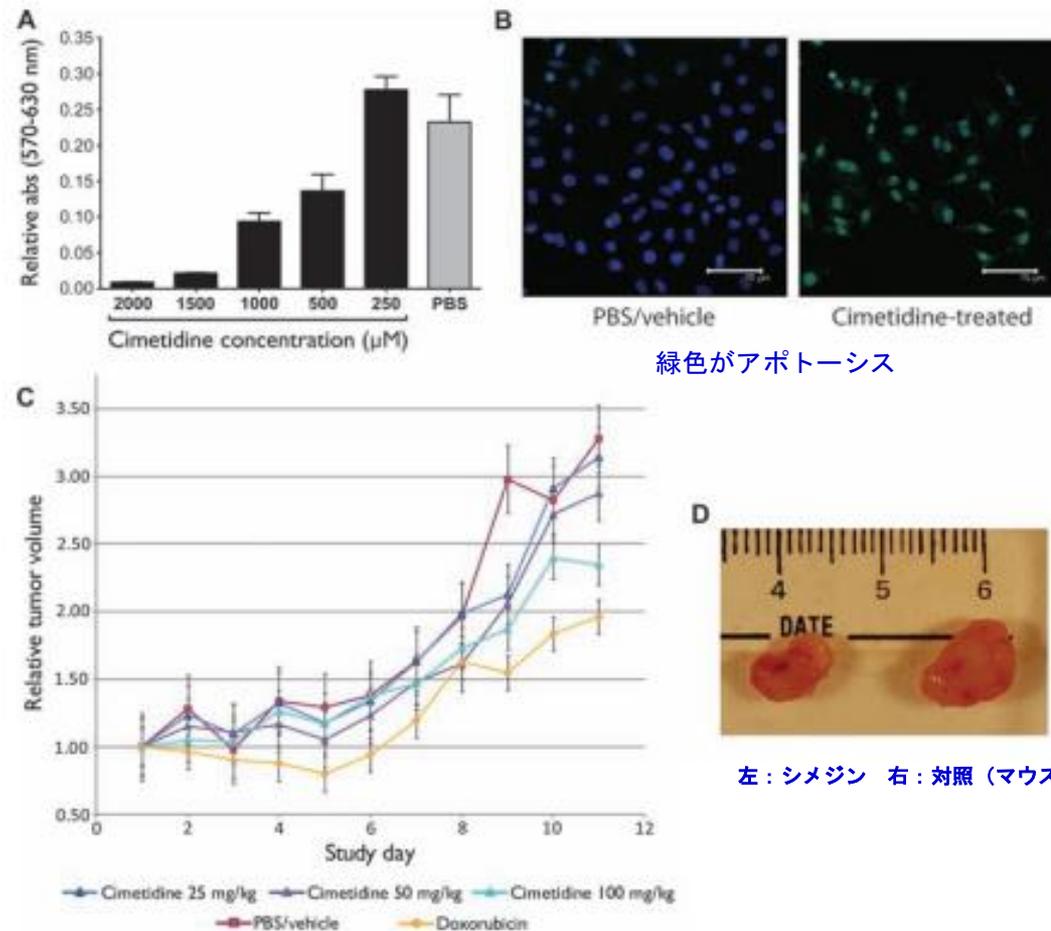
| Drug group | Drugs |
|-------------------------------------|---|
| <u>PI3K inhibitors</u> | LY-294002 and wortmannin |
| <u>HSP90 inhibitors</u> | Geldanamycin, raloxifene, monorden, and sodium phenylbutyrate |
| <u>HDAC inhibitors</u> | Vorinostat, HC toxin, and trichostatin A |
| Salicylate anti-inflammatory 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 |

動物実験での実証

シメチジン(cimetidine) →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



Transcriptional Profiling による疾患-薬剤ネットワーク (Hu, Agarwal)

遺伝子発現プロファイルの類似性を、相関係数、ESによってリンク

疾患-疾患、薬剤-薬剤、疾患-薬剤のネットワークを発現プロファイルより構成

疾患 (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

Solar keratosis 日光性角化症

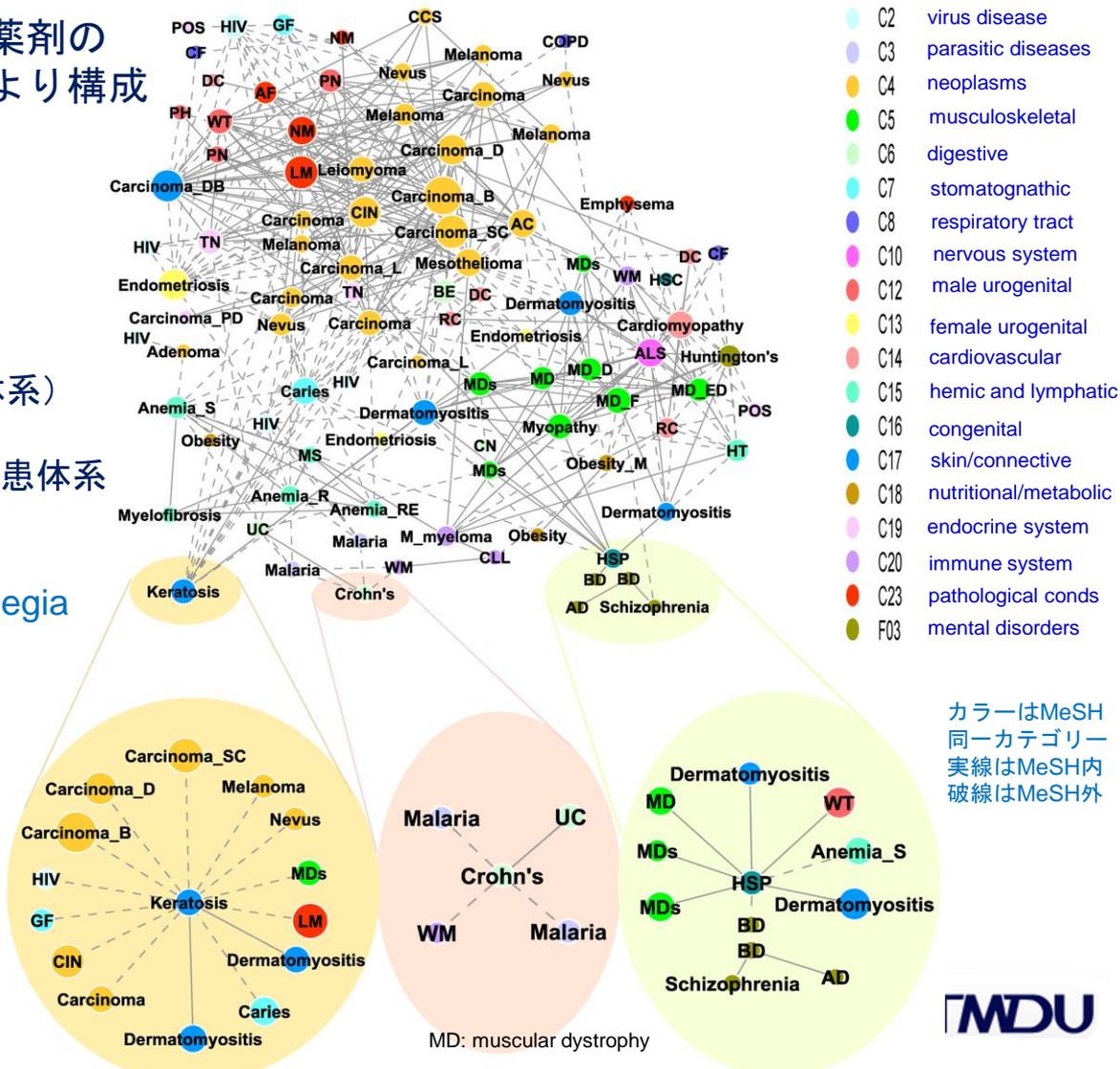
⇒ cancer(squamous)

<疾患 - 薬>

有効性：マラリア治療薬

⇒ Crohn's disease

ハンチントン病に種々の薬剤

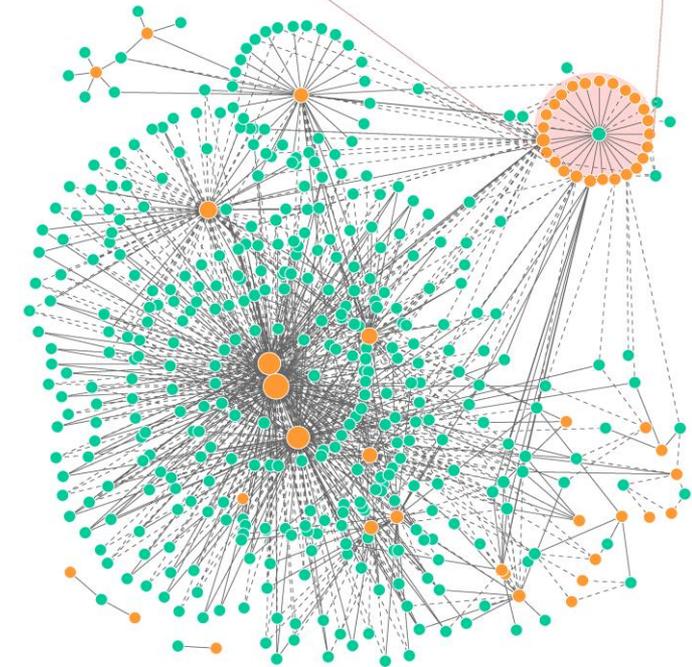
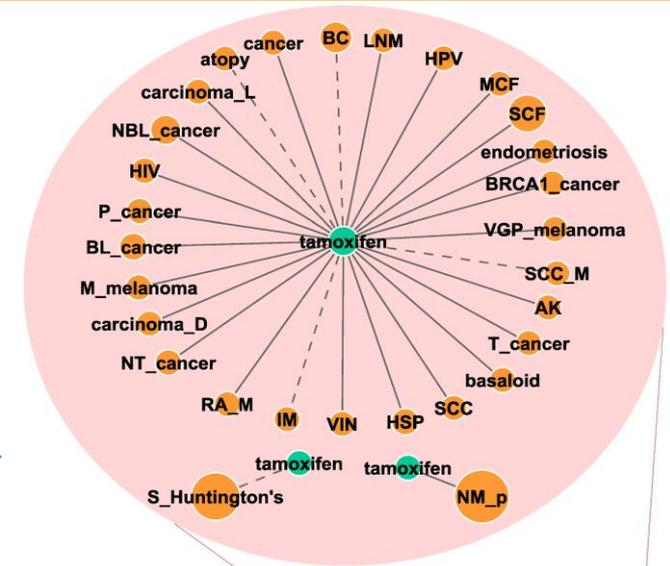


Transcriptional Profiling による Drug-Diseaseネットワーク

疾患－薬剤および薬剤－薬剤ネットワーク
(Disease-drug network: 右図)

橙色 49 疾患
緑色 216 薬剤
906 疾患－薬剤結合

Tamoxifen (breast cancer)
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

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⇒差別的発現 **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, **百万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ー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

LINCSの問合せ画面

--- LINCS Canvas Browser ---

Gene Lists

Up List

- EEF1A2
- UBE2S
- FAM64A
- FGFR1
- PAXIP1
- SPARC
- SNRPA1
- ADAMTS1
- EIF4EBP1
- PFKP
- BTG2
- CDK16
- ERRFI1
- ARPC4
- IFI30

clear

Down List

clear

Up Down

Search Example Enrich

Aggravate Reverse

Top 50 Consensus Experiments (Down/reverse)

| Overlap | Info (Perturbation, Dose, Time, Cell, Batch) |
|---------|---|
| 0.5000 | Tyrphostin AG 1478.56.78 μm 24 h A375 CPC006 |
| 0.5000 | PD0332991.2 μm 24 h MDAMB231.LJP001 |
| 0.5000 | PD0332991.10 μm 24 h MDAMB231.LJP001 |
| 0.5000 | PD0332991.10 μm 24 h MCF10A.LJP001 |
| 0.5000 | Aminopurvalanol A.10 μm 24 h PC9 CPC002 |
| 0.5000 | 3,5-dichloro-2-hydroxyphenylphenyl)benzenesulfonami |
| 0.4800 | PD0332991.2 μm 24 h BT20 |
| 0.4800 | PD0332991.10 μm 24 h BT20 |
| 0.4800 | MLN2238.10 μm 24 h BT20 |
| 0.4800 | 2-(6,6-dimethoxy-3-oxo-1,2,3,4-tetrahydro-1H-benzodiazepin-5-yl)carbamoyl)phenylpropan-1-ol.3.10 μm 24 h A375 |

Showing 1 to 10 of 47 entries

Average Change - Time Point - Drugs - Dose

IL1 100 ng/μl, 6 h in BT20

contrast:

Avg. Z-score:

Select a cell line: BT20

Select a batch: LJP004

Multiple Selections:

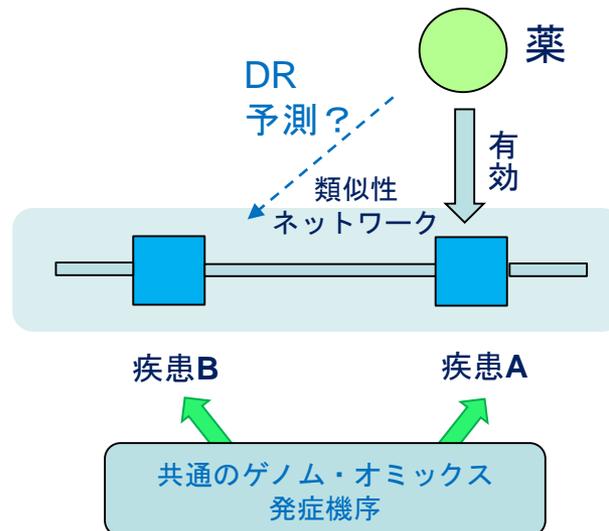
疾患ネットワーク創薬

疾患(類似性)ネットワーク空間での
ビッグデータ創薬/DR

ビッグデータ創薬/DRの基本原理2

疾患ネットワークに基づいたDR

- 従来の疾患体系 nosology
 - Linne以降300年に亘って表現型による疾病分類
 - 臓器別・病理形態学別の疾患分類学
- ゲノム・オミックスレベルでの発症機構による疾患分類
 - 発症機構類似性を基準に疾患ネットワーク
 - ゲノム・オミックス医学の疾病概念が基礎



疾患形成のゲノム・オミックス機序

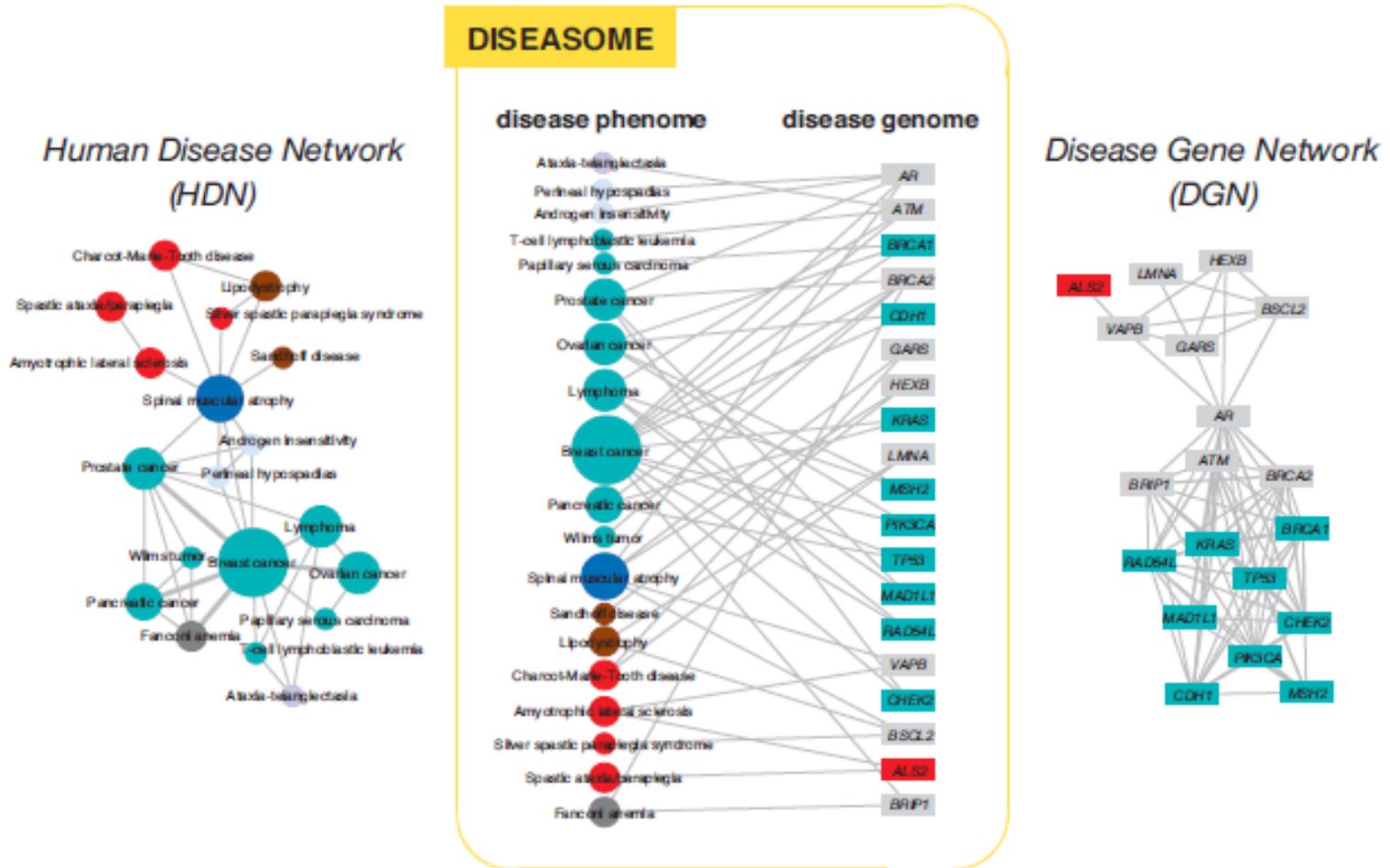
- 疾患関連遺伝子型（第一世代型）
 - 原因遺伝子、疾患感受性遺伝子の変異・多型が主要発症機序
- 疾患オミックス型（第2世代型）
 - 疾患オミックスプロファイルの変容が主要発症機序
 - Transdisease omics
- 疾患分子ネットワーク型（第3世代型）
 - 分子ネットワークの歪みが主要発症機序
 - がんなどで遺伝子型（肺腺がん等）でない通常のがん

第1世代型

Diseasomeと疾患遺伝子

- OMIMから 1,284 疾患と 1,777 疾患遺伝子を抽出
- ヒト疾患ネットワーク (HDN)
 - 867疾患は他疾患へリンクを持つ 細胞型や器官に非依存
 - 516疾患が巨大クラスターを形成
 - 大腸がん、乳がんがハブ形成
 - がんはP53 やPTENなどにより最結合疾患 がんなどは後天的変異
 - 疾患を網羅的に見る見方：臓器や病理形態学に非依存
 - リンネ (12疾患群分類) 以来300年続いた分類学を越える
- 疾患遺伝子ネットワーク (DGN)
 - 1377遺伝子は他遺伝子へ結合
 - 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を持つ

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



1つ以上の疾患関連遺伝子を共有する疾患

1つ以上の疾患を共有する疾患関連遺伝子

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

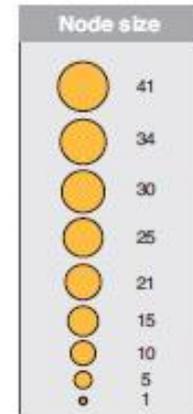
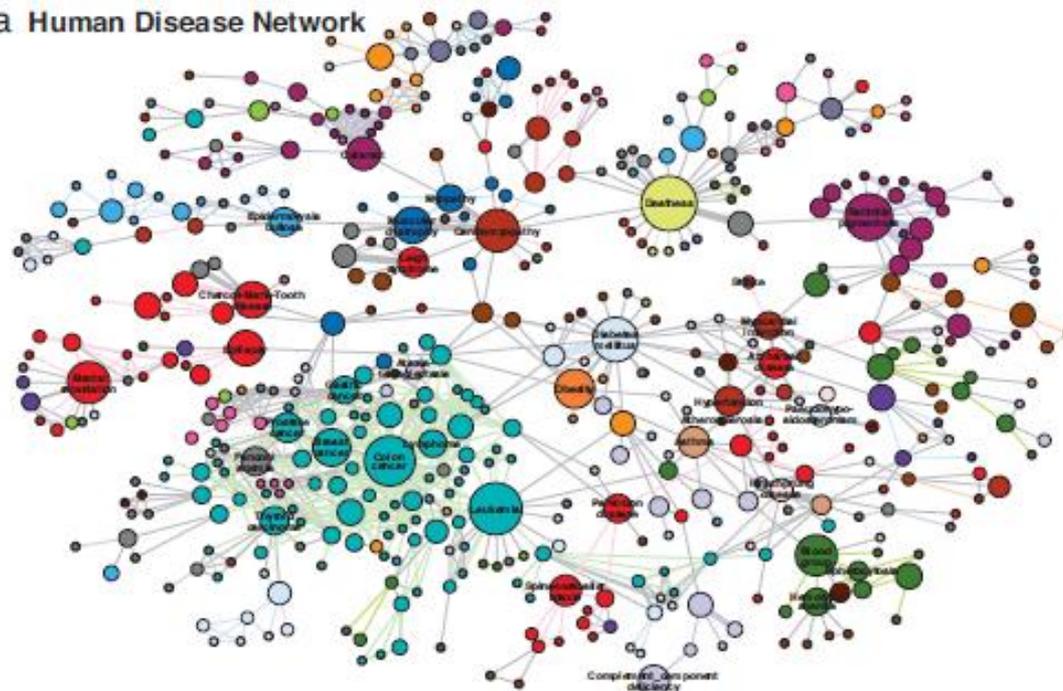


疾患 ネットワーク (HDN)

Nodeの直径
疾患に関与している原因
遺伝子の数に比例

リンクの太さ
疾患間で共有している
原因遺伝子の数

a Human Disease Network

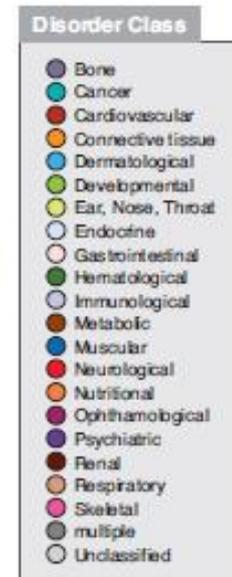
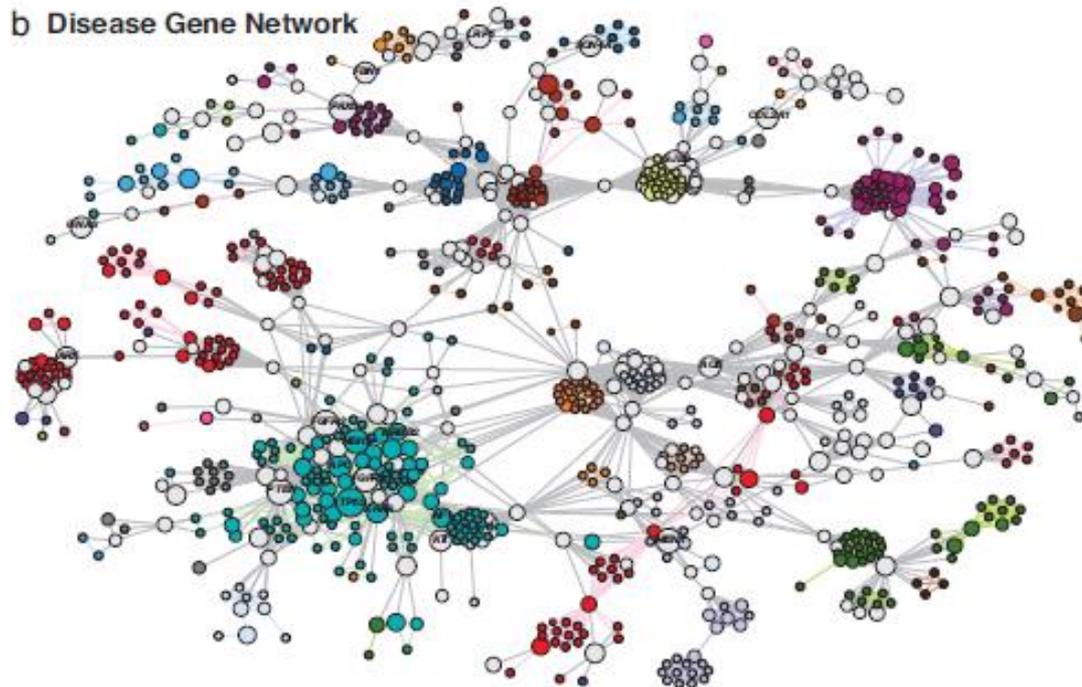


疾患遺伝子 ネットワーク (DGN)

Nodeの径
その遺伝子を原因にして
いる疾患の数に比例

2つ以上の疾患に関与し
ていると明灰色の遺伝子
ノード

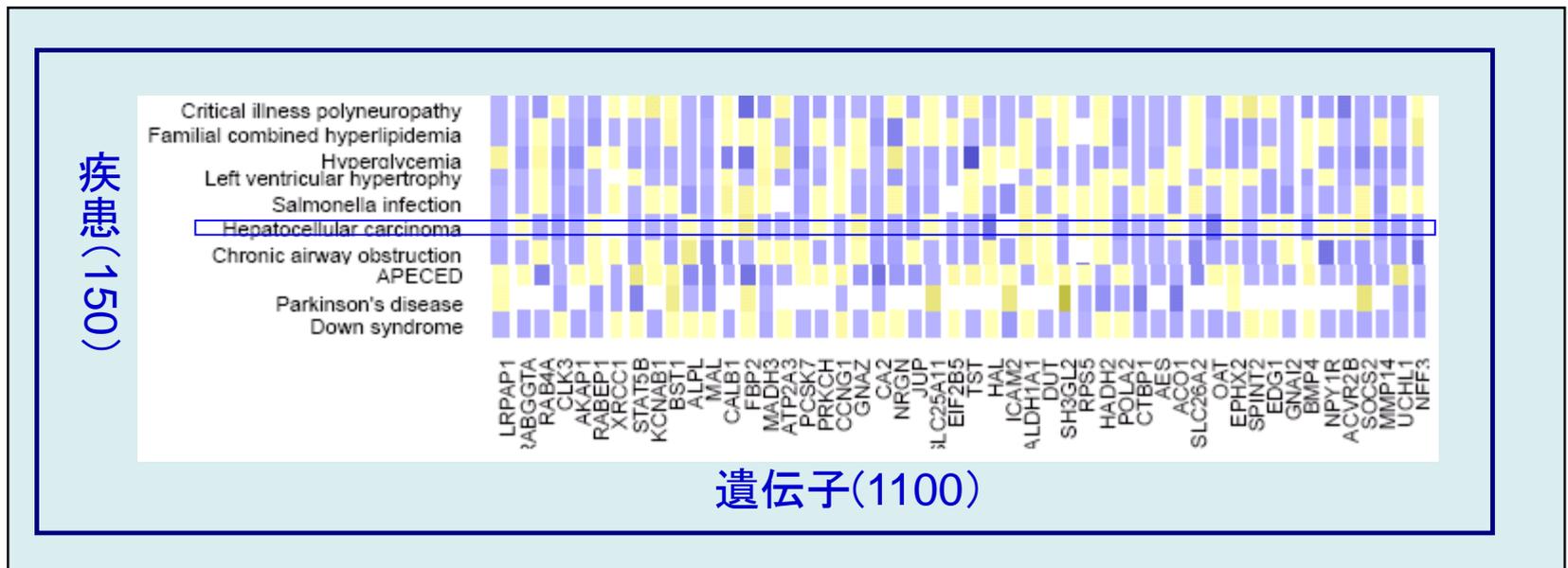
b Disease Gene Network



第2世代型

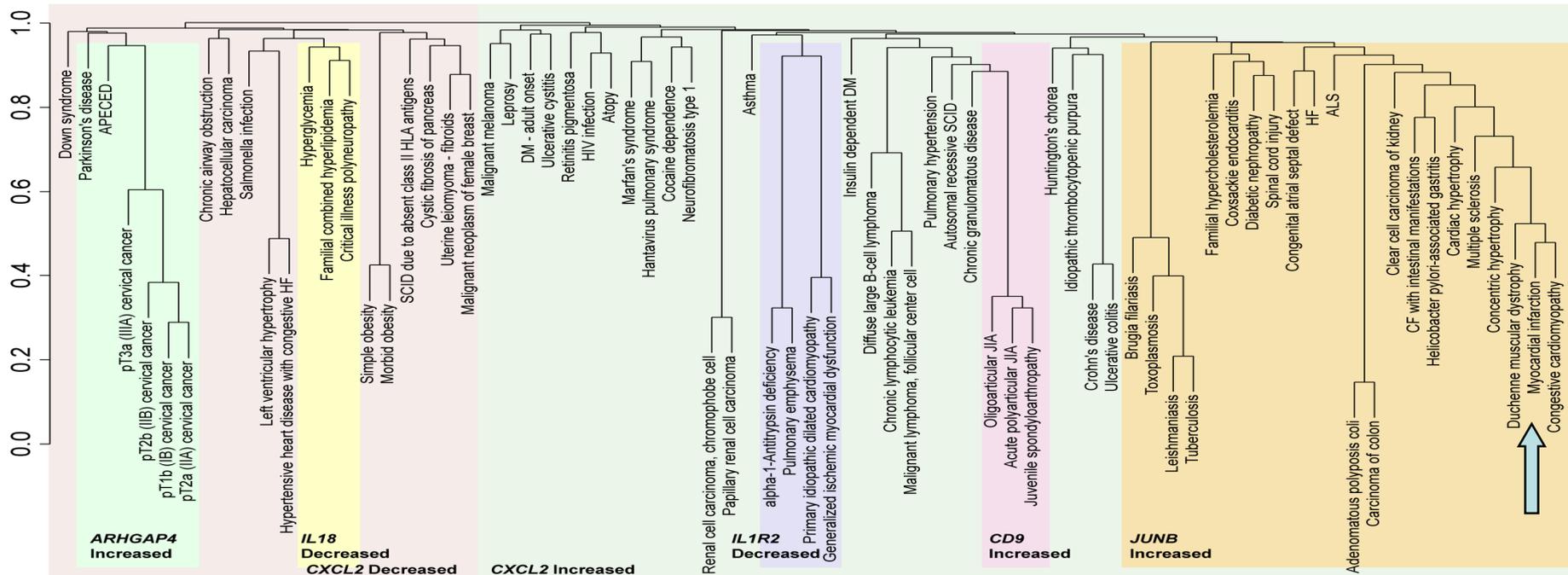
GENOMED (A. Butte et al)

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



Gene-Expression Nosology of Medicine

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



Transcriptomeの変化をPPIに投影した疾患ネットワーク (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4620に分解したPPIネットワークの機能moduleでの疾病での平均発現変化をもとに疾患ネットワーク構築

- 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

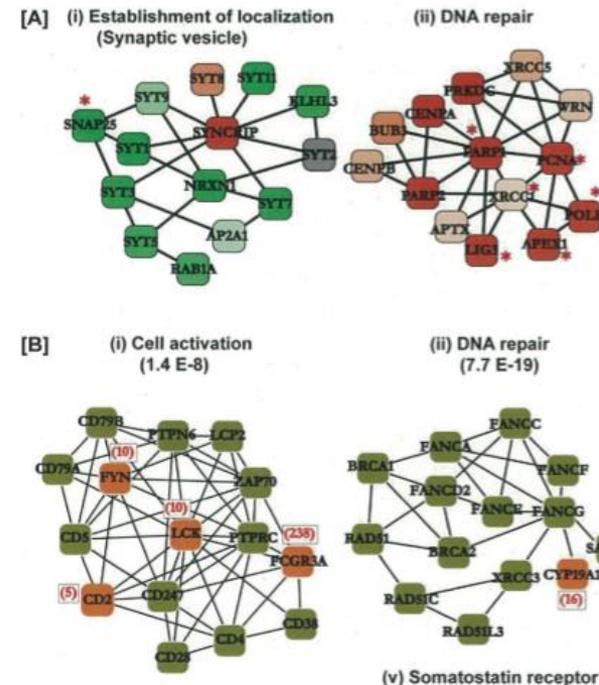
- **138の有意な相関**: ランダム化ネットに対し有意
- $p < 0.01, FDR = 0.1$
- **疾患類似性**: 肺がん群: 修復パスウェイ, 精神疾患: SNP-25

- 138の疾患相関のうち

- 17は少なくとも1つの共通薬 (14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

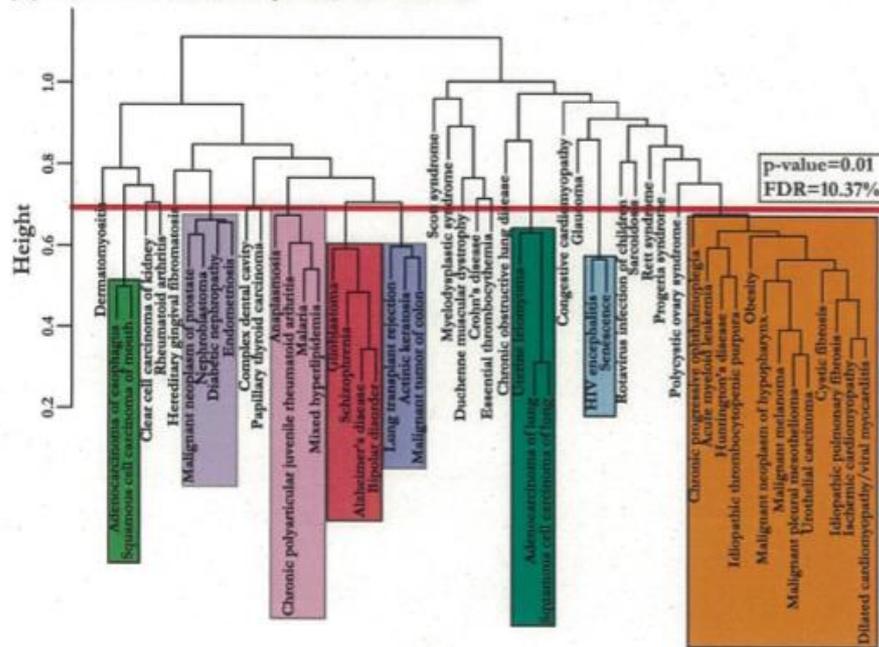
- 共通疾患シグネチャ



Transcriptomeの変化をPPIに投影した 疾患ネットワーク (Butte)

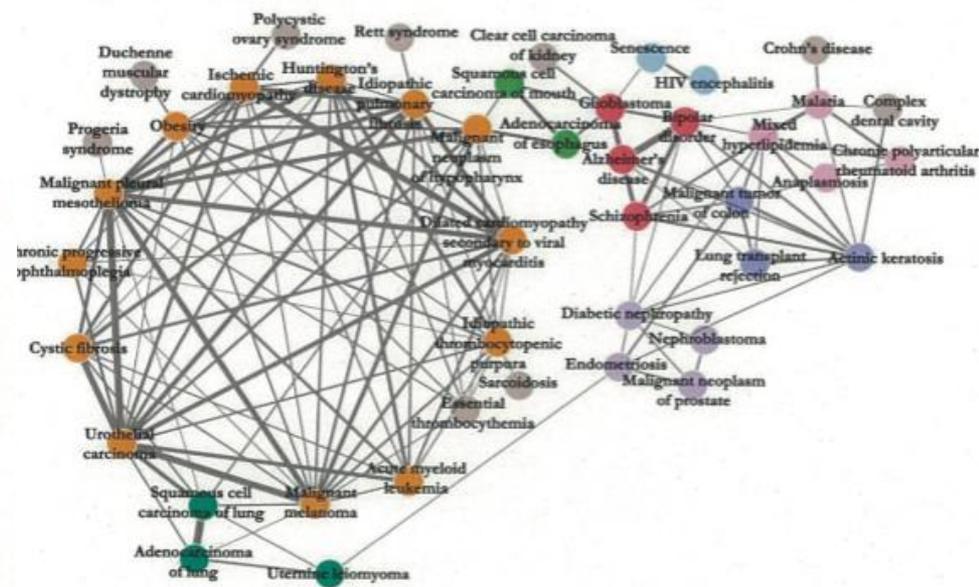
- アルツハイマー症、統合失調症、双極性障害がグループ化
- 子宮筋腫と肺がん、マラリアとクローン病
- 17のがんが1つの群ではない。がんの異質性
- 疾患ネットワーク間の遺伝子共有は高くない (遺伝子外効果)

[A] Hierarchical relationships between diseases



階層的クラスタリング

[B] All significant disease correlations

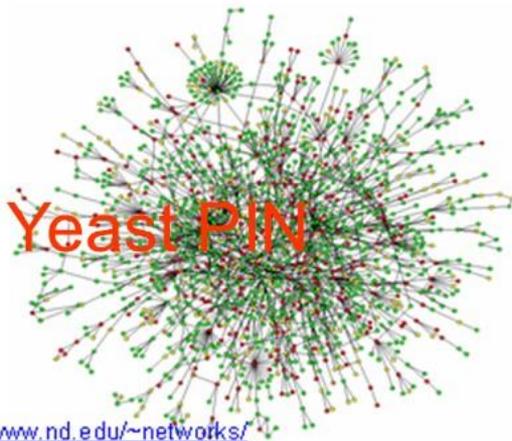
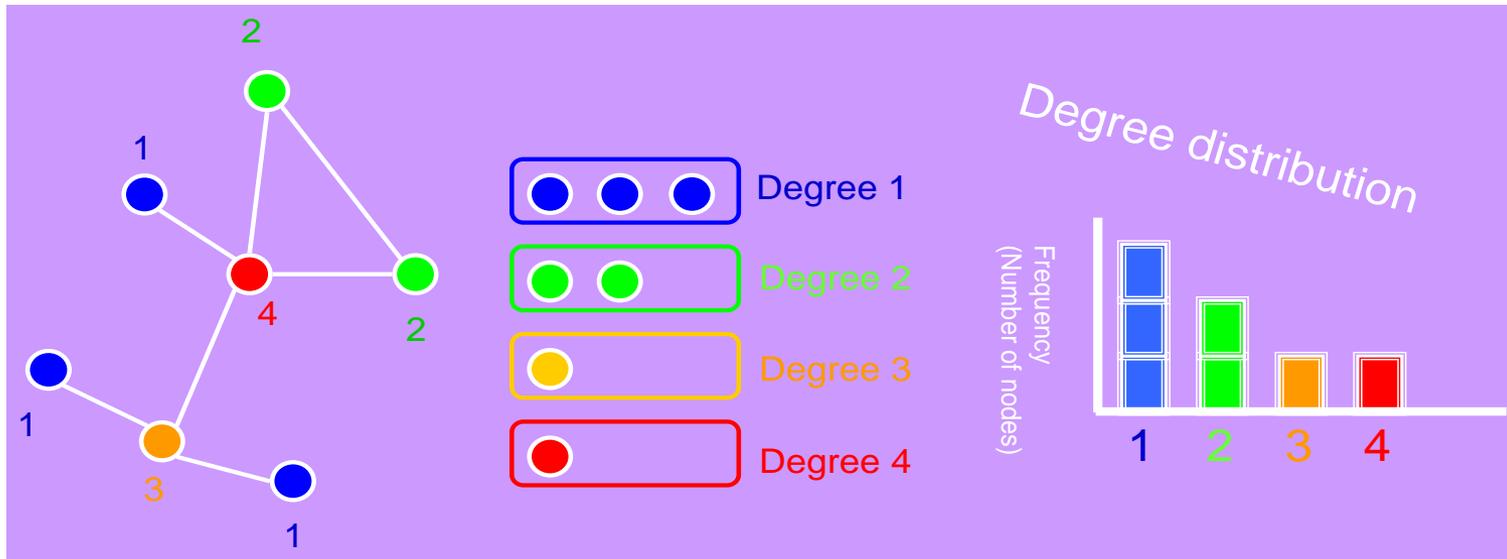


疾患ネットワーク

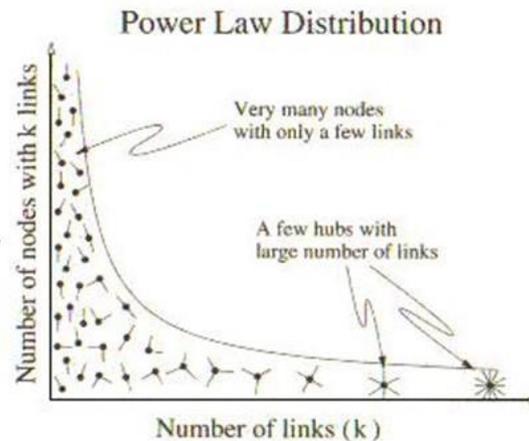
Interactome 創薬

タンパク質相互作用ネットワークに
基づいた創薬/DR

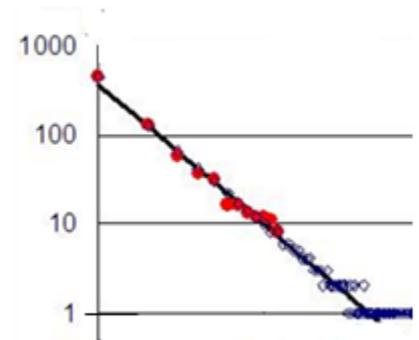
タンパク質相互作用ネットワーク(PIN)では数少ない相互作用が集中したタンパク質(hub)と相互作用が1や2の多数の末端タンパク質(branch)が存在する



<http://www.nd.edu/~networks/>

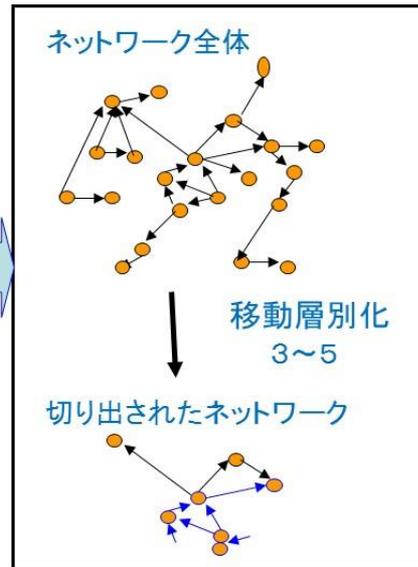
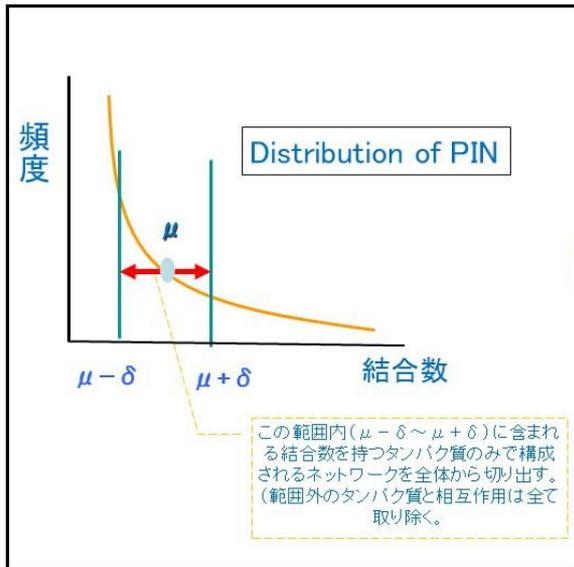


<http://www.macs.hw.ac.uk/~pdw/topology/ScaleFree.html>

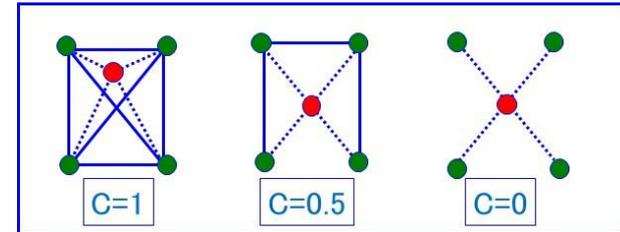


Log-log変換で直線

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



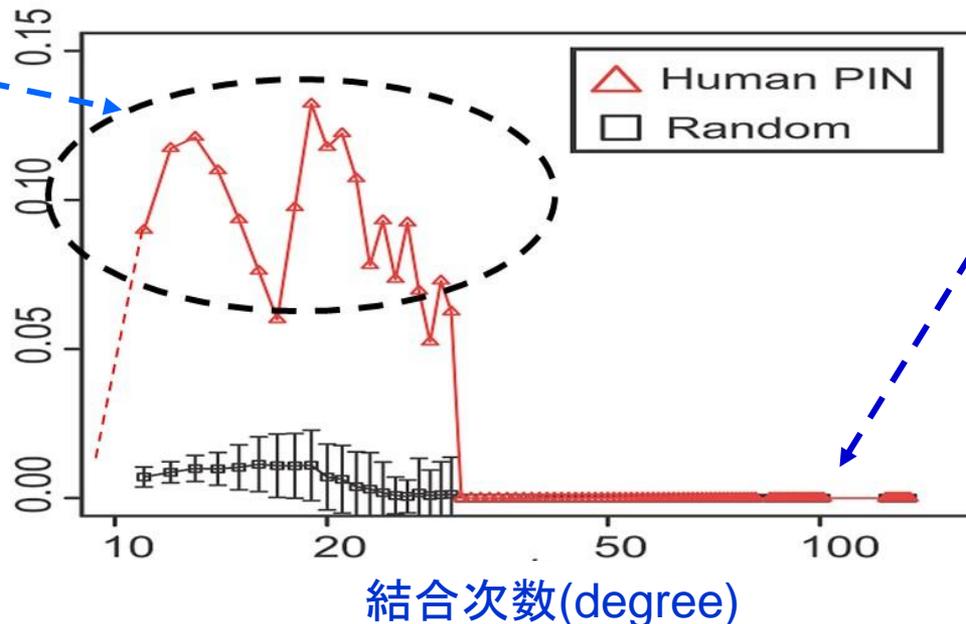
クラスター係数



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

中程度の結合数 (7~42) を持つタンパク質は多数の密なモジュールを構成

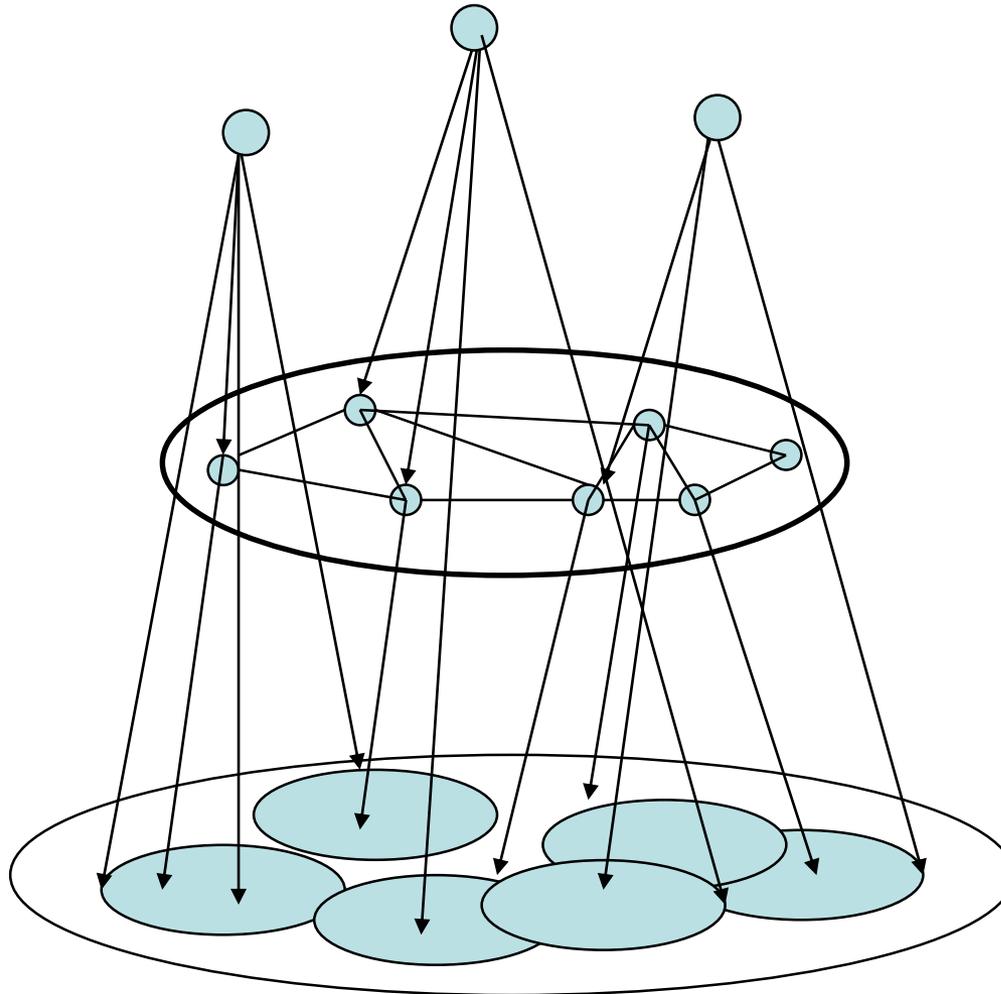
クラスター係数



高い結合次数を持つノード (スーパーハブ) はお互いに密に結合しない

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

高デGREEハブ

度数

> 31 ヒト

> 39 酵母

中間層

中程度デGREE

度数

6 ~ 30 ヒト

6 ~ 38 酵母

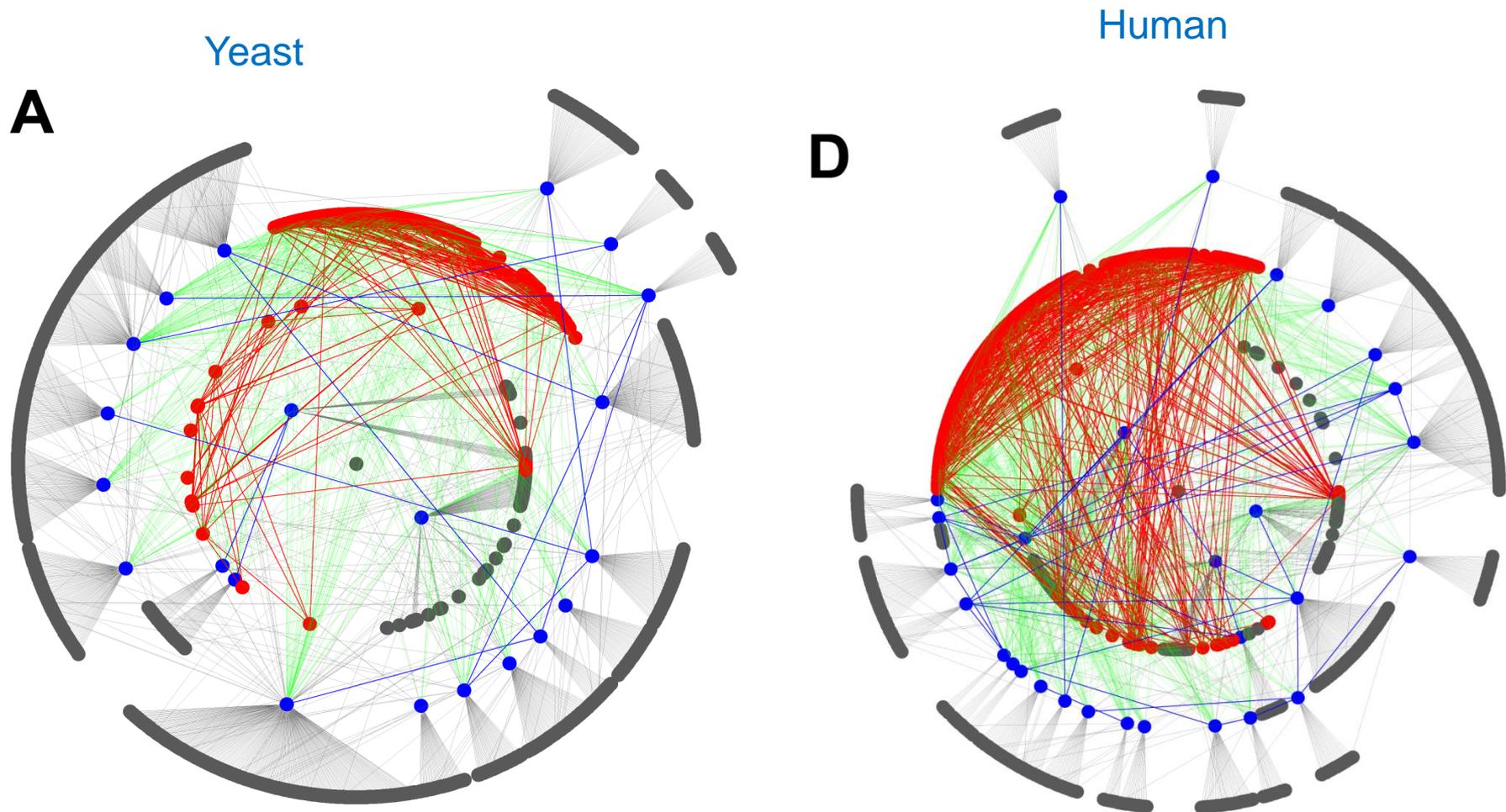
低層

低デGREE

ブランチ

度数 < 6

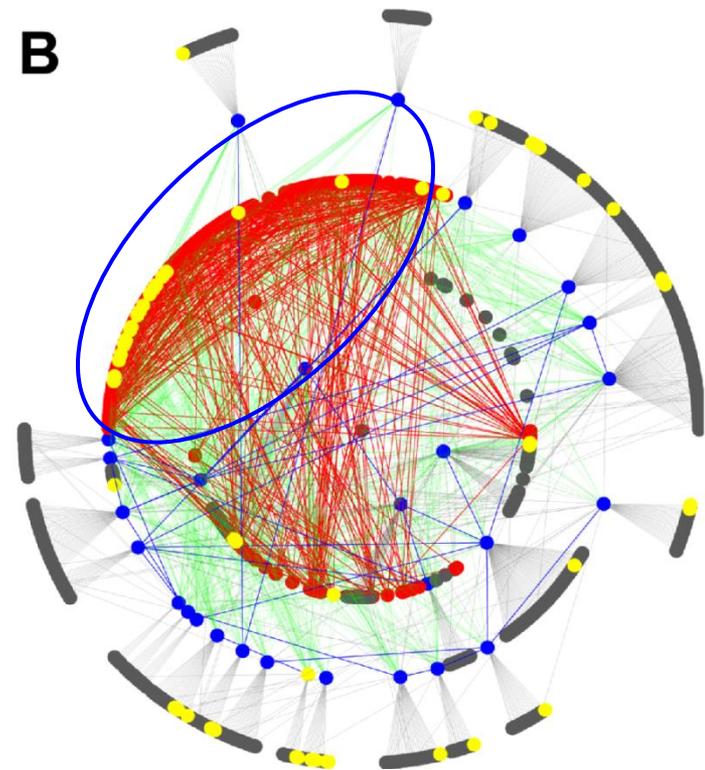
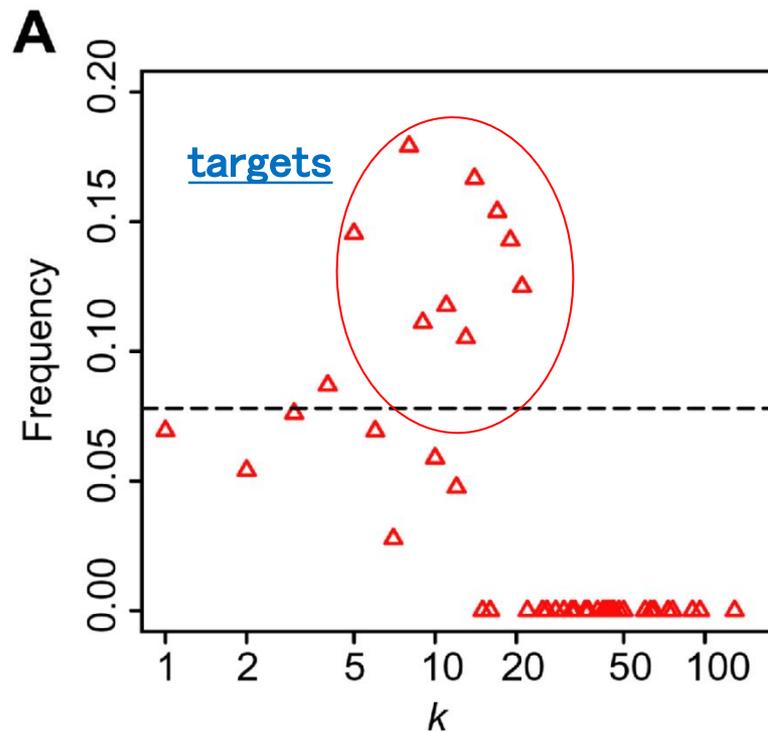
InteractomeのCloud Topology (3環トポロジー)



Middle-degree ノードは PPI backbone を形成する。

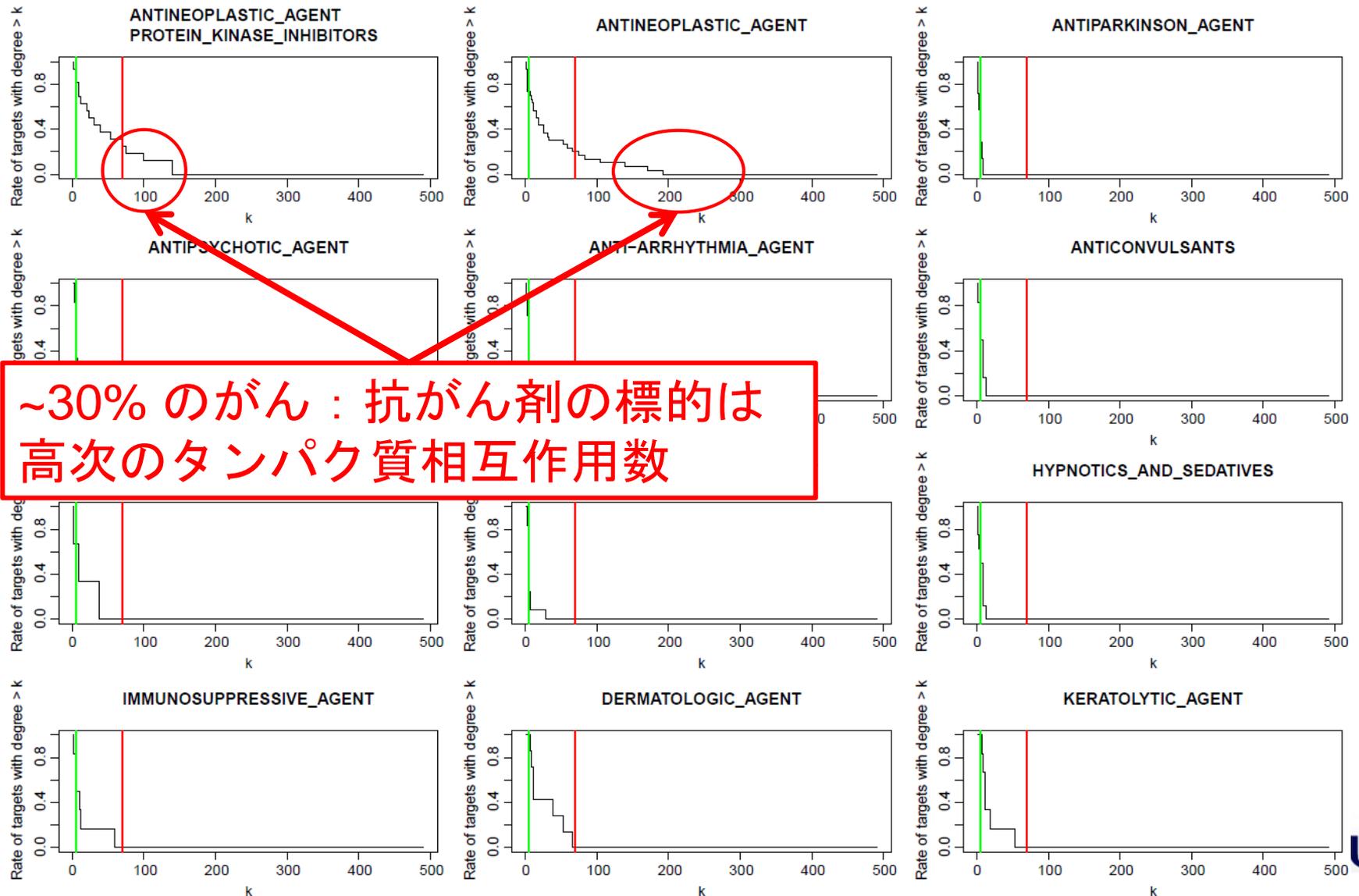
灰色, 赤, 青は、それぞれ低層、中層、高層のdegreeのノードをそれぞれ表す。

薬剤標的分子と結合度数



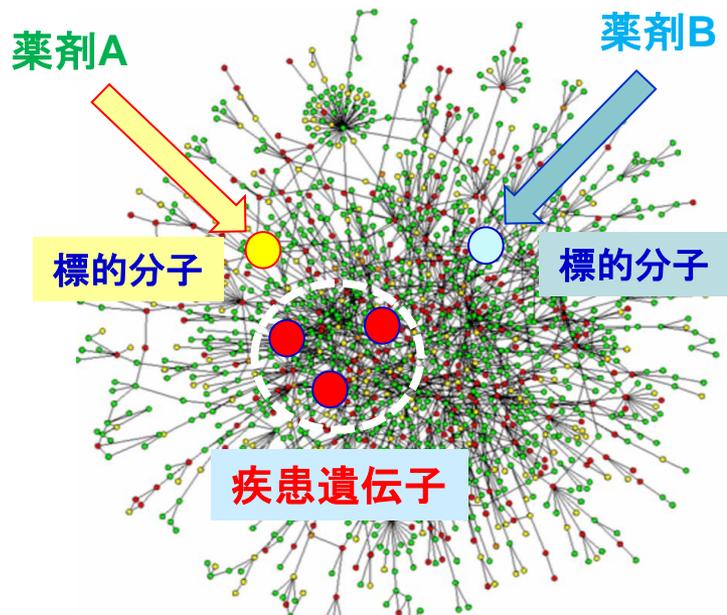
中層レベルのノードは治療薬として最適な標的である。それゆえ、多くの市場にある薬剤標的は、ヒトのバックボーンタンパク質に集中している

がんの疾患遺伝子は高次結合ハブのタンパク質が多い



Interactome創薬/DR

- タンパク質相互作用ネットワーク（PPI）空間/世界での創薬/DR戦略
- Interactomeの世界での検討
- **薬 剤**：薬剤の標的分子（タンパク質）によって PPI世界と繋がる
- **疾 患**：疾患特異的発現遺伝子をタンパク質へ翻訳、PPI世界と繋がる
- PPI世界内での薬剤と疾患の「代理人」の**距離・親近性**を基準に、薬理作用のインパクト力を評価
- random walkingで総合的な近さの評価を行う



タンパク質相互作用
ネットワーク
PPI: interactome

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)

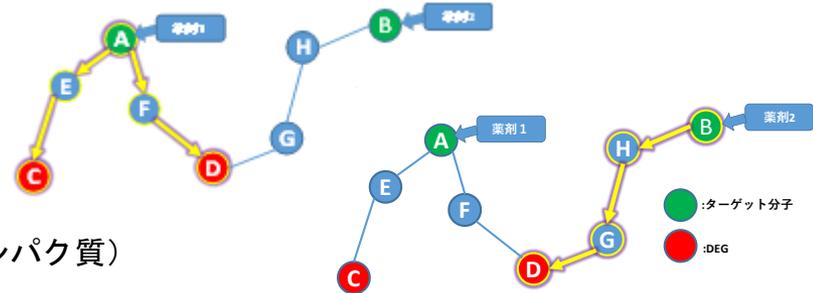
- **HPRD**

- 37,070 PPI, 9465 タンパク質

- **STRINGS**

- **薬剤⇒標的分子 : DrugBank**

- 7,759 薬剤、4300タンパク質
- 12,604 薬剤-標的分子 (4,452薬剤, 1,617タンパク質)



- **疾患遺伝子（肺腺癌）**

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

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

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

- 薬剤の標的分子と疾患遺伝子の間のネットワーク距離の総合評価

- 「再出発ありランダム歩行」RWRでネットワーク距離を評価

- 標的分子からランダム歩行を繰り返す（出発点から再出発あり）

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

- 一定の時間が過ぎると、定常状態になり、歩行で滞在確率分布は変化しない。

- 定常状態での疾患遺伝子ノードに滞在している確率の総和が薬剤の評価になる

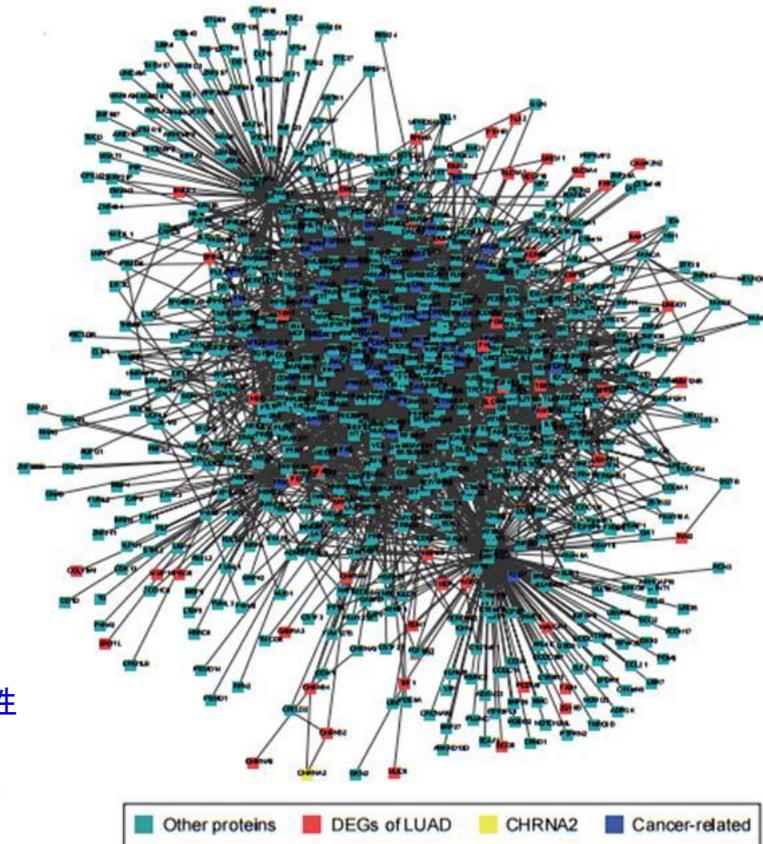
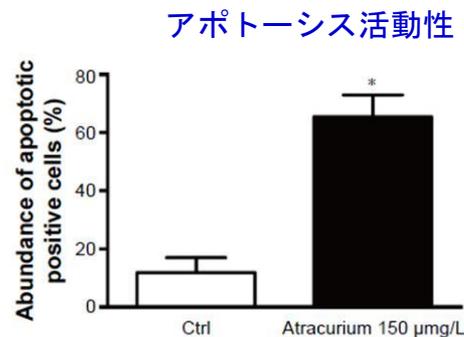
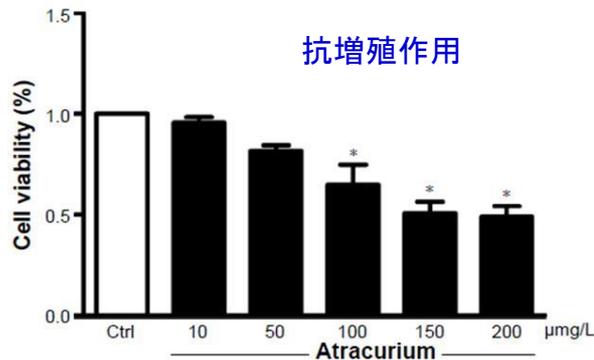
$$\mathbf{P}^{s+1} = (1-\gamma)\mathbf{M}\mathbf{P}^s + \gamma\mathbf{P}^0$$

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

Interactome DR 結果の検証

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

- HPRDとSTRINGSの両方のランダム歩行で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

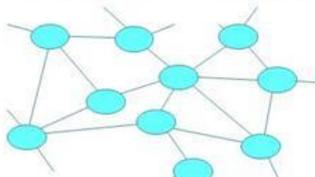


＜疾患-薬剤-標的＞の
多層ネットワークによる
ビッグデータ創薬/DR

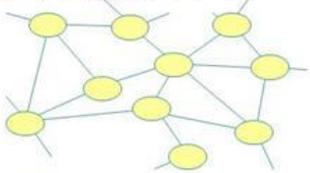
疾患ネットワークとDrug Projection Map

DR informaticsの構築

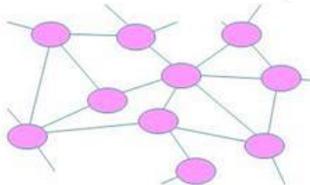
第1世代疾患ネット (原因遺伝子親近性)



第2世代疾患ネット (OmicsProfile親近性)

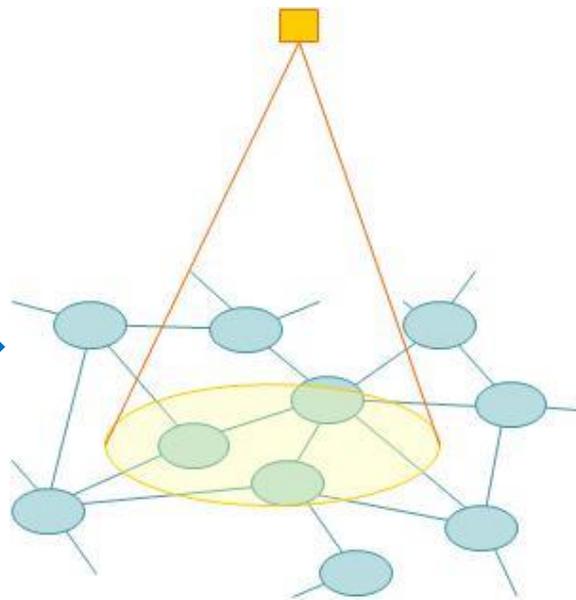


第3世代疾患ネット (Pathway親近性)



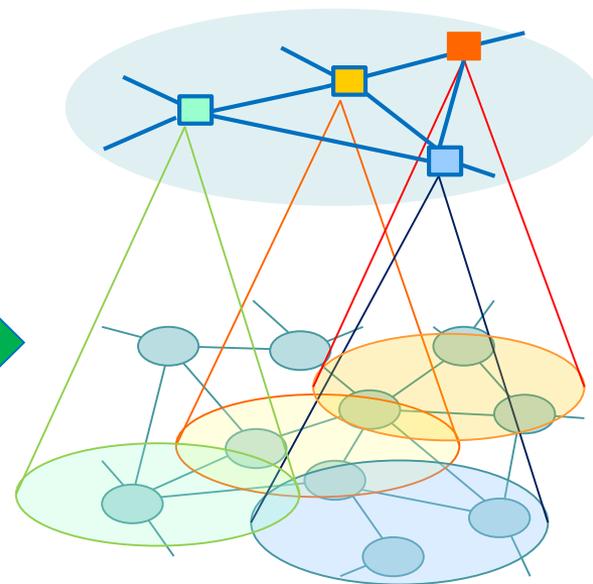
統合化

薬剤



疾患ネットワーク

薬剤ネットワーク

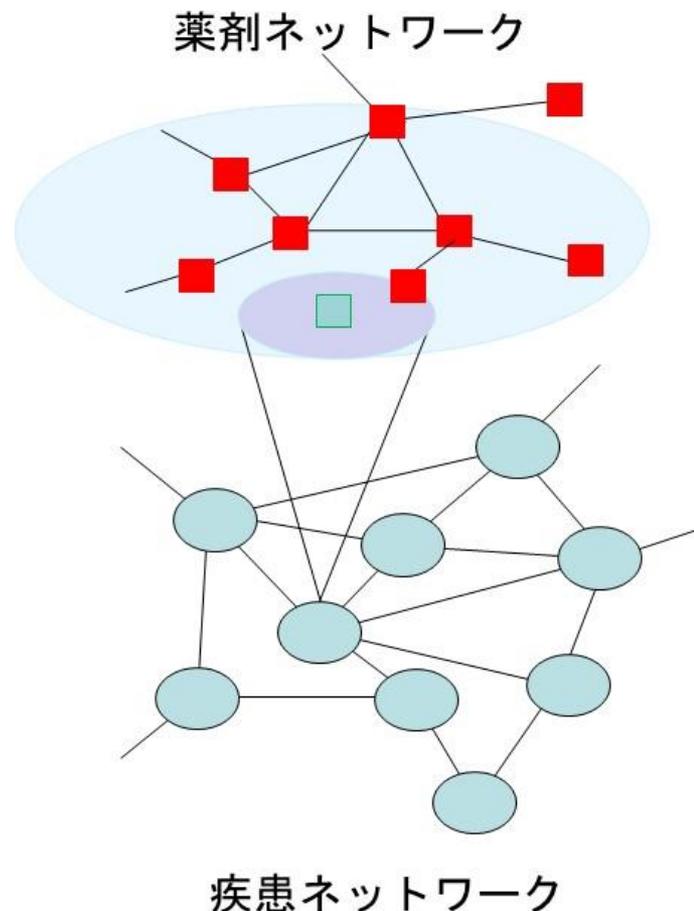
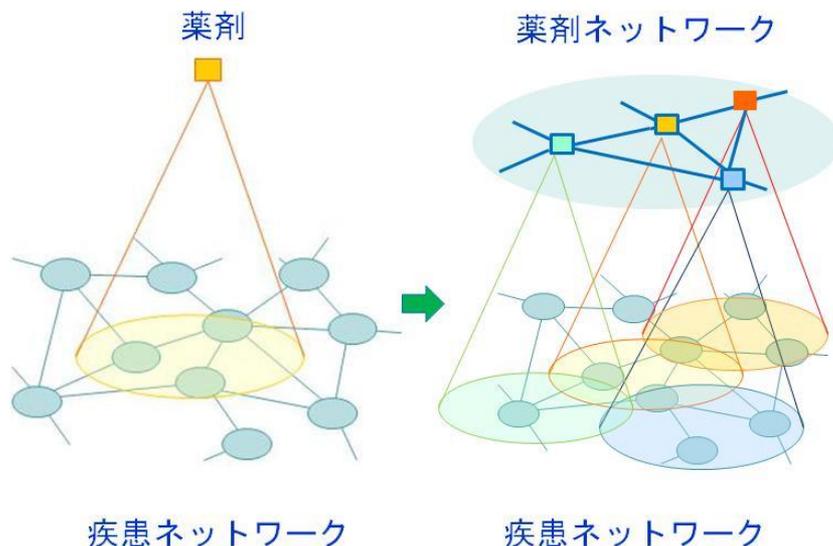


疾患ネットワーク

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
 - 多層的な疾患ネットワークとその統合
 - 医薬品の有効性・毒性の近傍Projection
⇒ DRにおけるfeasibilityは証明
- 創薬への発展
 - 薬剤・化合物空間のネットワークは既に確立
 - cMapでは不十分・LINCS(2014)
 - 疾患ネットワークの確立が重要
 - 疾患から逆投影。創薬の可能性探索
- 疾患トポロジーと薬剤トポロジーの双対性
 - Dual Topology-based Drug Discovery

疾病から薬剤ネットワークへの逆投影
Double Topology 双写像 創薬方法論



Heterogeneous Network 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

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

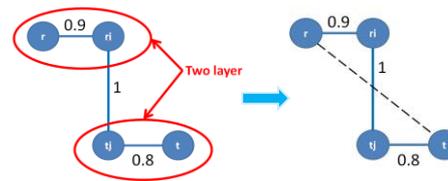
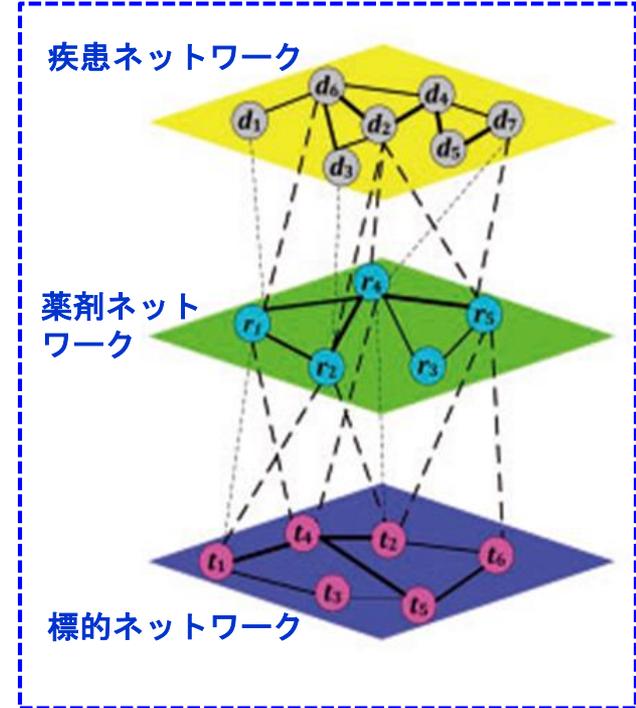
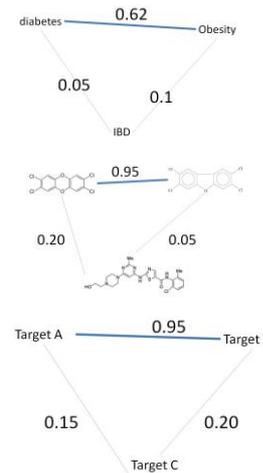
$$w(d, t) = \sum_{r_i \in R} \sum_{r_j \in R} w(d, r_i) \times w(r_i, r_j) \times w(r_j, t)$$

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

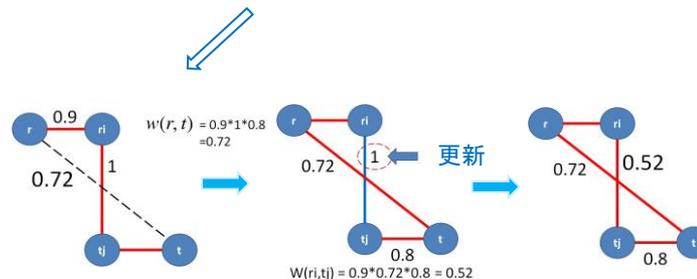
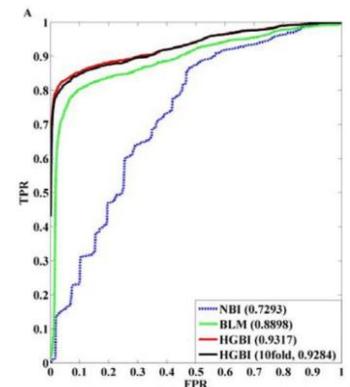
結合係数更新のマトリックス表示

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

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



従来の方法より
DR推定精度高い
ROC曲線



ビッグデータ・バイオリソース の利用

—RCT, EBMからの呪縛の解放—

Real-World-Dataを用いた 創薬/DR戦略

—RCT, EBMからの呪縛の解放—

個別化（層別化）医療の概念の普及とRCTの限界

- 個別化・層別化の概念の浸透
- RCTの治験集団とReal World Dataの乖離
 - 全ての個別化パターンを網羅した治験集団は現実には不可能
 - 現在の治験集団
 - 大半のRCTは医療現実の外の「人工的な環境」
 - 高齢者・妊婦はいない、欧米では黒人とくに青年は含まれない
- 将来へ向けたプラットフォームの確立
 - 母集団に近いReal World 医療データが収集可能
 - ⇒データの大規模化の「n = All」の実際
 - Real World Data時代の臨床研究のプラットフォーム
- 我が国での推進
 - 電子カルテの整備が遅れているので、急にRWDの創薬利用は困難
 - 米国はHITECH法のお陰でmeaningful use 基準で90%の普及
 - まずは、疾患レジストリー、疾患型Biobankからはじめる
 - On demand型のリソース提供は始まっている（病院併設型Biobank)



「学習する医療システム」 Learning Health System

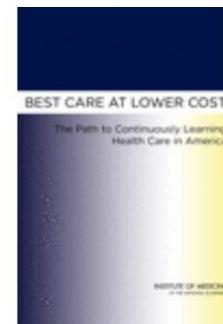
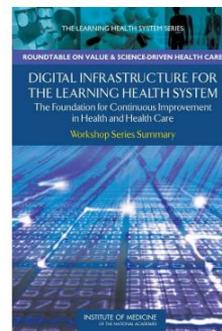
新しい生物医学知識が臨床実践に給されるまで17年
臨床データを用いて医療を実施しながら医療を改善

- IOM “Clinical Data as a Basic Staple of Health Learning”
- 医療システムのデジタル化（IT化）は必然の傾向である
- 「ルーチンの医療活動から集められたデータ（形式的臨床研究と違って）がLHSを支える鍵である」
- データを共有することによって学習して医療システムを改善
- RCTは「黄金基準」であるが、通常の医療システムの外で実施されている。医療が実際対象とする患者集団を代表しているのか。
- RCTは時間が掛かり費用もかかる
- 有効な知識の蓄積の速度が加速する

IOM(Institute of Medicine)のレポート
2007年にEBM/RCT（無作為試験）に
変わるパラダイムとして提案

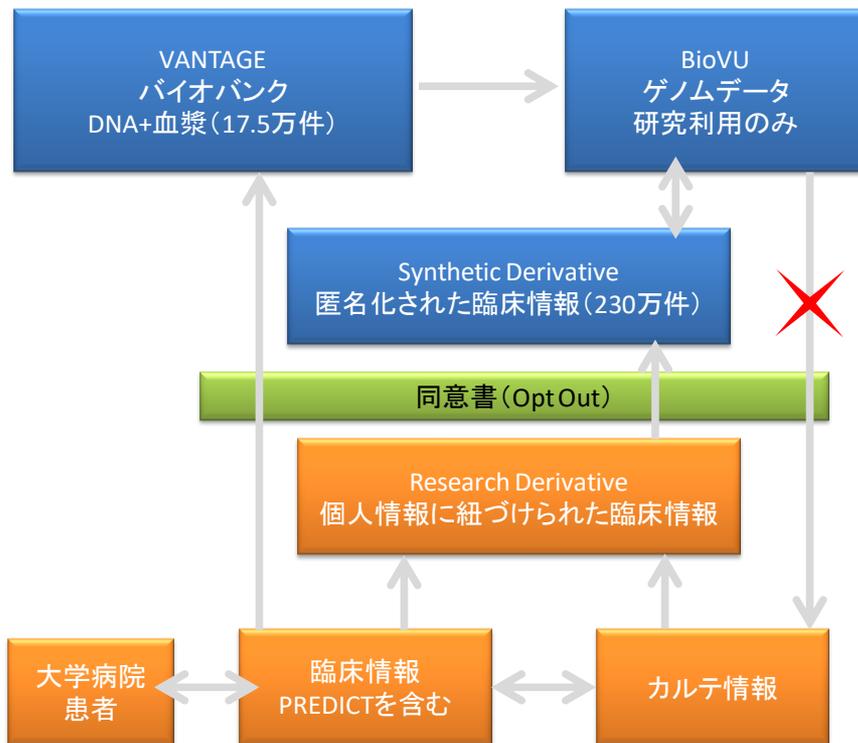
Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care

Best Care at Lower Cost: The Path to Continuously Learning Health Care in America



LHSの代表例 BioVU

ゲノム情報と電子カルテ情報を用いた Vanderbilt大学病院の医療情報システム



電子カルテ

Synthetic Derivative : 電子カルテから匿名化臨床表現型のデータベース 230万件。Opt out 形式

バイオバンクと遺伝子解析

BioVU : Synthetic Derivativeと連結可能な Genome DNA情報

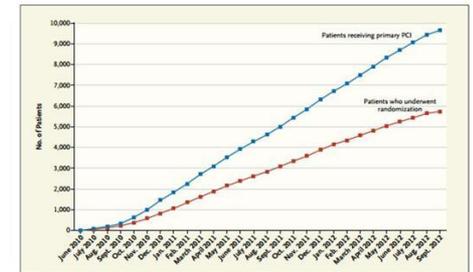
VANTAGE Core : 検体17.5万件、血液検からDNA抽出・ゲノム解析、バイオバンク運営

PREDICT : 臨床レベルの遺伝子解析情報により、薬物副作用防止などを実現するシステムを自らの医療システムにより知識抽出して実現する

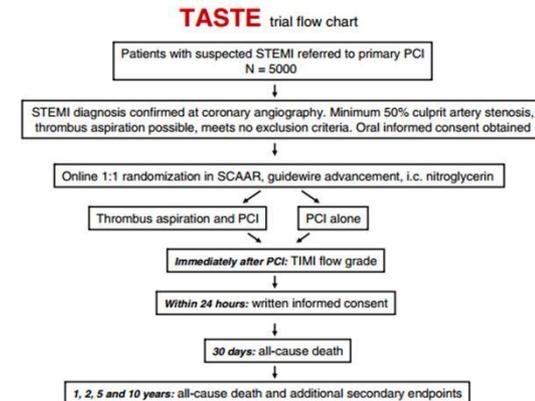
クロビドグレル（抗血栓剤）の遺伝子多型に関してABCB1, CYP2C19、さらにPON1の多型が知られていたが、ヒトを対象とした臨床実験の報告はなかった。SDから循環器疾患で clopidogrelの投与歴の対象者（ケース群）およびコントロール群を選出。BioVUから遺伝型を決定する。この条件に合致するケース群は255件。解析の結果、CYP2C19*2とABCB1の関与は有意。PON1は非有意が判明した。

Biobank (registry) 準拠の 創薬過程・治験

- スウェーデンの**TASTE**(STsegment-Elevation MI in Scandinavia)
Registry-based randomized clinical trial (RRCT)
- 国家的網羅的なRegistry登録者から治験対象者を選ぶ
 - SCAAR**(Swedish Coronary Angiography Registry)
- 選んだ集団で心筋梗塞のPCI*療法で
 - 血栓吸引を行った後、PCIを行う
 - PCIのみを行うの
 - 2群にランダム化割付し
 - 術後30日での生存をendpointとして治験を行う
- 治験のエンドポイントは疾患レジストリーの追跡
- これまで小規模の治験では相反する結論
- Observational研究：Population型コホートでは困難
- 経費は通常なら100万円程度を**50ドル**で済んだ。
 - *経皮的冠動脈形成術 (percutaneous coronary intervention)

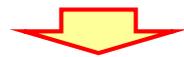


Rapid Randomization in the TASTE Trial, with Enrollment of Most Patients Receiving Primary Percutaneous Coronary Intervention (PCI). Adapted from the Institute of Medicine (www.iom.edu)/IOM/duffy/Activity/PCI/Quality/YSK/LS/ST%20Workshop/Presentation/Charger.pdf. The incremental cost of the Thrombus Aspiration in ST-Elevation Myocardial Infarction in Scandinavia (TASTE) trial was \$100,000, or \$50 for each participant who underwent randomization.



RRCTの特徴

- 質の高い大規模臨床レジストリーと前向き無作為化治験の長所を結合
 - 被治験者選択が容易
 - 迅速な登録
 - 非登録患者の制御
 - 非常に長期にわたる追跡が可能
 - アルツハイマー症など
 - 経費が掛からずデザインが単純
- 疾患レジストリーの方で
 - 患者鑑別
 - 無作為化
 - ベースライン情報の収集
 - エンドポイントの探索を行ってくれる



ビッグデータ研究とRCTの融合: 将来の治験方式

ゲノム・オミックス医療の 次世代の展開とlife-long healthcareへ

第2世代のゲノム医療へ

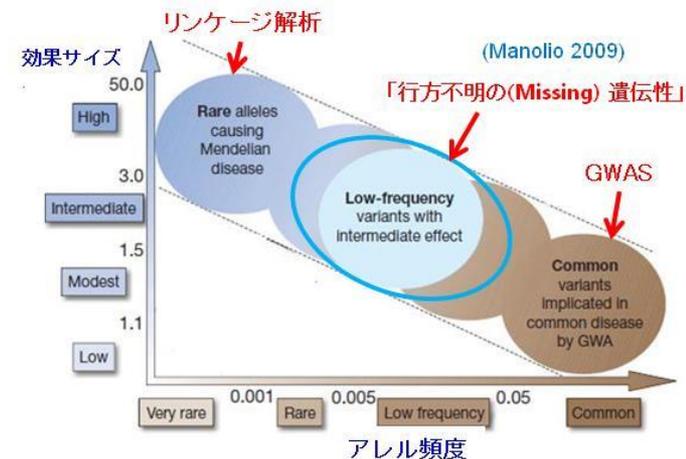
成功した臨床実装

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

しかし

多因子疾患の機序/発症予測は無着手である

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



大半の疾患の基礎としての 「遺伝素因X環境要因」の相互作用

一部の単一遺伝病を除き、大半の疾患
(Common diseases)の発症は

疾患発症の相対リスク=

遺伝要因(G:genome) X 環境要因(E:exposome)

相互作用は加算的でもなく乗算的でもない
＜(G,E) 組合せ特異的な効果＞である

GWASでSNPの相対リスクが低い
(1.1~1.3)理由: GxE組合せ特異
的效果を環境要因の全てに亘って
平均しているからである



発達プログラム説 DOHaD

Developmental Origin of Health and Disease)

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



オランダ
飢饉 (1944)

環境因子



Epigenome変化



遺伝子発現調節

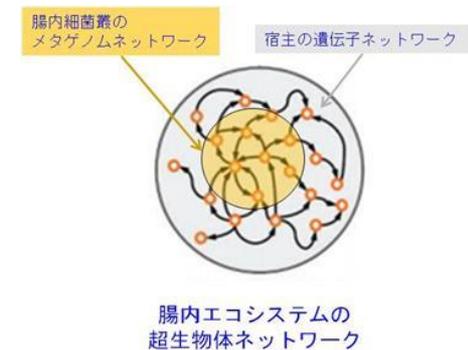
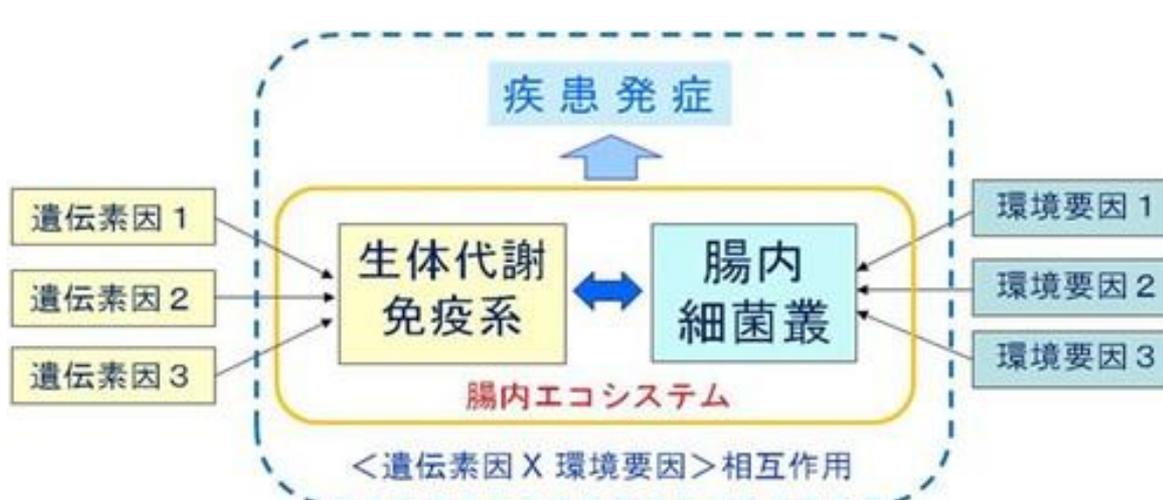


疾病発症



腸内細菌叢microbiome：メタゲノム

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



**メタゲノム
超生物ネットワーク**

第2世代のゲノム・オミックス医療

- 生涯的全体性においてその個人の疾患可能性の全体性を把握し、個別化予防、個別化治療に取り組む
製薬産業もlife-long healthcareの産業へ
- ゲノム・オミックス情報と医療・健康
- **第1世代ゲノム医療**
 - ゲノムの変異・多型性の個別性に基づく
- **第2世代のゲノム医療**
 - 多因子疾患が対象、環境情報との相互作用
 - エピゲノム機構、メタゲノム機構
 - <トランス・ゲノム機構>へ

ご清聴ありがとうございました

