

医薬品開発を革新する オミックスデータの最新の活用法

東京医科歯科大学 名誉教授（生命医療情報学）
東北大学 東北メディカル・メガバンク機構 特任教授
機構長特別補佐（情報・システム担当）

田中 博



Topics

1. 疾患ゲノム・オミックスの「ビッグデータ」と創薬・育薬

- ① ゲノム・オミックス・ビッグデータの基軸的な特徴
- ② 医療ビッグデータの時代の到来と米国の最新の状況
- ③ 「医療ビッグデータ」分析のための「データ原理」
—— Sparse ModelingとDeep Learningに共通するもの
- ④ ビッグデータ創薬・AI創薬の先進事例
- ⑤ Real-World-Dataを用いた創薬・育薬戦略の将来性
—— 創薬Rapid learning戦略は有効か

2. 計算DR（Drug Repositioning）の有効性と将来展開

- ① シグネチャ逆位法によるDRの現況と我々の研究成果
- ② 「疾患ネットワーク」と「薬剤ネットワーク」の概念
- ③ 両ネットワークを用いた双対写像関係による計算DR戦略

3. ゲノム・オミックス情報を用いたビッグデータ創薬の展望

疾患ゲノム・オミックスの 「ビッグデータ」と創薬・育薬

医療ビッグデータ時代の到来

医療「ビッグデータ」へ

- (1) 次世代シーケンサなどによる「ゲノム/オミックス医療」による情報蓄積
- (2) モバイルヘルス(mHealth) によるWearable センサ情報の継続的蓄積
- (3) Biobankによるゲノム・コホート情報



個別化医療・層別化医療、総じて
健康・医療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
 - 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
 - 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
 - システム分子医学・Precision Medicine
- 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

旧来のタイプの
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「ビッグデータ革命」

～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：10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” ($p \gg n$)

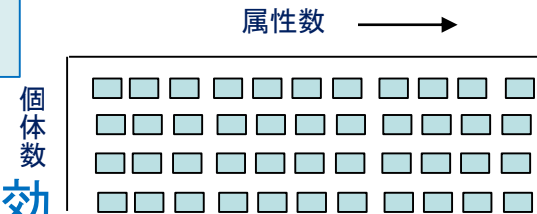
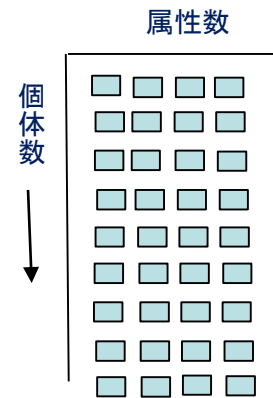
1個体に関するデータ属性種類数が膨大

属性に比べて個体数 少数:従来の統計学が無効

とくに多変量解析:GWASで単変量解析の羅列

— 目的：例えば医療の場合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”
 - 個別化医療を実現するために<個別化・層別化パターン>を網羅的に調べる：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」統計学に限定されない臨床研究
 - Real World Data: ビッグデータ知識生成（BD2K）
- 創薬の戦略パラダイムの転換
 - ビッグデータ創薬の可能性
 - 創薬・育薬のReal World Dataの利用
 - Transdisease Omics, Drug networkのDual Network Topologyによる創薬

医療ビッグデータの時代の到来と 米国の最新の状況

米国では
「新しいタイプのビッグデータ」による
医療・創薬の革命は、すでに5年の歴史がある。
まず、その革命がどの様に
始まったか見てみよう

次世代シーケンサのインパクト

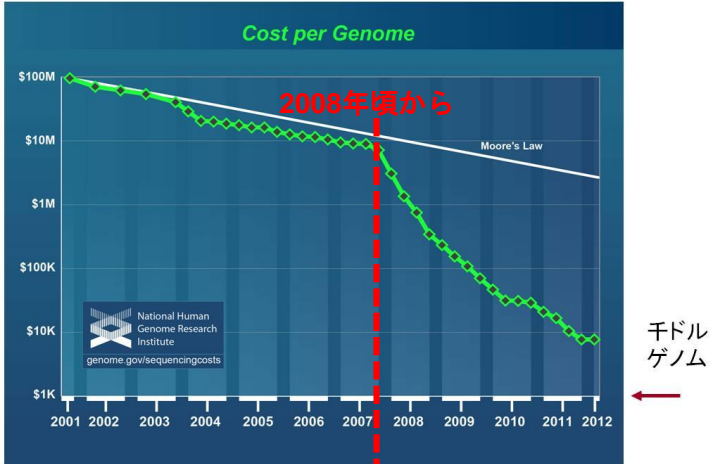
次世代シーケンサを始めとするhigh-throughput分子情報収集の急激な発展

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	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



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!

米国におけるゲノム医療の開始

第1世代の（生得的）ゲノム医療が中心
次の2つの潮流が同時に2010年に開始

(1) 原因不明先天的疾患(undiagnosed disease)

原因遺伝子の臨床の現場で(POC)の診断

次世代シーケンサの爆発的發展を受けて

Wisconsin 医科大学での全エキソーム解析

(2) 薬剤の代謝酵素の多型性の検査

臨床の現場で電子カルテの警告(診療支援)

Vanderbilt大学病院の先制ゲノム薬理

注：初めから少数の予想される遺伝子の変異を調べる候補遺伝子アプローチはすでに「遺伝子医学」で行われていた。網羅的でデータ駆動的な検査（ゲノム網羅的アプローチ）によって変異を見出す医学である

ゲノム・オミックス医療の進展とビッグ・データ

2005~ NGS登場 (454 Life sci)
2007~ シーケンス革命

2010

ゲノム医療臨床実装の開始
臨床WESの最初 (MCW)
先制PGxの最初 (VU)

- MCW Nic君原因不明腸疾患 WES XIAPの変異同定・骨髄移植
- Vanderbilt preemptive PG (PREDICT計画) 開始

Wisconsin医科大学
臨床シーケンス初例
大きなインパクト

第1世代

Early adopter
時期

Baylor医科大学
Mayo Clinicなど
後続病院多数

2013
前後

ゲノム医療の国家的取組み
NIH "BD2K" initiative 開始
各種ゲノムコンソーシアム

ビッグ
データの
概念

NIH "Big Data to Knowledge" 計画 (2012/13)
ACGM incidental finding list 56 genes (2013)
NACHGR report "Future is here" (2013)
CPIC guideline, EGAPP guideline 2013.14

第2世代

国規模の計画/全国Consortium
時期

2015

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

ゲノムオミックス医療 すでに数十の医療
施設でG/O医療が病院の日常臨床実践

NIH "BD2K" COE in Data Science, DDI (2014)
ASCO "CancerLinQ", Cancer Common
"Precision Medicine (Obama)" 1 M genomic cohort

ゲノム医療の最初の臨床実装 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(A. Mayer)
- Nicの全エキソンの配列を次世代シーケンサ決定
- MCW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



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)



Froedtert 病院

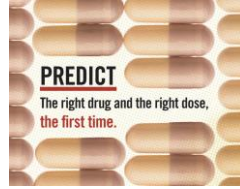


Baylor医科大学



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



Clopidogrel Poor Metabolizer Rules

Genetic testing has been performed and indicates this patient may be at risk for inadequate anti-platelet response to clopidogrel (Plavix) therapy

This patient has been tested for CYP2C19 variants, and the presence of the *2/*2 genotype has identified this patient as a **poor metabolizer** of clopidogrel. Poor metabolizers treated with clopidogrel at normal doses exhibit higher rates of stent thrombosis/other cardiovascular events.

Treatment modification is recommended if not contraindicated:

- Prescribe prasugrel (EFFIENT) 10mg daily and stop clopidogrel (PLAVIX) startdate, 10 AM

Due to increased risk of bleeding compared to clopidogrel, prasugrel should not be given to patients:

- that have a history of stroke or transient ischemic attack *** Not known; please check StarPanel
- that are greater than 75 years of age
- whose body weight is less than 60 kg

Click here for [more information](#)

If prasugrel (EFFIENT) not selected, please choose desired action:

- Increase maintenance dose of clopidogrel (PLAVIX) 150 mg daily, startdate, 10AM
- Maintain requested daily dose of clopidogrel (PLAVIX) 75 mg daily, startdate, 10AM

If not using prasugrel, please select a reason:

- Contraindicated for prasugrel
- Potential side effects
- Patient opts for clopidogrel
- Other (Specify)

Click here for [more information](#)

Cancel Order

NOTE: The Vanderbilt P&T Committee has recommended that prasugrel (if not contraindicated) should replace clopidogrel for poor metabolizers; if this is not possible consider doubling the standard dose of clopidogrel (or, use standard dose clopidogrel). However, there is not a national consensus on drug/dose guidance in this population.

Back Home Close

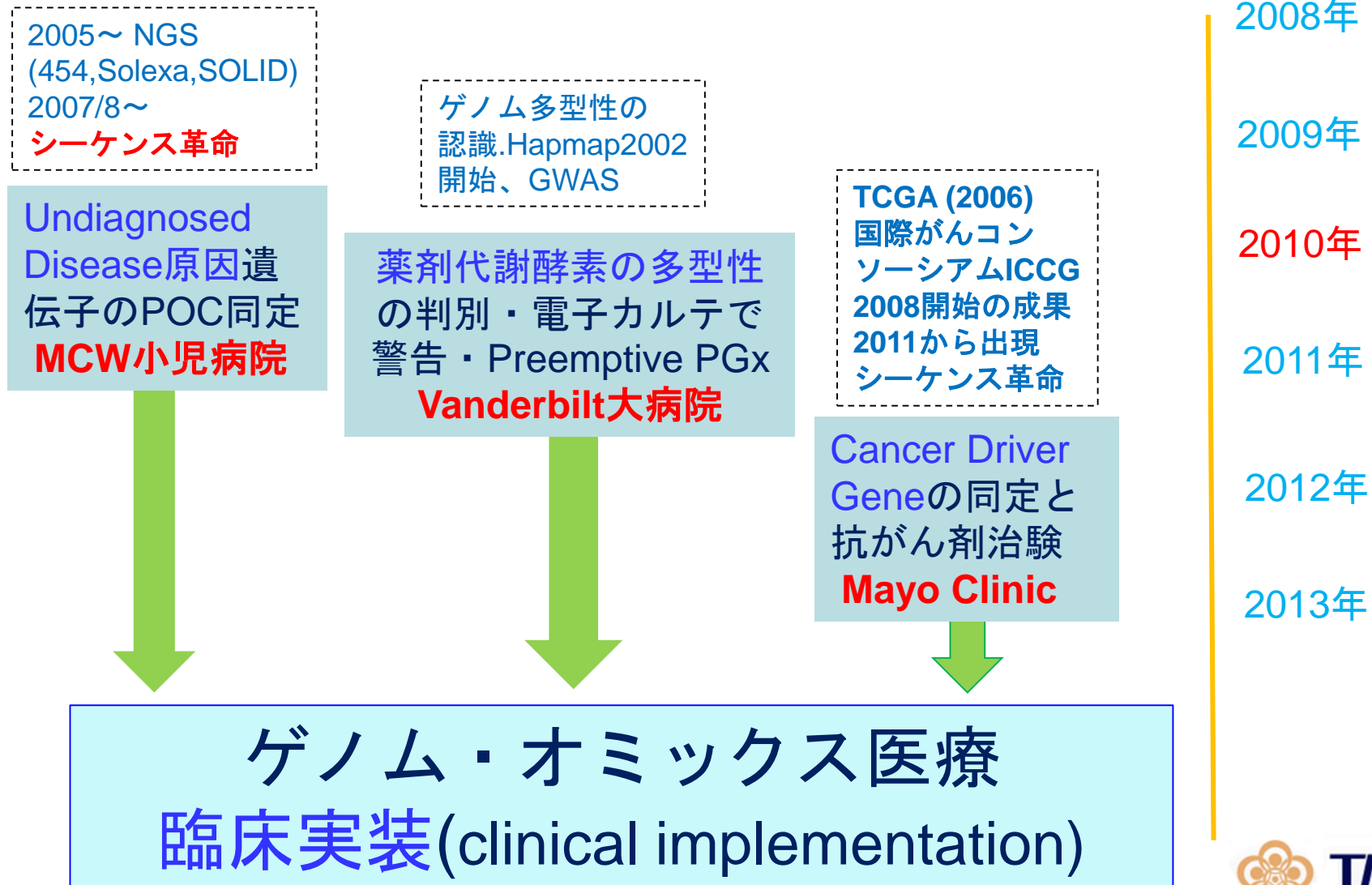
クロピドグレル処方
電子カルテの警告画面
商品名プラビックス：抗血栓剤
ステント留置手術の後に処方

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

ゲノム・オミックス医療の 3つの流れ



ゲノムオミックス医療臨床実装化の第3の流れ

著名ながんセンター Dana Faber /MD Andersonなど

第3の要素が加わる

難治性がんのドライバー変異の同定する



組織限局的な後天的ゲノム変異のクリニカル配列解析
国際がんゲノムコンソーシアム (ICGC : 2008年から)
50種のがんを500症例の全ゲノム配列解析
2012頃から成果発表と始まった(我が国も肝臓がん)
患者個人70余の変異、全集合で3000を超える変異
がんを推進させるDriver変異と偶発的なPassenger変異

Mayo Clinic

- 全患者に全ゲノム配列解析 : 10万人患者 (診療圏) データベース構築
- 先制的ゲノム薬理学 (Preemptive PGx) 検査の初期の実施
- 特別に診断する“診断オデッセイ” : Clinical Sequencing 原因不明遺伝病



ゲノム/オミックス医療－米国の状況

現状 米国ではすでに**数十の医療施設**で
ゲノム/オミックス医療が病院の日常臨床実践

NHGRI Working Groupのリスト

- Wisconsin大学病院
 - 原因不明の遺伝疾患の診断
- Vanderbilt大学病院PREDICT計画
 - 薬剤代謝酵素の多型性
- Mayo Clinicの臨床ゲノムシーケンス
 - PGx
 - がんおよび非常に稀な遺伝病原因探索
 - 10万人ゲノムDB
- その他、右表にあるように多数の病院
- 分子情報と臨床情報の融合を目的として統合データベース
 - Mofit Cancer Center (Oracle HRI)
 - 製薬会社Merkと病院の契約

Institution	Major Projects
MC Wisconsin	Using whole genome sequencing to establish diagnosis in patients with currently undiagnosed genetic disorders
Mount Sinai	<ul style="list-style-type: none"> • CYP2C19 testing for antiplatelet rx post percutaneous coronary intervention • Personalized decision support for CVD risk management incorporating genetic risk info
Northwestern	Using pharmacogenomics evidence (from GWA genotyping) to guide prescriptions in primary care and assess risk for other conditions such as HFE/hemochromatosis
Cleveland Clinic	Tumor-based screening for Lynch syndrome, endometrial cancer
UCSD	<ul style="list-style-type: none"> • Screening for actionable mutations in malignant gliomas and glioblastomas for biomarker based RCTs • Targeted rx (such as RET inhibitor) of metastatic solid tumors based on tumor mutation status
Morehouse	• Exome sequencing of 1200 early onset severe African American hypertension cases and 1200 controls
Duke	<ul style="list-style-type: none"> • Computer-based family hx collection and CDS tool with 1-yr follow-up for perceptions, attitudes, behaviors related to thrombosis and breast, ovarian, and colon cancer • SLC01B1*5 genotyping and statin adherence • Effect of genetic risk info on anxiety and adherence in T2DM

Institution	Major Projects
Alabama	Planning stages for projects in risk assessment, pharmacogenetic analysis, identification of families for further research
Baylor	Whole exome and whole genome sequencing in Mendelian disorders to improve diagnosis
Geisinger	<ul style="list-style-type: none"> • Selection for gastric bypass surgery vs other wt loss means based on genetic variants predictive of long-term benefit from surgery • IL28B variants and response to hepatitis C treatment • KRAS and BRAF mutational analysis in thyroid cancer patients
Ohio State	<ul style="list-style-type: none"> • Personalized genomic med study of CHF and HTN pts randomized to genetic counseling vs usual care • CYP2C19 testing in interventional cardiovascular procedures for clopidogrel
Harvard	Whole genome sequencing with integration in EMR and CDS; pilot of 3 patients to start
U Penn	Genotyping for assessment of MI risk in Preventive Cardiology program
St. Jude's	Pre-emptive PGx genotyping in children
Vanderbilt	Pre-emptive PGx genotyping for clopidogrel, warfarin, or high-dose simvastatin
U Maryland	Develop and apply evidence-based gene/drug guidelines that allow clinicians to translate genetic test results into actionable medication prescribing decisions
Mayo	<ul style="list-style-type: none"> • PGx driven selection/dosing of antidepressants • CYP2C19 genotyping for antiplatelet rx post PCI
Inter-Mountain	Tumor-based screening for Lynch syndrome

医療ビッグデータ時代の到来（米国）

ゲノム医療の実践

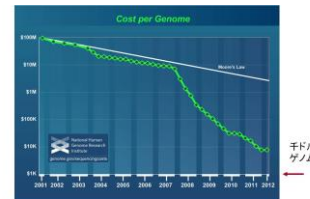
第1段階 ゲノム医療の発展

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)

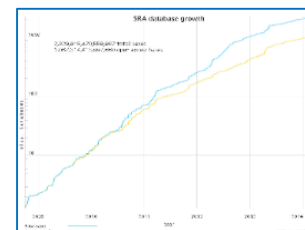
米国では数十の著名病院で実施

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

ゲノム・オミックス医療の進展とビッグ・データ

2005~ NGS登場 (454 Life sci)
2007~ シーケンス革命

2010

ゲノム医療臨床実装の開始
臨床WESの最初 (MCW)
先制PGxの最初 (VU)

- MCW Nic君原因不明腸疾患 WES XIAPの変異同定・骨髄移植
- Vanderbilt preemptive PG (PREDICT計画) 開始

Wisconsin医科大学
臨床シーケンス初例
大きなインパクト

第1世代

Early adopter
時期

Baylor医科大学
Mayo Clinicなど
後続病院多数

2013
前後

ゲノム医療の国家的取組み
NIH "BD2K" initiative 開始
各種ゲノムコンソーシアム

ビッグ
データの
概念

NIH "Big Data to Knowledge" 計画 (2012/13)
ACGM incidental finding list 56 genes (2013)
NACHGR report "Future is here" (2013)
CPIC guideline, EGAPP guideline 2013.14

第2世代

国家政策/全国Consortium
時期

2015

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

ゲノムオミックス医療 すでに数十の医療
施設でG/O医療が病院の日常臨床実践

NIH "BD2K" COE in Data Science, DDI (2014)
ASCO "CancerLinQ", Cancer Common
"Precision Medicine (Obama)" 1 M genomic cohort

臨床表現型 eMERGEプロジェクト

electronic Medical Record + Genome (NIH grand)

電子カルテからphenotyping

- **phase I (2007-2011) 臨床表現型情報のタイピング**
 - 電子カルテを通して臨床phenotypingするときの形式
 - EMR : 臨床phenotypingとbiorepositoryに基づくGWASが可能か (EMR-based GWAS)。ELSI側面も検討
 - eMERGE-I: Mayo Clinic, Vanderbilt大学, Northwestern大学など 5 施設

- **phase II (2011-2015) 臨床実装**
 - 電子カルテと遺伝情報の統合(実装)
 - 電子カルテへのゲノム情報の統合
 - PGxの臨床応用に関する試行プロジェクト
 - 結果回付 Return of Result (RoR)
 - 4施設がeMERGE-IIより加わる
 - いくつかの小児病院とMount Sinai/Gesinger

- **phase III : 2015より始まる**

- **CSER consortiumと連携**

- “Clinical Sequencing Exploratory Research” コンソーシアム
NHGRIにより予算化



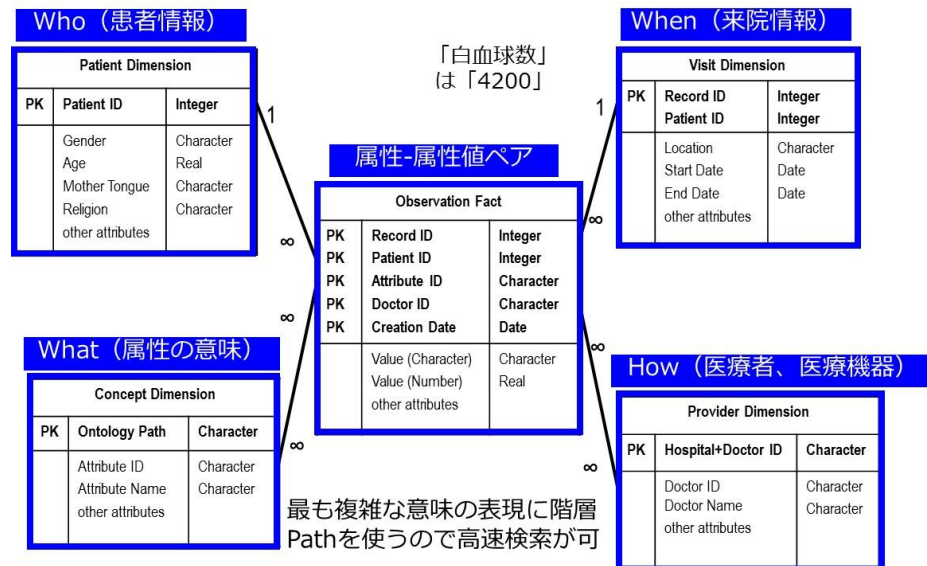
臨床データの表現型形式化（Phenotyping）の問題

i2b2 (Informatics for Integrating Biology and the Bedside)

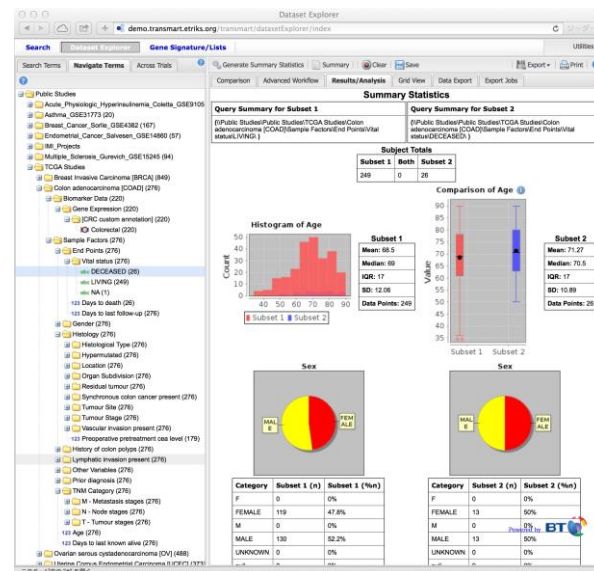
- 格納すべきあらゆる情報を主語 (subject) 述語 (predicate) 目的語 (object) のトリプレットで形式化、
- オントロジーとの組み合わせで検索可能とする、特徴的な設計
- Star Schema: データベーススキーマの1つ、その中心に位置する observation_fact テーブルに集約される。

tranSMART - トランスレーショナル生物医学研究のプラットフォーム

- tranSMART Foundationにより開発されているオープンソース(GPL3)のプラットフォーム: データマート方式
- 転帰 (outcome) などにより集団を抽出し、ヒートマップ, 相関解析, クラスタ分析, 主成分分析, 生存時間分析などの解析が可能 (IMI: Innovative Med. Initiative)



Node	Description
Biomarker Data	Measurements of biomarkers such as RBM antigens, gene expressions, antibodies and antigens in ELISA tests, and SNPs.
Clinical Data	Primary and secondary endpoints, and other measurements from the study.
Samples and Timepoints	Tested samples (such as tissue or blood) and time periods when the samples were taken.
Scheduled Visits	Periodic stages of the trial during which patients are seen.
Design Factors	Compounds involved in the study, dosages, and regularity with which the compounds were administered. Note: With clinical trials, this node is typically named Treatment Groups.
Sample Factors	Patient information, such as demographics and medical history.



国家戦略としての「医療ビッグデータ」

NIH「ビッグデータから知識へ」計画

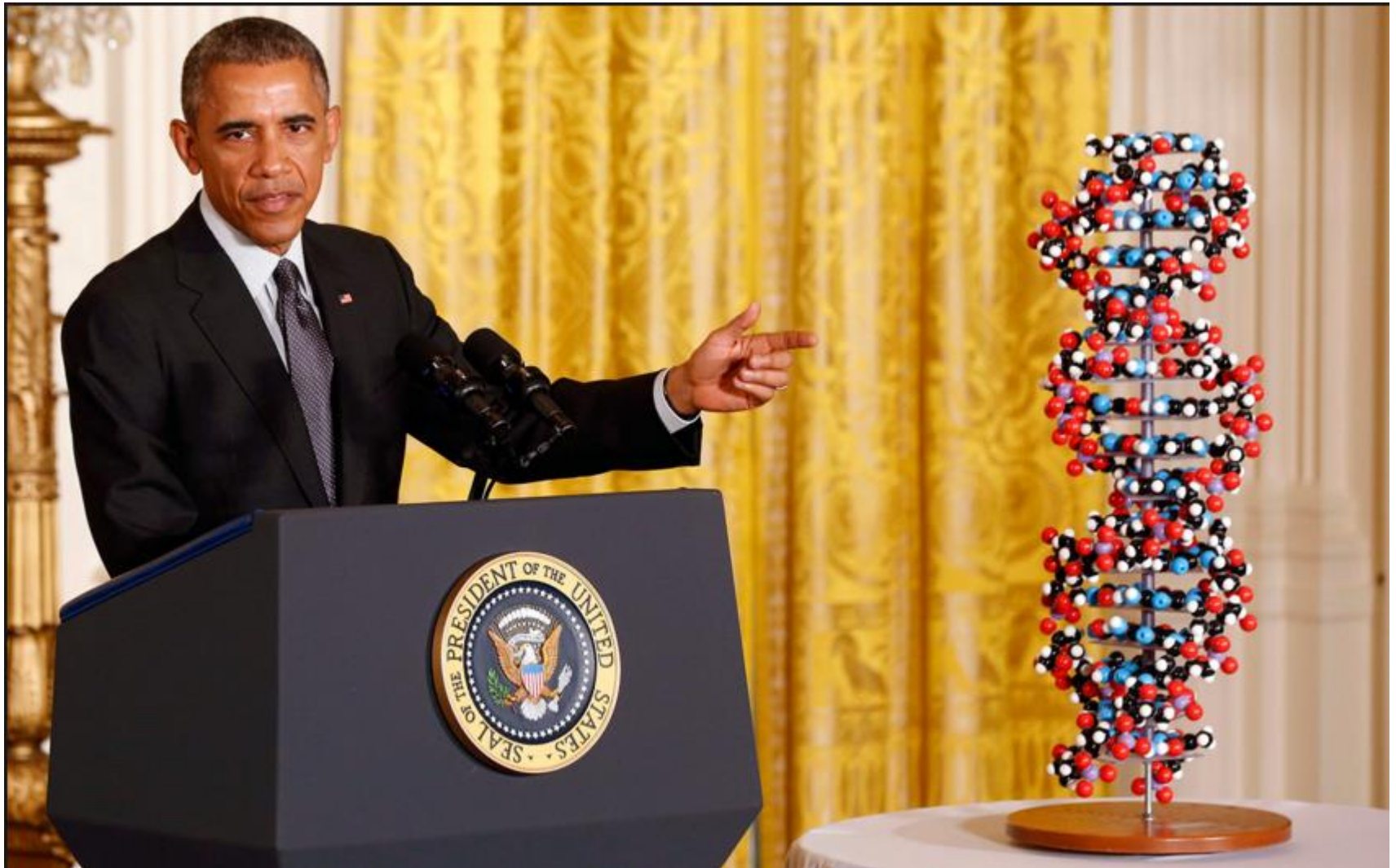
“Big Data to Knowledge” (BD2K) initiative

- **BD2K: “Big Data to Knowledge” Initiative 開始**
 - 次世代シーケンサによるゲノム・オミックス医療の普及により、臨床シーケンス情報蓄積の大量化蓄積に対応して政策開始
 - 研究費の配分**2013年**に提案。計画実施は2014年から
 - **データ科学のための副長官**（Associate Director of Data Sciences）を医療情報の世界から任命 **Bourne, PhD.**
- **Francis Collins長官談「NIH全規模での優先計画」**
 - 生命医療研究に喫緊の重要性を持つ、指数的に増大する生命医療データを活用する。
 - 「ビッグデータの時代は到来した」（Collins）
 - NIHがこの革命を作り上げる。**様々な異なったデータ種類に対するアクセスの統合・分析に主導的な役割を果たす。**

国家戦略としての「医療ビッグデータ」

- ゲノム・オミックス医療情報の全国的連携を目指して
 - 各先進病院で保持しているゲノム・オミックス医療情報の全米的な連携へ 臨床ゲノムオミックス医療DB
- NIH : BD2Kの2014年のGrandとしてのDDI (掘起し)
 - 医療におけるデータ科学の全米COE創設
 - Center of Excellence in Data Science
 - Univ. Pitts: Center for causal modeling and discovery of biomedical knowledge from big data
 - UCSC: Center for big data in translational genomics
 - Harvard: Patient-centered information commons
 - その他、コロンビア大学、イリノイ大学など11施設 32M\$
 - Data Scientist 人材養成への予算措置
 - データ発見索引 DDI (Data Discovery Index) Consortium
 - Data discovery index coordination consortium (DDICC)
 - データベースカタログの発展・Pub MEDのDB版
 - UCSD: BioCADDIEを中心にDDI開発の準備を担当
 - BioCADDIE : Biomedical and healthCAre Data Discovery and Indexing Ecosystem
- 米国はすでに戦略的に対応している。わが国は？

オバマ大統領 Precision Medicine Initiativeを開始



2015年1月 大統領一般年頭教書演説

Precision Medicine とは何か

個人の遺伝素因・環境素因に合わせた (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、目指していたのは診断/治療の個人化ではなく層別化であることを明確化

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
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を明快に強調。電子カルテの臨床表現型 (Clinical Phenome)も疾患発症後には不可欠。3つの成因の重視
- (2) 日常生理モニタリング情報の包摂
モバイルヘルス(mHealth)・wearable sensorによる大量継続情報収集の重視
- (3) ゲノムコホート・Biobankの重視
Precision Medicineを実現する基礎として、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

Obama大統領一般年頭教書 Precision Medicine Initiative



- 2015年一般年頭教書で発表
 - 精密医療、層別化医療、個別化医療
precision medicineの推進
 - 250億円 (215M\$) の予算
 - 130M\$: NIH, 100万人コホート
 - 70M\$: NCI, がんのドライバー変異
 - 10M\$: FDA, データベース開発
 - 5 M\$: ONC標準規格, 情報 privacy, security
 - 100万人のゲノムコホート研究
GxE 発症相互作用
 - mHealthの推進
- ASHG (米国人類遺伝学会, Oct 6-10)
 - Francis Collins PMIを講演
 - コホート 欧州に出遅れたが巻返す



F.Collins



Biobankとゲノムコホート

• 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
- **ゲノム/オミックス個別化医療、創薬の情報基盤**
 - **疾患型BioBank**：全国的・全世界規模で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（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。疾患ゲノムコホート
- **個別化予防の情報基盤**
 - **Population型BioBank**：「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、生涯を追跡するゲノム・コホート

• 欧米のBiobank

- **英国 UK biobank**
 - 50万人の健常者。40～69歳（2006-2010, 62Mポンド), 2011-16, 25Mポンド
 - 健診データ（血液・尿・唾液サンプル、生活情報）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。
 - 最初の対象は稀少疾患（患者・家族）、がん患者、最初はEnglandのみ
- **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**
 - 100万人のゲノムを集める

Biobank/ゲノムコホートへの期待

- 疾患型バイオバンク/ゲノムコホート
 - 個別化医療の層別化パターンの網羅的抽出
 - 病院ゲノム・オミックス医療DBを補う
- Population型（健常者）コホート
 - (1) 前向きコホートの長所により発症要因同定
疾患発症相対リスク
＝遺伝子要因 × 環境生活習慣要因
上記の**相相互作用**を評価(exposome)
「個別化予防」
 - (2) 「健康から疾患発症に至る過程」を多数収集
「**先制医療受攻状態**」(vulnerable period)同定
⇒ 先制医療薬の開発
⇒ QOL的にも医療経済的にも良策

わが国でのゲノムオミックス医療の臨床実装

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

「ゲノム医学実現推進協議会」(中間報告) 2015.7

「全国遺伝子医療部門連絡会議(10.18)」NGS臨床応用セッションに会員の中で「臨床応用を実施している部門は12施設」

アンケート結果が発表。東大病院ゲノム医学センターなど25~40%程度の原因遺伝子同定

AMED : 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。米国UDP,

英国DDD(Deciphering Developmental Disorders), カナダForge (Finding of Rare Disease Genes)

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

ゲノム医療では、米国と水を空けられている。

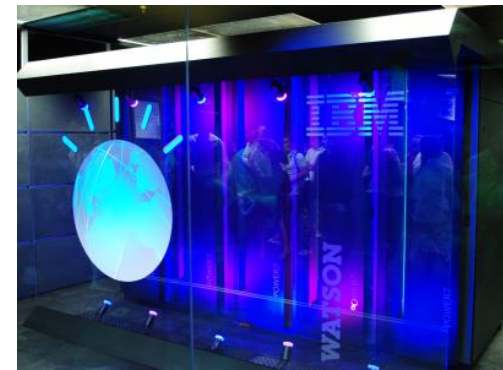
しかし、Biobank/Genomic Cohortでは我が国の状況はそれほど遅れてはいない。

医療ビッグデータ分析のための 「データ原理」

Sparse ModelingとDeep Learningに
共通するもの

ビッグデータと機械学習

- The ASCO (米国臨床癌学) **CancerLinQ** initiative
 - 診療の現場(EHR)から大量の診療データを集め分析
 - 新しい臨床治験へのガイドライン作成
 - 17万人のがん症例データベースを構築。各がんについて1～2万人の症例を集める
 - 学習システムを構築し治療知識を統計学習、ニューロネットを駆使して学習。
- **BigDataにおけるLearning systemの不可欠性**
 - 2013年に、CancerLinQのプロトタイプを完成、10万人以上の乳がんを蓄積、完全規模へ継続構築中
- IBM Watsonのがんセンターへの普及
 - Memorial Sloan-Kettering Cancer Center (MSKCC) とWatsonを母体にThe Oncology Expert Adviser software (OEA)開発
 - 他にNew York Genome Centerとglioblastoma (グリア芽細胞腫) 知識生成
- Cancer Commons initiative
 - Rapid learningのインフラ整備
 - 目的：患者の個別症例と最新の知識を更新
 - 個々の患者の”Donate Your Data”(DYD)登録
- Google X project, “Human Longevity Inc.”



IBM Watson
Learning Big
Data

「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

ビッグデータ方法論的スパース仮説

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)

ビッグデータ次元縮約

ビッグデータ解析に向けた 2つの機械学習・AI方法の適用

- **数理的知識処理** : データマイニング、探索統計学の数理的枠内で次元縮約
 - ⇒ スパースモデリング(疎性モデル) によるデータ行列の強制的「次元落ち」(L_1 正則化)
- **ニューロネットワーク** : Deep Learningによる特徴量抽出を用いた次元縮約
 - ⇒ Deep LearningのAutoEncode機能を用いた実質的な独立次元抽出に基いた解析・予測

数理的知識処理

スパースモデルによる次元落ちの正則化

従来の重回帰分析

$\mathbf{x} = (x_1, \dots, x_p)$ と目的変数 y に関して n 組のデータ $\{(y_i, \mathbf{x}_i); i = 1, \dots, n\}$

$$y_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \varepsilon_i, \quad i = 1, 2, \dots, n$$

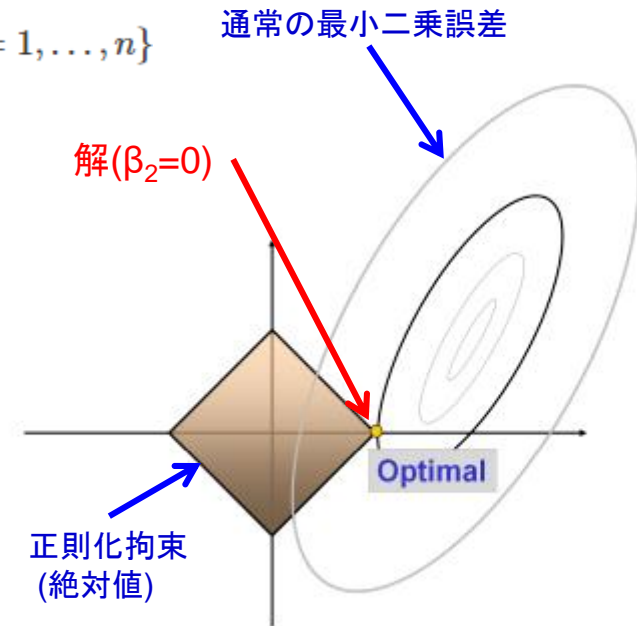
Lasso(L_1 型正則化重回帰分析)

$$\hat{\beta}^{\text{lasso}} = \arg \min_{\beta} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \beta)^2, \quad \text{subject to} \quad \sum_{j=1}^p |\beta_j| \leq t.$$

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{n} \|\mathbf{X}\beta - \mathbf{Y}\|_2^2 + \lambda_n \sum_{j=1}^p |\beta_j|.$$

通常の最小二乗

次元抑制項 (正則化項)

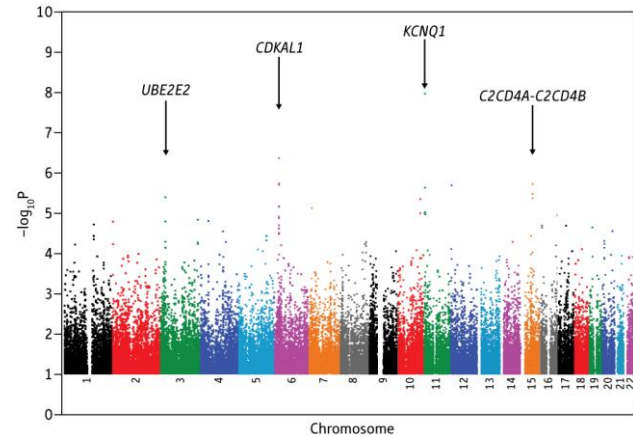
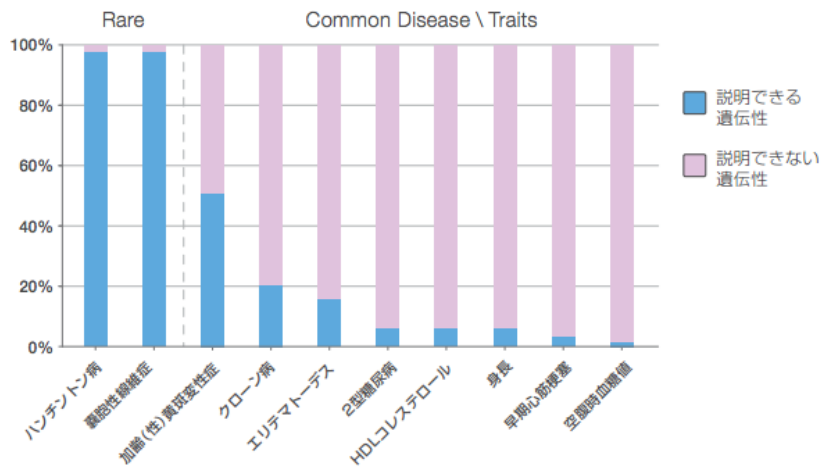


寄与の低い β_j は0になる \Rightarrow 変数選択と次元落ちが同時に達成できる

様々な変法 : Larsアルゴリズム(λ を ∞ から減少), elastic net, adaptive lasso, grouped lasso

GWAS限界へのスパースモデルの挑戦

Missing HeritabilityとGWASスパースモデル



- $\pi_{\text{explained}} = h^2_{\text{explained}} / h^2_{\text{all}} < 1$
20%~30%しか説明できず

我々の見解

Gene-gene 相互作用

細胞分子ネットワーク 効果

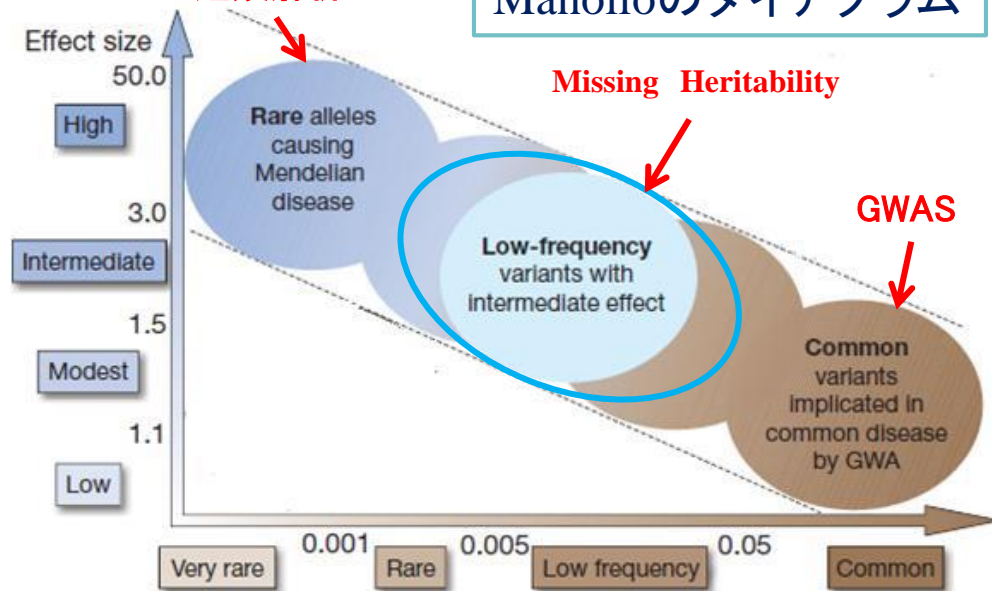
Gene-Environment 相互作用

相互作用を1項目のみで評価
他の相互作用項の効果で相殺

オッズ比

連鎖解析

Manolioのダイアグラム



アレル頻度

様々なスパースモデルの利用

- GWASへの応用

GWASにおけるgene-gene interactionの取り込み
(主効果と相互作用)

- Correlated SNPs (Ayers and Cordell, 2010)
- 検出力がさらに増加し、false-discovery rate (FDR) が低くなった (He and Lin, 2011)
- Pathwayに含まれているSNP間だけ相互作用を認める (Lu, Latourelle, 2013)

- 遺伝子発現プロフィールへの応用

- Biomarker (差別的発現遺伝子) が明確化

- 主成分分析にスパース正則化

- 主成分の解釈が容易になる

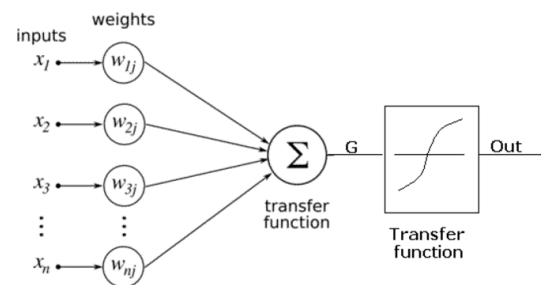
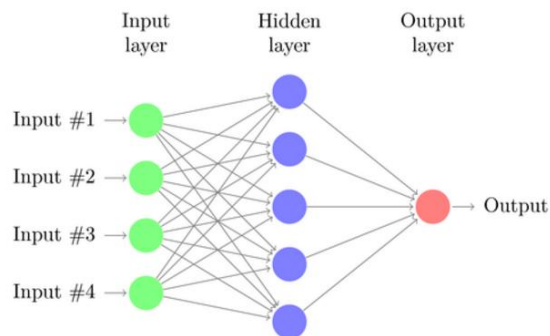
- 次を最小化

$$Q_\lambda(v_1, X) = \frac{1}{2} \text{trace}[(X - z_1 v_1^T)^T (X - z_1 v_1^T)] + \sum_{j=1}^p p_\lambda(|v_{1j}|),$$

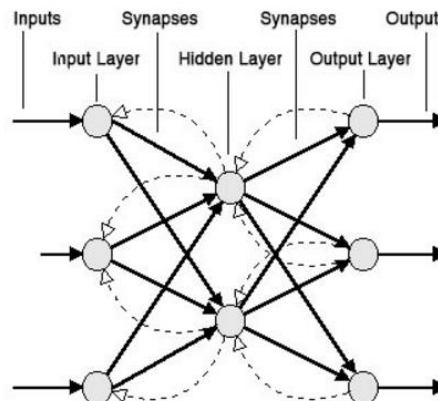
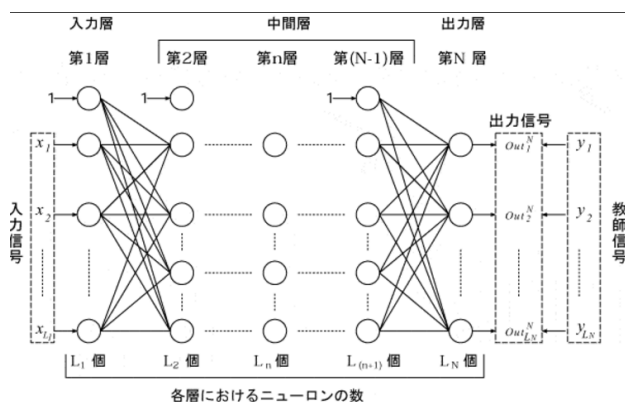
- 判別分析でも正則化により次元縮約

Deep Learningによる基軸成分の抽出

古典的Neural Network(1970年代)



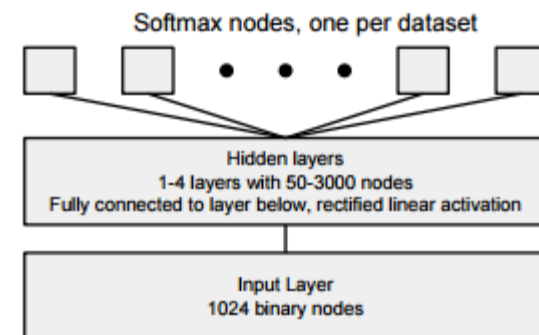
多層Neural NetworkとBack projection (1980年代)



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の変更。

Deep learning : 創薬からの注目

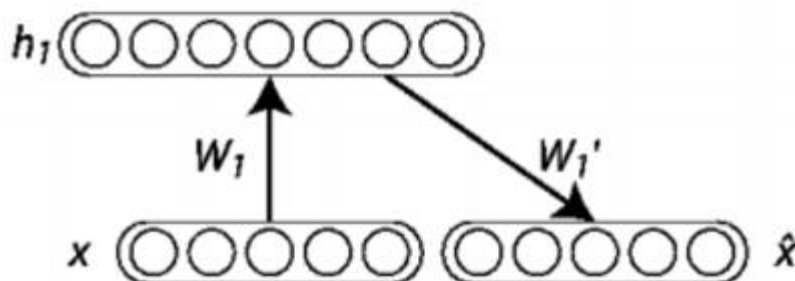
- 創薬を巡る状況
 - 平均14年、約2000億円 (\$1.7 B) の費用
 - 市場化された新薬の減少
 - 創薬に費やす期間・コストを低減したい
- Kaggle (データサイエンス競技会)にMerck社が出題
Molecular Activity Challenge (2012).
 - 15データセットから異なった分子の生物学的活動を予測するモデルの開発コンテスト
 - 勝利したモデルは深層学習 deep learning を用いたモデル
- Google in collaboration with Stanford (2015)
 - Stanford 第学の Pande 研究室と共同研究
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"



Deep learning どこが新しいか

Greedy Layer-wise Training (2006, Hinton)の提案

- (1) 最初に「教師無しデータ」を利用して、各レイヤーのパラメータを一層ずつ調整。
- (2) 最初の層を学習する場合は入力を変換し逆変換をかけ元の入力と比較し一致するようにパラメータを更新。
xを入力、1層目の変換関数をf, その逆変換の関数をg ; $g(f(x))$ を計算し、**xと $g(f(x))$** が一致するようにパラメータを学習する。
パラメータが十分な数があれば元の入力をそのまま返すような関数が学習される(恒等写像)が、パラメータに**正則化**をかけて学習することにより、少ない表現力で入力の情報を表現するようにパラメータが調整される。入力情報を最も良く表現できるような関数が抽出。
基本的な特徴情報が取得される。
- (3) **autoencoder** : 変換をかけて元の信号に戻せるように学習する方法
- (4) 第一層の結合係数は**固定して**次の階層の学習に入る
- (5) 最後の層が学習できれば、最後は逆伝播で微調整する



$$\text{input: } X \quad \text{code: } h = W^T X$$

$$\text{loss: } L(X; W) = \|W h - X\|^2 + \lambda \sum_j |h_j|$$

正則化項

Artificial Intelligence-based (AI)創薬

- 標的分子選択と妥当性検証
 - 適切な分子標的の選択
- Virtual screening と選択 ←
 - 適切な化合物に対するクラス判定
 - 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-近隣法, logistic回帰など優位
 - Deep Learningで構造活性相関を学習する
 - 特徴量の抽出、薬理機構への
 - 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)

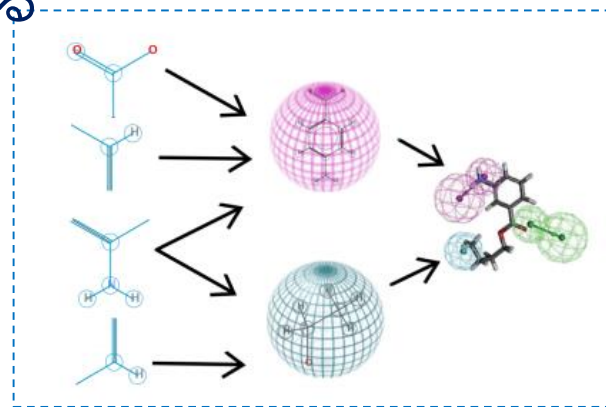
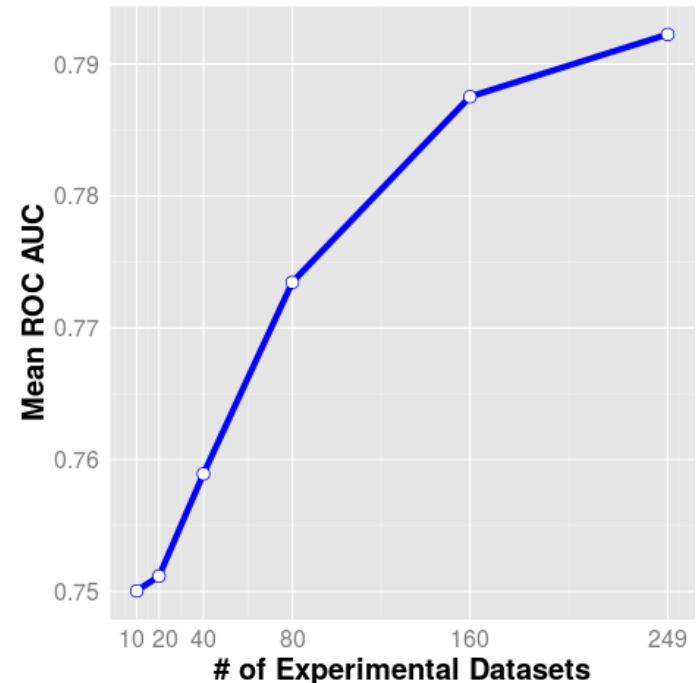
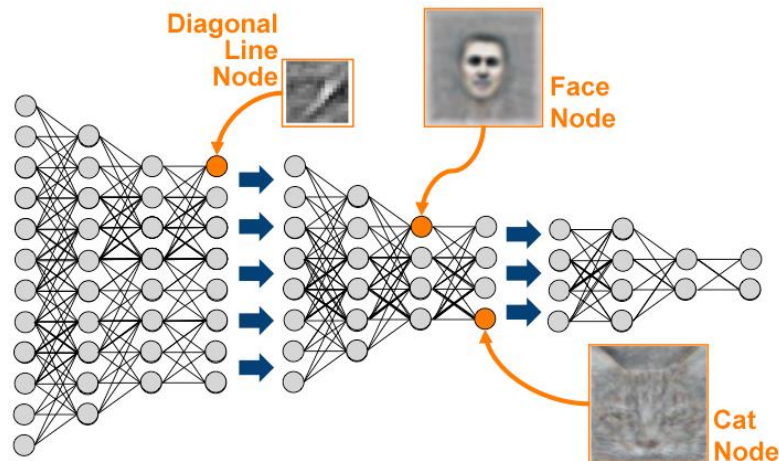


Figure . Hierarchical nature of fingerprint features: by combining the ECFP features we can build reactive centers. By pooling specific reactive centers together we obtain a pharmacophore that encodes a specific pharmacological effect.

スパース仮説

- スパースモデルもDeep Learningも次元縮約的な特徴量を探るビッグデータ解析のための仮説的原理に基いている。
- ただし従来型の次元縮約とは違って、縮約次元は大きく標本数 n に影響される（縮約次元は n の関数）データを多数集めれば集めるほど認識が深まる。
- いつまでのOpen性(open-endedness)がある。



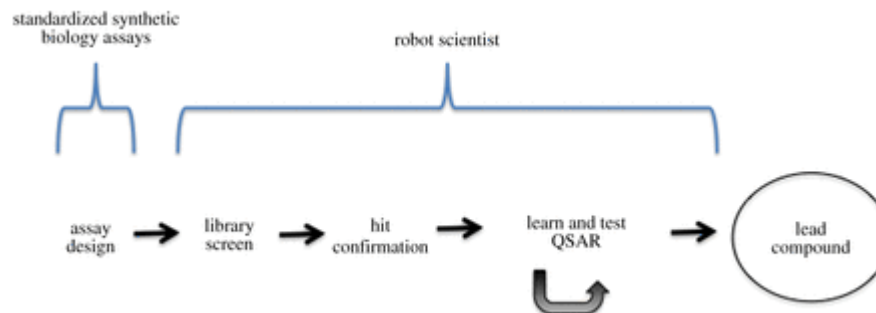
データ数が増加すると内部的な縮約次元が増加し推定精度が上昇する

そのほかのAI創薬の話題

- 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（詳細不明）
- 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



Real-World-Dataを用いた 創薬育薬の戦略の将来性

—RCT, EBMからの呪縛の解放—

「学習する医療システム」 Learning Health System

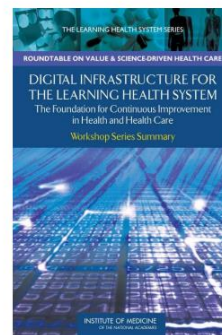
新しい生物医学知識が臨床実践に給されるまで17年
臨床データを用いて医療を実施しながら医療を改善

- 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

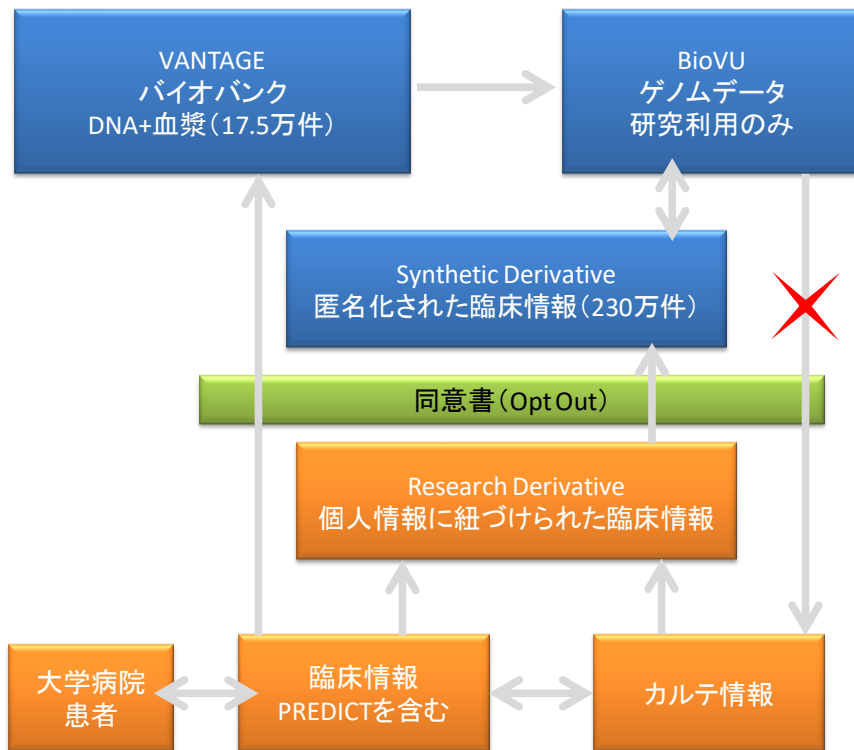
*Digital Infrastructure for the
Learning Health System: The
Foundation for Continuous
Improvement in Health and
Health Care*

*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。

個別化（層別化）医療の概念の普及とRCTの限界

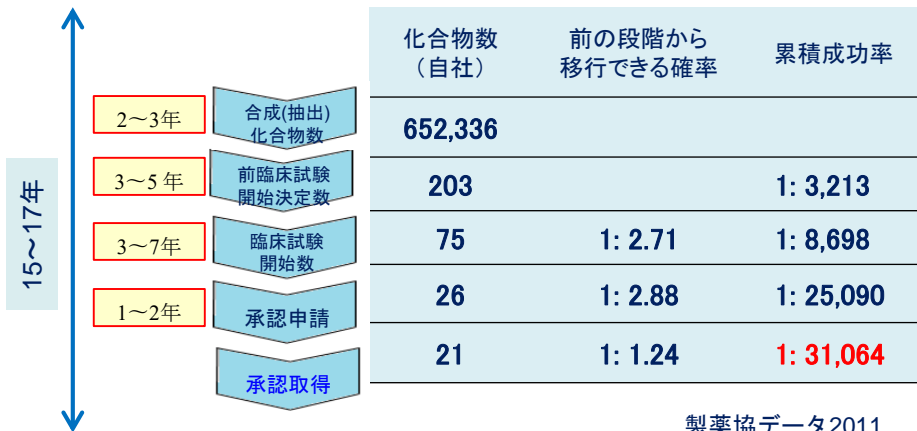
- 個別化・層別化の概念の浸透
- RCTの治験集団とReal World Dataの乖離
 - すべての個別化パターンを包摂した治験集団は現実には不可能
 - 現在の治験集団
 - 大半のRCTは医療現実の外の「人工的な環境」
 - 高齢者・妊婦はいない、欧米では黒人とくに青年は含まれない
- 将来へ向けたプラットフォームの確立
 - 母集団に近いReal World 医療データが収集可能
⇒データの大規模化の「相転移」
 - Real World Data時代のプラットフォームを形成。
⇒製造後第3相試験でReal World Dataを使うか（抗がん剤の例）
 - 我が国の戦略 段階的移行 BioBankの利用段階



計算DR(Drug Repositioning)の 有効性と将来展開

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

オミックス創薬の原理

- 薬剤特異的遺伝子発現 (Drug-induced SDE)

- CMAP : Connectivity Map

- 薬剤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集合)
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
- 最近LINCS 100万サンプルへ拡張

- 疾病特異的遺伝子発現

(Disease-associated SDE)

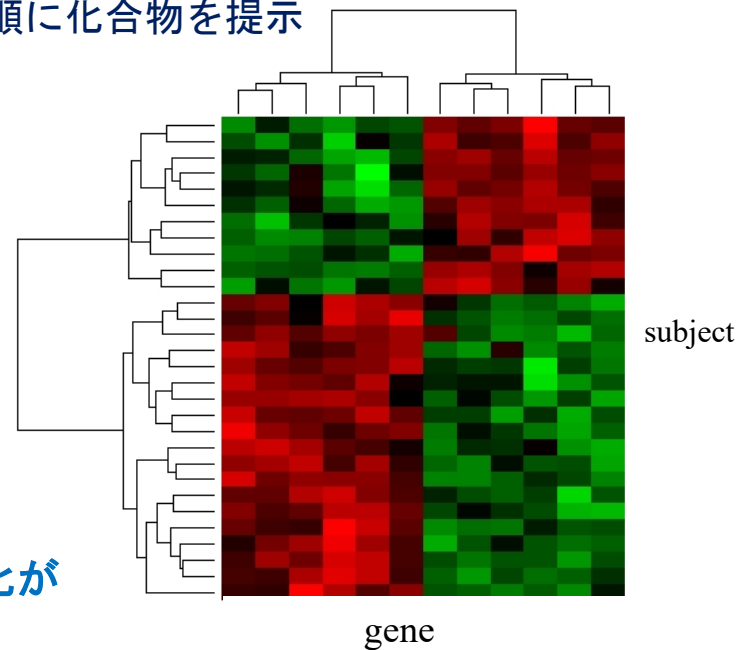
- GEO (gene expression omnibus),

- 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

本来は、分子ネットワークの疾病/薬剤特異的变化が基本 (第3世代網羅的医学)。

遺伝子発現プロファイル変化

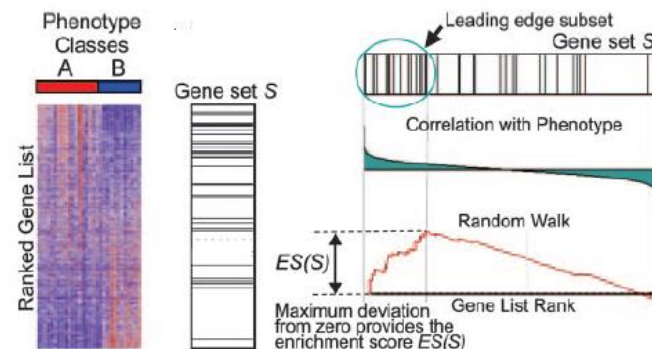
≈ 分子ネットワーク変化



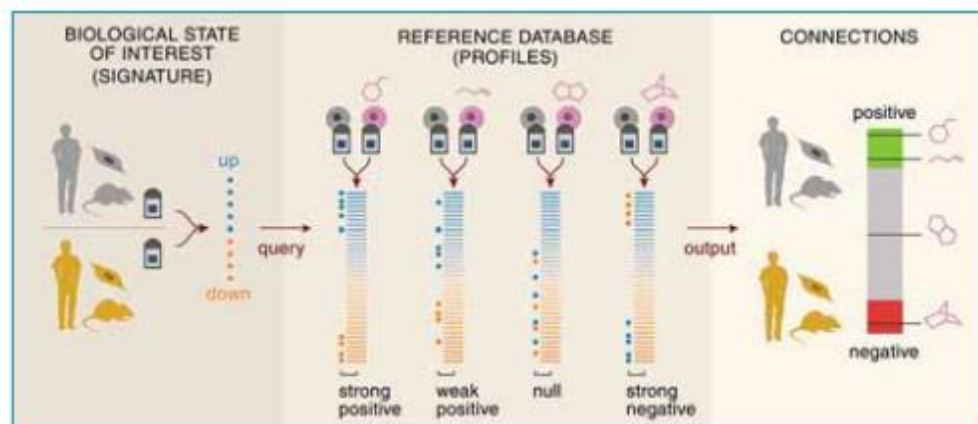
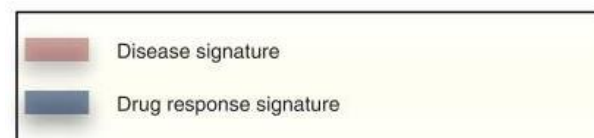
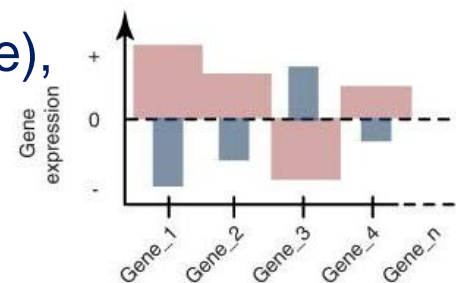
遺伝子発現プロファイルによる有効性予測

- 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) : ES score
 - 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

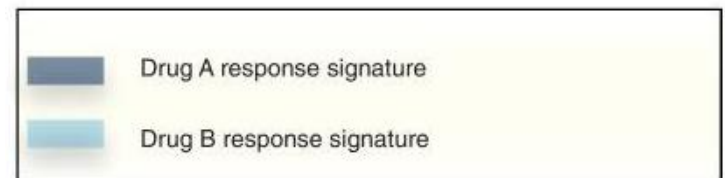
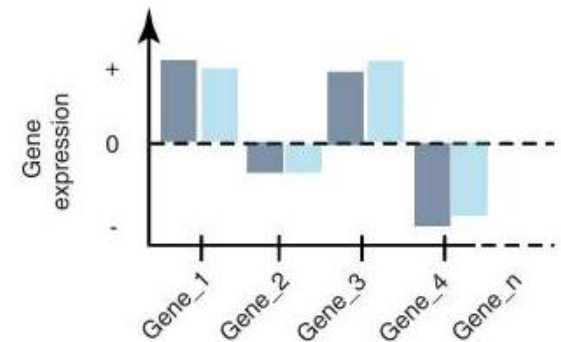


GSEA



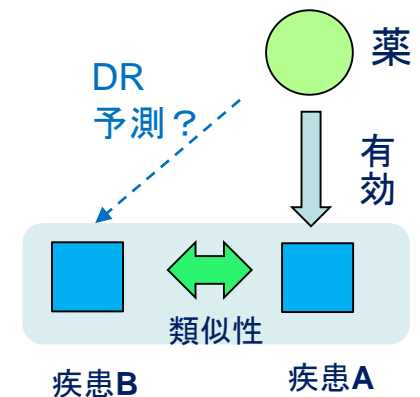
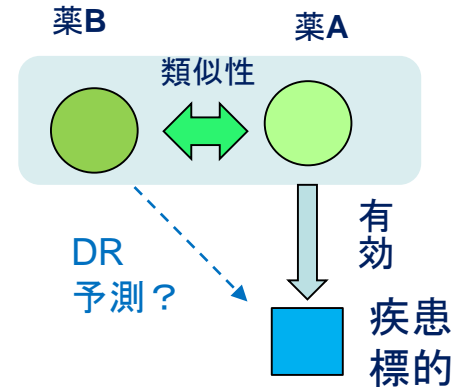
遺伝子発現プロファイルによる毒性予測

- 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



合理的DRへのアプローチ

- 医薬品中心 Drug-based (drug-centric)
 - 医薬品の構造・特徴の類似性に基づいて別の医薬品の適応を予測
 - ① 化合物の化学的構造・特徴の類似性
 - ② 薬物投与時の遺伝子発現プロファイル
- 疾患中心 Disease-based(disease-centric)
 - 疾患の発症機序の類似性に同一の医薬品が別の疾患の適応を予測
 - ① 疾患原因/感受性遺伝子の共有
 - ② 疾病遺伝子発現プロファイル
 - ③ 疾患を起こす分子ネットワークの類似性
- 両者の融合的アプローチ



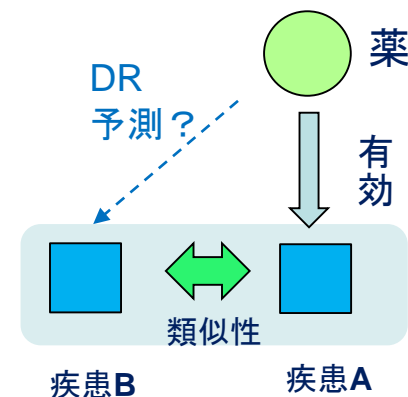
疾患ネットワークに基づいたDR

第1段階 疾患ネットワークの構築

- 疾患のゲノム・オミックス機序の類似性をネットワークで表したものの

第2段階 DR候補予測

- 既存の<疾患-薬剤>を決め
近傍の疾患をDR候補とする

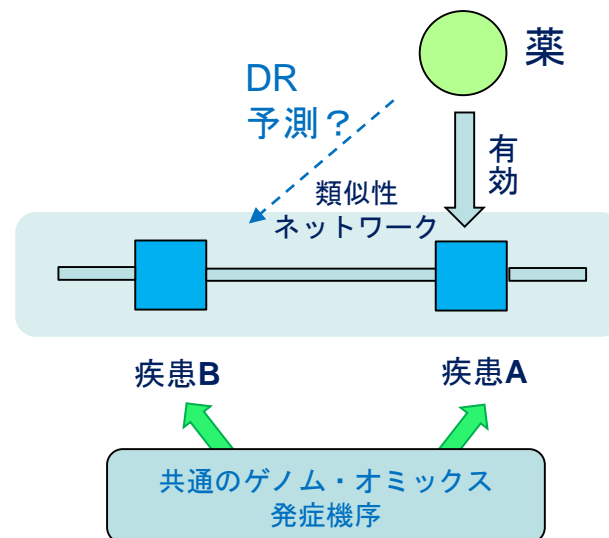


第3段階 有効性・毒性の確認

- シグネチャア逆位法、連座法で有効性、毒性を予測

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



疾患の成立機序における主要機序

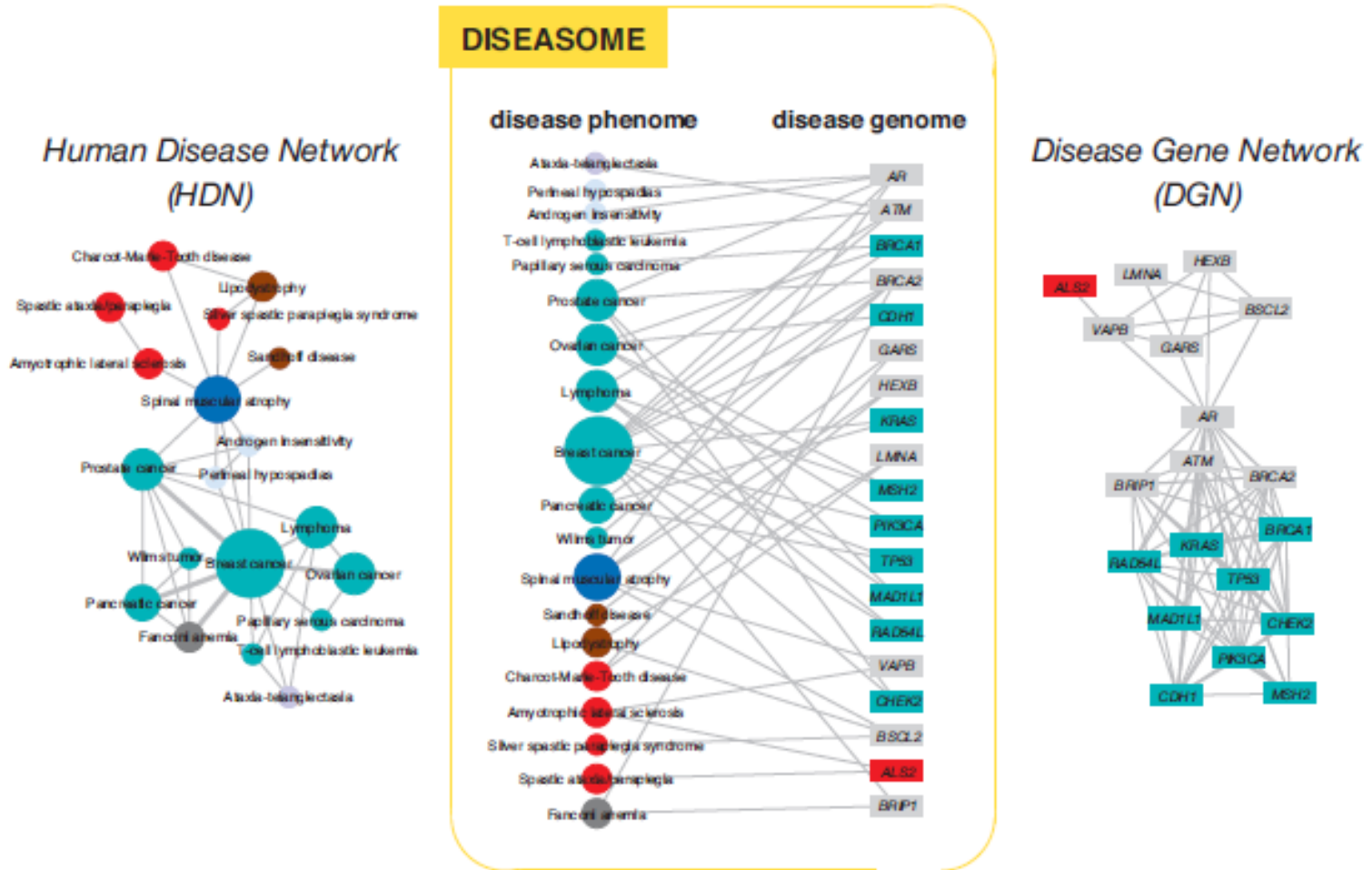
- 疾患関連遺伝子型（第一世代型）
 - 原因遺伝子、疾患感受性遺伝子の変異・多型が主要発症機序
- 疾患オミックス型（第2世代型）
 - 疾患オミックスプロファイルの変容が主要発症機序
 - Transdisease omics
- 疾患分子ネットワーク型（第3世代型）
 - 分子ネットワークの歪みが主要発症機序
 - がんなどで遺伝子型（肺腺がん等）でない通常のがん

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

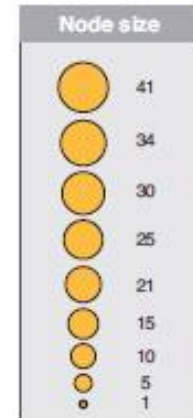
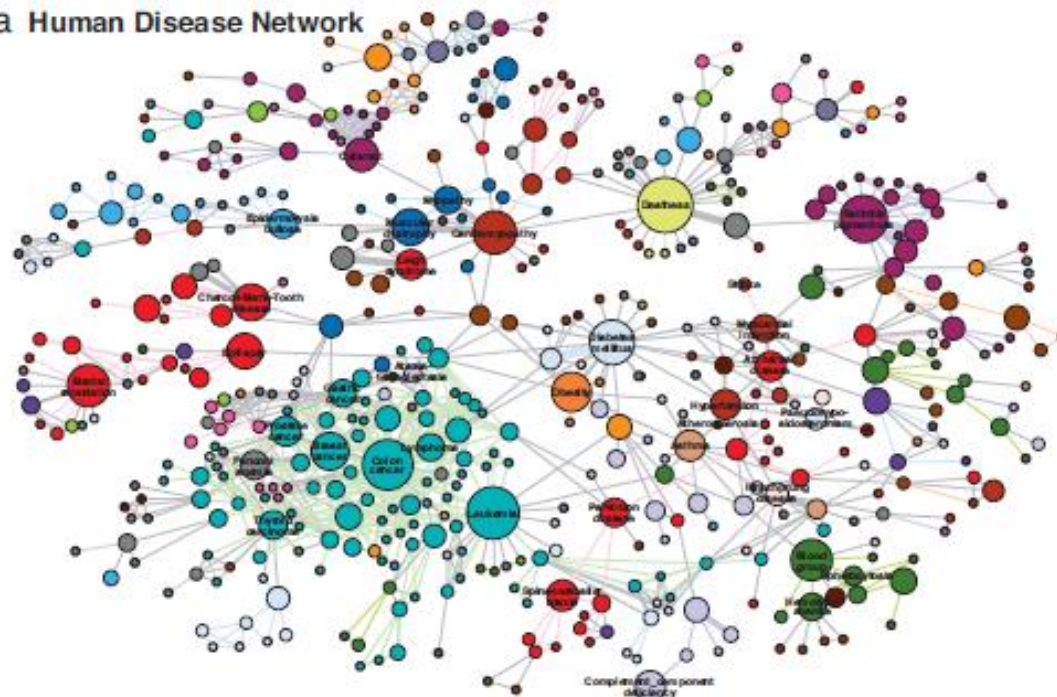
Kwang-Il Goh*, Michael E. Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-Laszlo Barabasi The human disease network PNAS2007

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

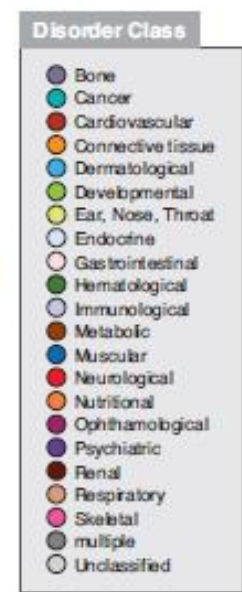
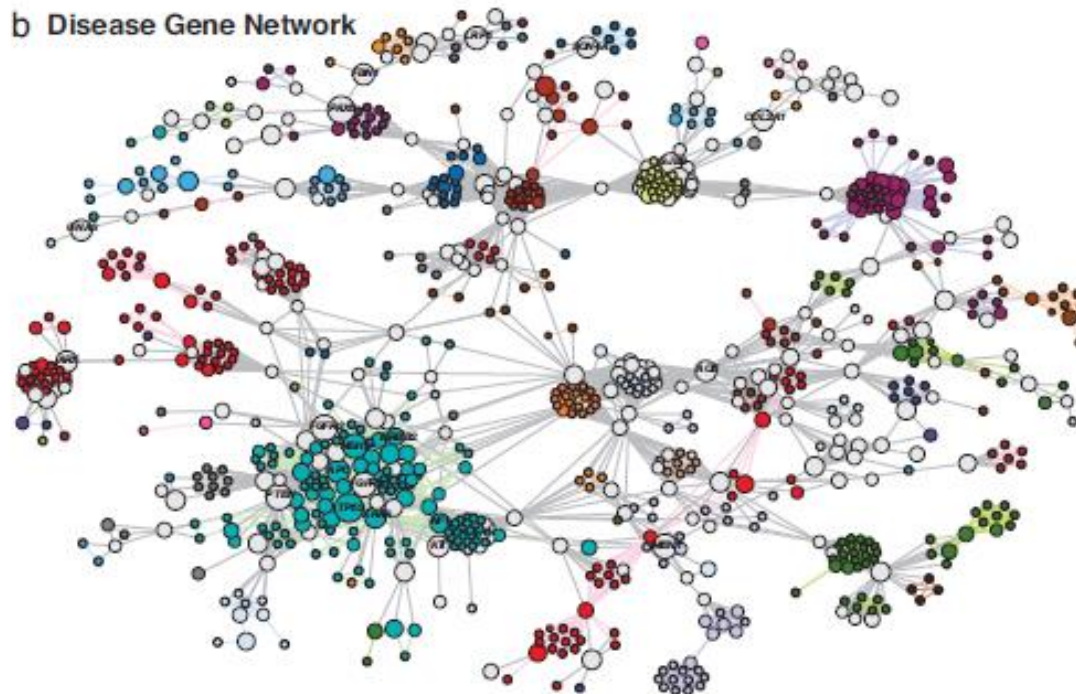


疾患遺伝子 ネットワーク (DGN)

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



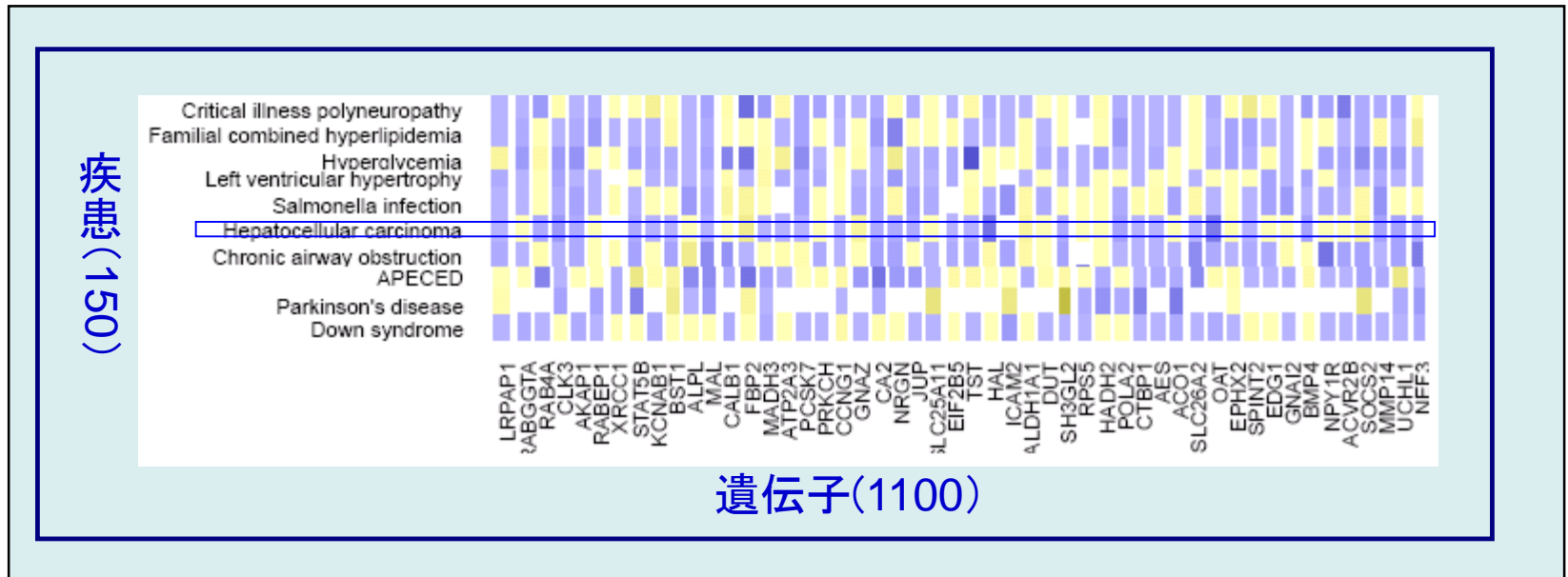
Diseasomeを巡る状況

- Mendel疾患から複雑疾患、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、代謝ネットワークの利用
 - 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)
 - パスウェイ準拠型の疾患ネットも
- **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、臨床表現型ネットワークと分子ネットワークを繋げる機構**
 - 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

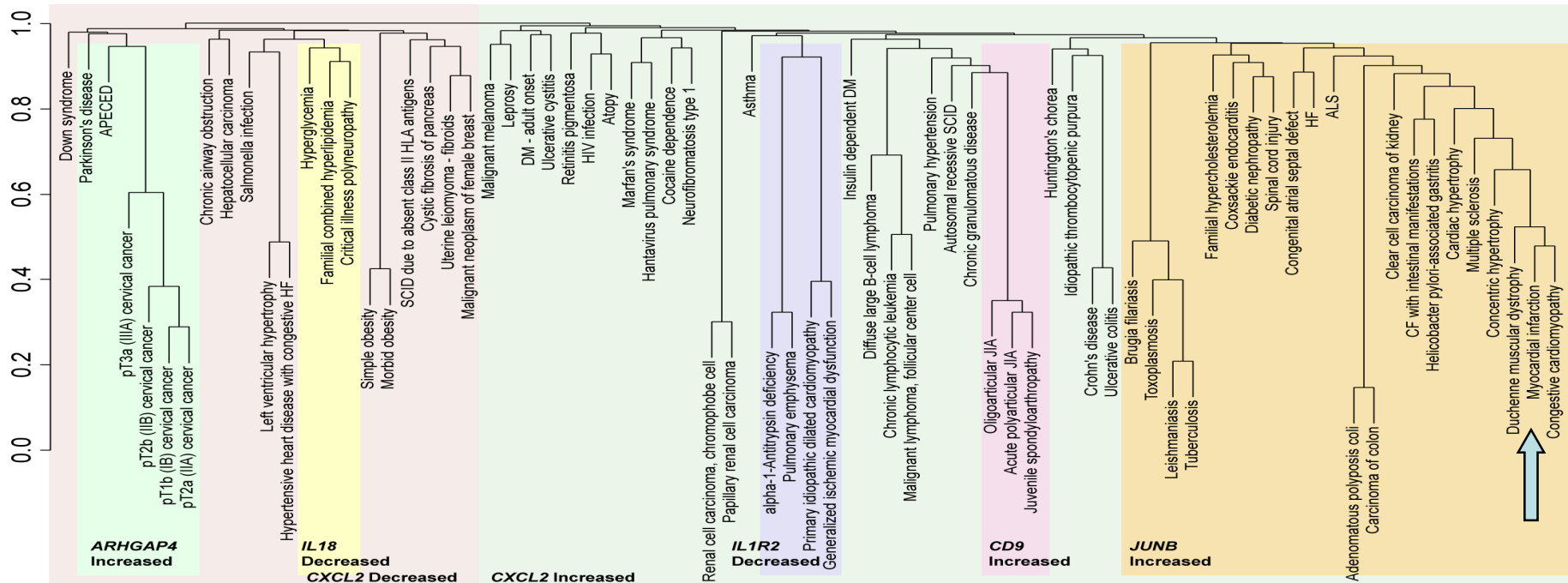
GENOMED (A. Butte et al)

- 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



Transcriptional Profiling による疾患ネットワーク

Hu, Agarwal 遺伝子発現プロファイルとGSEA関連尺度によるリンク

疾患 (disease-disease) 645 nodes
 疾患-薬 (disease-drug) 5008 pairs

Solar keratosis 日光性角化症

⇒ cancer(squamous)

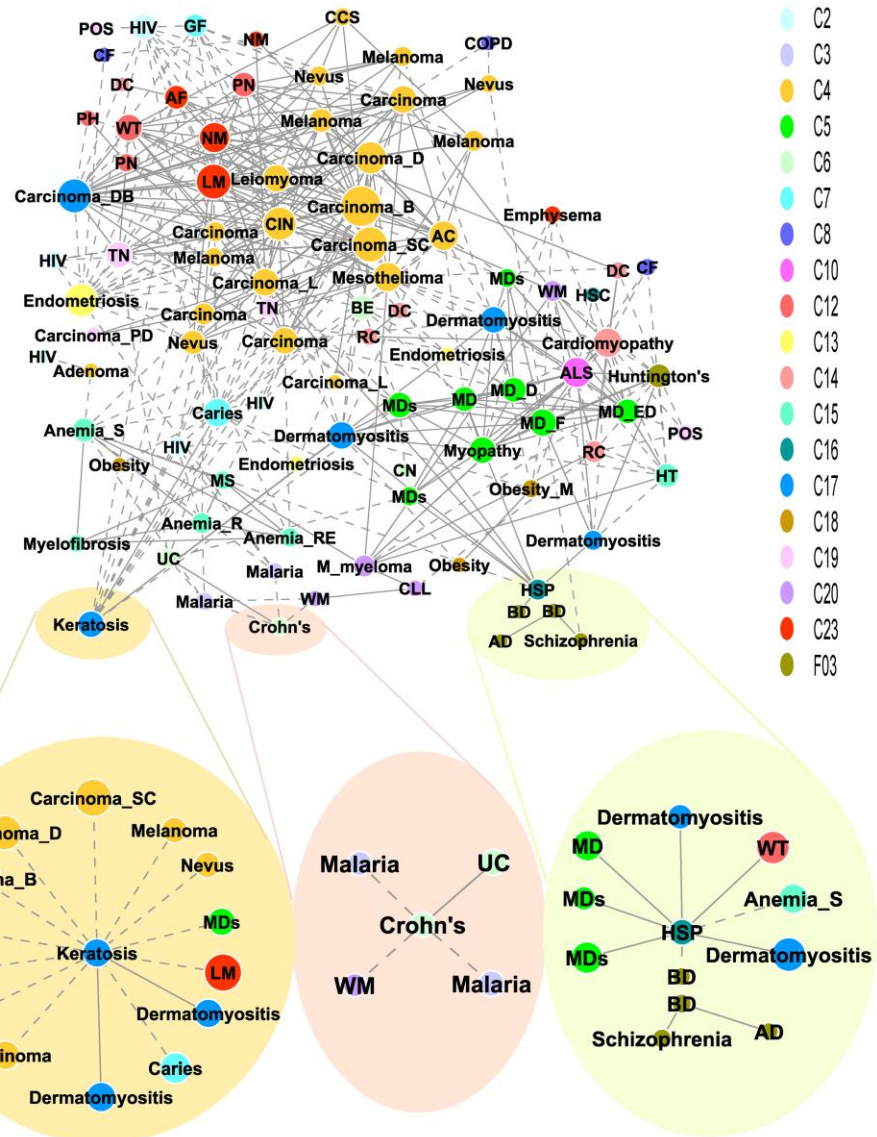
Crohn's disease

⇒ マラリア

Hereditary Spastic Paraplegia

(遺伝性痙攣性対麻痺)

⇒bipolar双極性うつ病



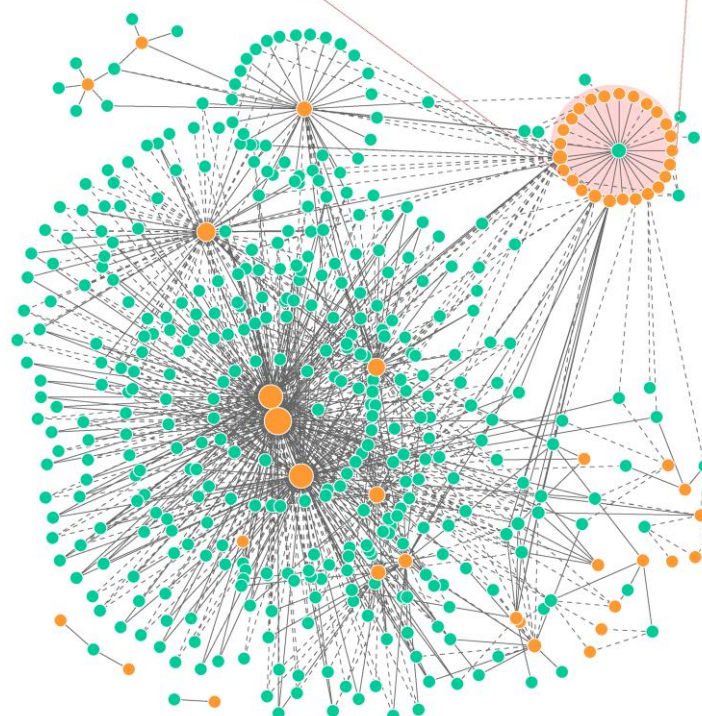
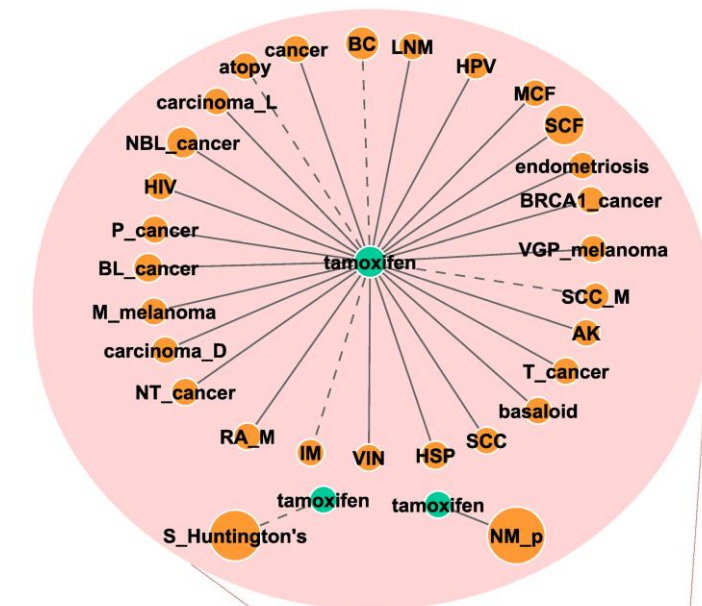
カラーはMeSH
 同一カテゴリー

Transcriptional Profiling による 疾患ネットワーク

Disease-drug network

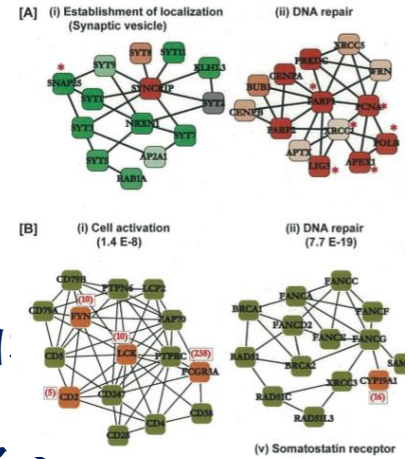
橙色 医薬品
緑色 病気

Tamoxifen (breast cancer)
(負の値をもっている)
⇒ アトピー,
⇒ マスト細胞分泌抑制、
アレルギーマジック抑制
(正の値をもっている)
副作用の予測
⇒ carcinogenic

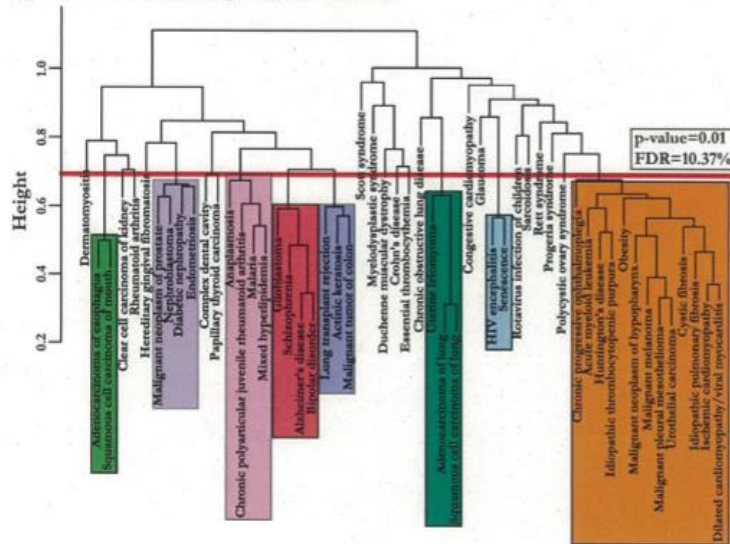


Transcriptomeの変化をPPIに投影した疾患ネットワーク (Butte)

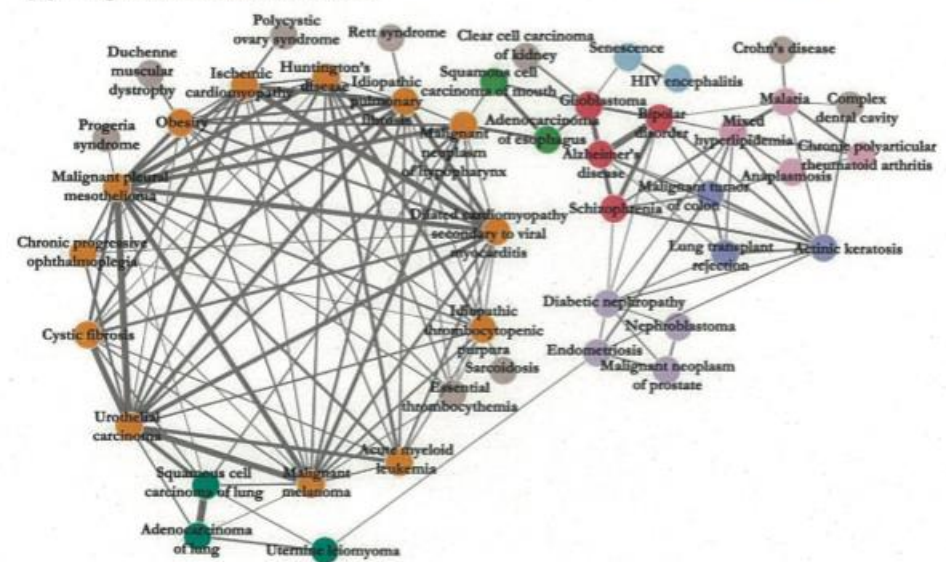
- 遺伝子発現プロファイルを直接使うのではなく 4620Moduleに分解したPPIネットワークでの疾患での平均発現変化をつかう
 - 遺伝子発現プロファイルより疾病によって変化するmoduleを調べる 病気に対するPPIの応答 マラリアとクローン病
 - moduleの遺伝子の変化を平均して遺伝子の代わりに moduleの発現平均スコアを用いる
- 疾患の大半を占める <共通疾患状態シグネチャ>



[A] Hierarchical relationships between diseases



[B] All significant disease correlations



最近の話題 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）を与えて、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では
 - 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、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（約22000個の遺伝子発現）をGEOから作ったモデルで推定する
 - 相互依存性高い⇒1000遺伝子に情報が含まれているので推定可能
- **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
- **ヤンセン(Janssen Pharmaceuticals) がL1000™ Expression Profiling技術を契約**
 - 25万種類の化合物について、この技術を供給するGenometry 社と契約

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-hydroxyphenyl(phenyl)benzenesulfonamide
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-tetrahydrophthalazine-5-yl)carbamoyl)phenol
0.4800	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

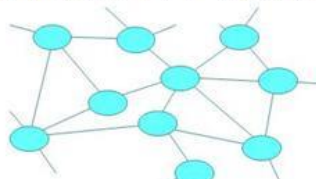
Select a batch: LJP004

Multiple Selections:

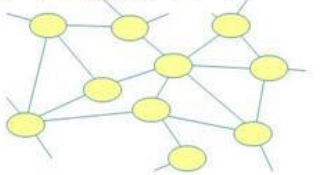
疾患ネットワークとDrug Projection Map

DR informaticsの構築

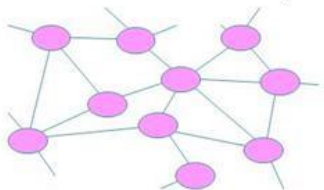
第1世代疾患ネット (原因遺伝子親近性)



第2世代疾患ネット (OmicsProfile親近性)

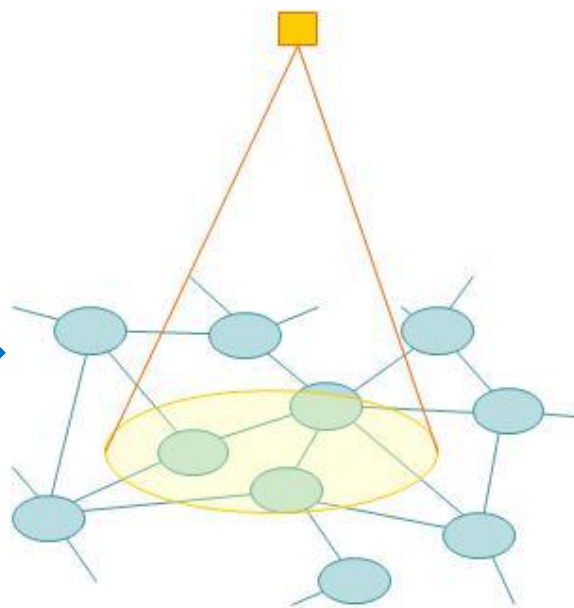


第3世代疾患ネット (Pathway親近性)



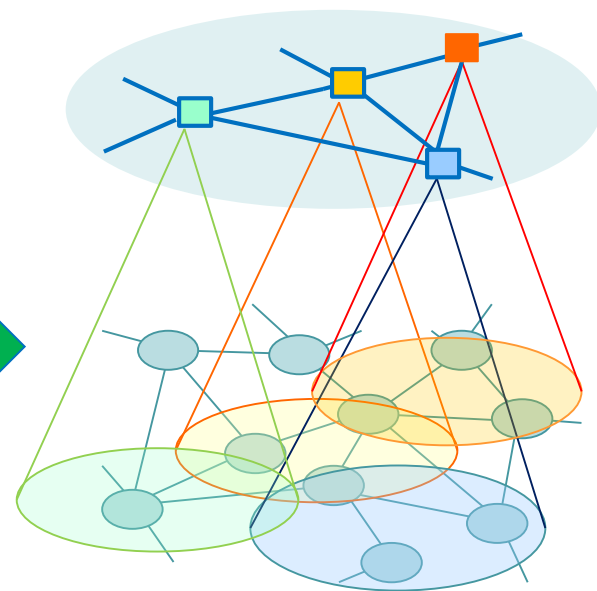
統合化

薬剤



疾患ネットワーク

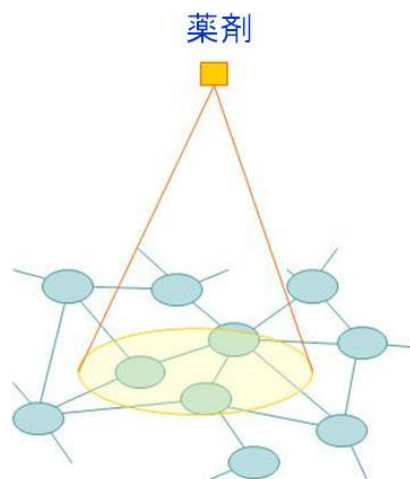
薬剤ネットワーク



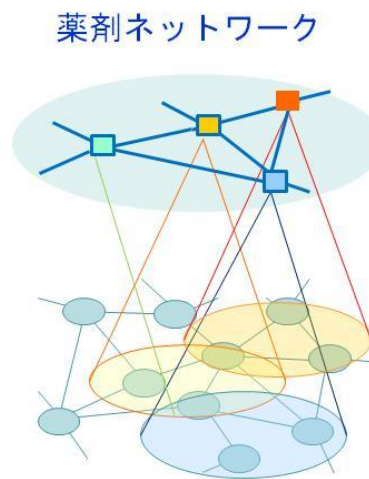
疾患ネットワーク

薬剤（化合物）ネットワーク

- 化合物空間(Chemical compound space: CCS)
 - 化合物空間ネットワーク
 - 類似性準拠(similarity-based compound network)
 - Magigiora (2014), Saito(2011)
 - 生物学的に適切な化合物部分空間
 - 薬剤特異的遺伝子発現の類似ネットワーク
 - LINCS準拠の薬剤添加時の遺伝子発現の類似性
- 薬剤・化合物空間ネットワーク



疾患ネットワーク

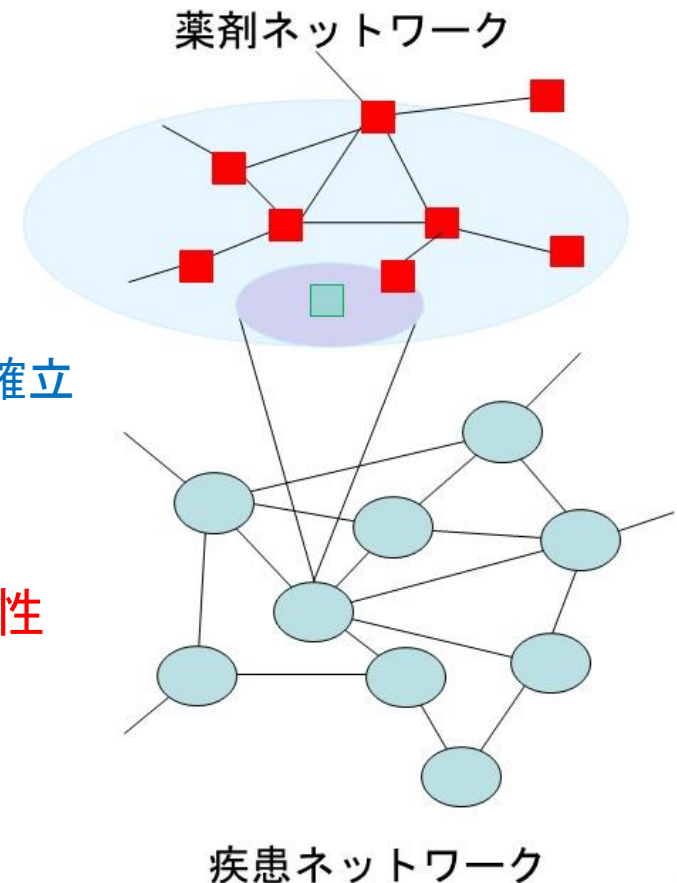


疾患ネットワーク

Drug SpaceとDisease Spaceの双対性

疾病から薬剤ネットワークへの逆投影 Double Topology 双対写像による創薬方法論

- 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)を利用
 - 疾患ネットワークの確立が重要
 - 疾患空間から逆投影。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



ゲノム・オミックス情報を用いた ビッグデータ創薬の展望

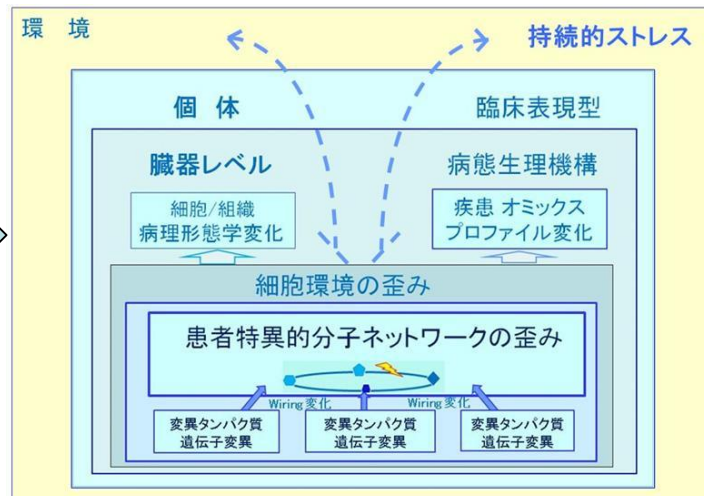
網羅的分子医学 第3世代 システム分子医学 (2010~)

システム生物学の疾患への応用
「疾患をシステムとして理解する」

単因子性の疾患を除いて、大半の疾患は1個や2個の遺伝子の変異ではなく多数の遺伝子の変異やタンパク質の機能異常による分子パスウェイ/ネットワークの調節機能不全や歪み
distortion (dysregulation) of molecular network

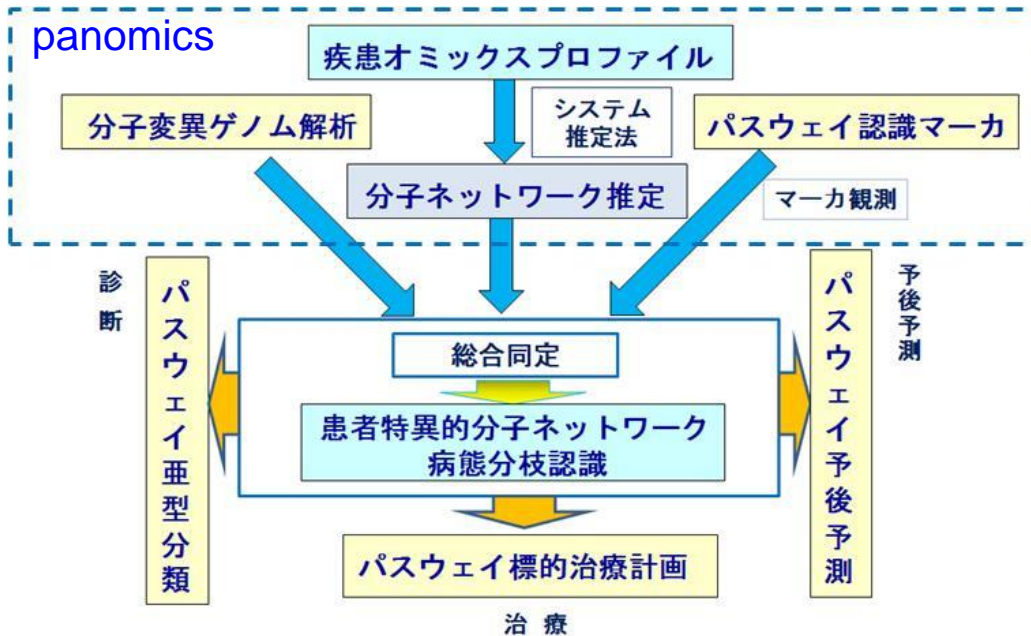
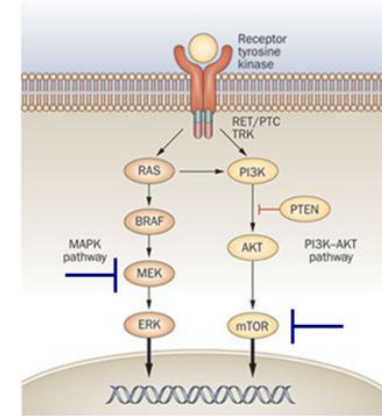
個別化医療・予測医療・先制医療・創薬のための疾患のシステムの理解に基づく医療

疾患オミックス
(網羅的分子表現型
molecular phenome)
成り立たせる基底としての
「細胞分子ネットワークの
構造変化」
疾病の理解における
第3のパラダイム



個別化医療の展開

- 個別化・分子標的創薬
 - 疾患（がん）を分子変異で層別化
 - バイオマーカ（分子変異 genomic biomarker)
 - 基本概念 **Oncogene addiction**
- システム分子医学的疾患認識へ（分子システム）
 - 「がんはパスウェイの病気である」
 - **Pathway Addiction**
 - Panomicsより患者特異的パスウェイ分枝を決定



1. 疾患オミックスプロファイルから
→ **患者特異的分子ネットワーク(個別化医療)の**
調節不全分枝 同定
Dysregulated pathway/subnetwork の同定

2. パンオミックスによる臨床的実践の戦略

- 遺伝子発現プロファイル** 推定法による分子ネットワークの同定 (80%)
- 次世代シーケンシング** 転写因子や信号パスウェイスイッチ分子の変異
- リン酸化プロテオーム** パスウェイバイオマーカ
リン酸化状況の認識

分子ネットワーク歪構造同定

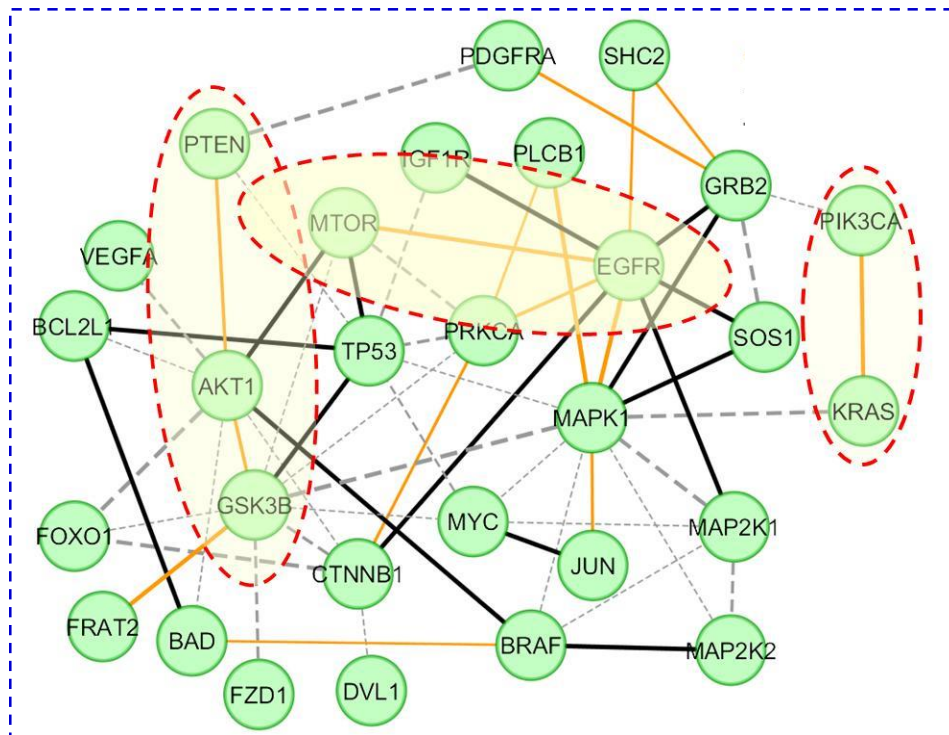
個人のpanomics1例だけではネットワークは同定できない
臨床表現型と合わせて層別化する。類似症例検索より推測

疾患ネットワーク同定方法論

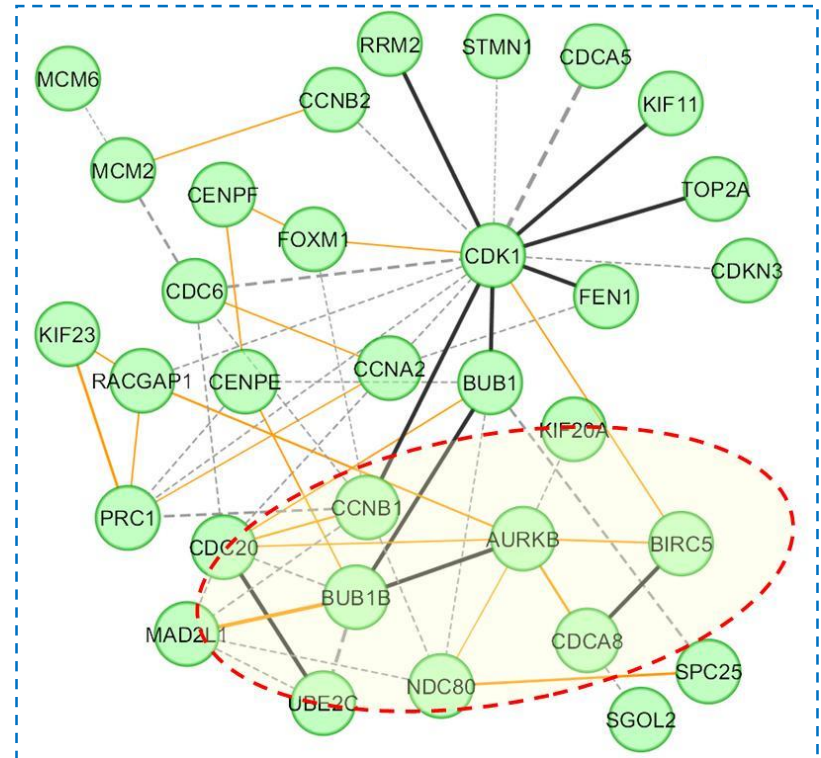
東京医科歯科大学病院 肝胆膵外科にて肝胆膵外科で手術を受けた40症例

— 予後良好群20例 vs. 予後不良群20例

細胞増殖分子ネットワーク推定
differential branches 増殖系の過剰活性

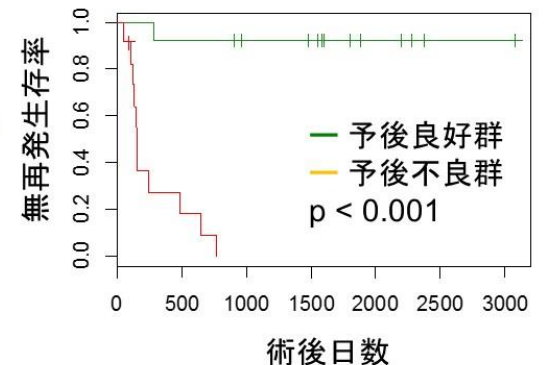


細胞周期分子ネットワーク推定
differential branches 結果 AurkBを起点としたNW



差異的発現分枝

— 予後不良群に特徴的
— 両者共通
- - - 既知PPIあり



薬理ゲノム学 (PGx) から システム薬理学へ

- **マルチオミックス情報収集の普及**
 - 次世代シーケンスやL1000技術の発展
 - 液性バイオマーカ (Liquid biopsy)の進展
- **疾患バイオバンクの発展**
 - ビッグデータの蓄積による「疾患の層別化」が進展
- **臨床生命情報学の発展**
 - 1患者情報による「患者特異的分子ネットワーク」同定



システム薬理学による創薬・治療

患者特異的分子ネットワーク同定に基づいた薬剤治療システムの戦略

分子ネットワークの上流や下流に注目した薬剤の開発

副側パスウェイの効果を抑制する薬剤など併用

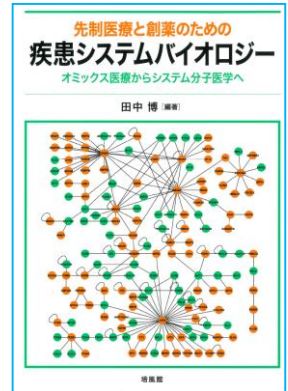
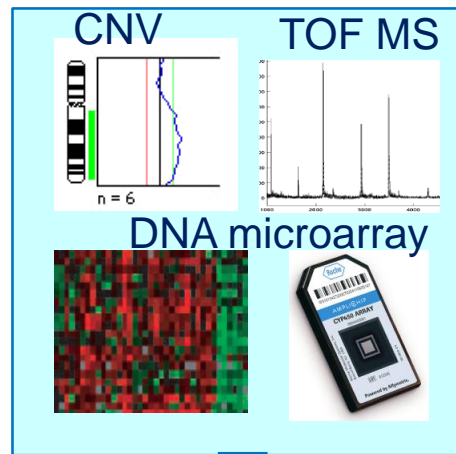
分子ネットワークの薬剤摂動や時間経過による構造変化に基づく薬剤治療

Precision Medicineとは 分子システム医学のことである

断層像
再構成技術



疾患
パンオミックス
プロフィール



スーパー
コンピュータ



逆計算

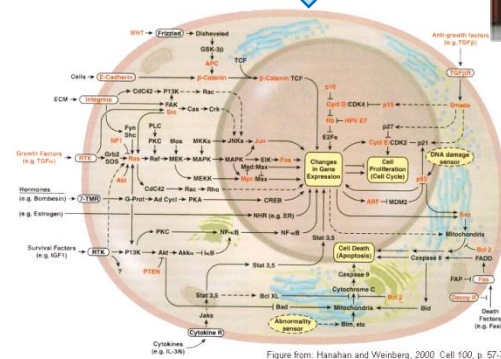


Figure from: Hanahan and Weinberg, 2000. Cell 100, p. 57-70

合理的な
診断・治療
予後予測

患者特異的分子ネットワーク
調節不全分枝同定



ご清聴ありがとうございました

