

# 医療ビッグデータ, Precision Medicine,創薬/DR

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

田中 博



# ビッグデータを取り巻く状況と 医療のパラダイム転換

# 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」

～何が新しいのか～

＜目的もデータ形式も従来型と違う＞

従来の医療情報の「ビッグデータ」

**Big “Small Data”**

医療情報・疫学調査 属性数：10項目程度

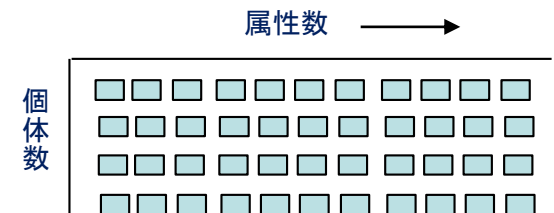
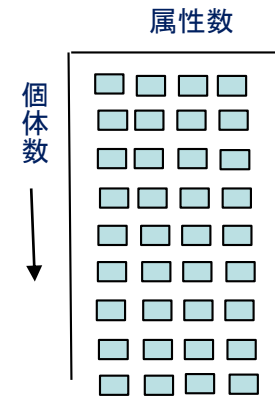
- 目的：Population MedicineのBig Data  
集合的見地から「医療の集合的法則」を見る

ゲノム・オミックス医療, mHealth の  
ビッグデータ

**Small “Big Data”**

1 個体に関するデータ属性数が膨大

- 属性に比べて個体数が少数
- 従来の統計学が無効  $p \gg n$
- 目的：Personalized Medicine のため  
多様な個別化パターンを多数を集める



新しいデータ科学の必要性

# わが国における従来型ビッグデータ

- 従来型電子化医療情報のビッグデータ
  - 患者診療電子化データをNation-wide（国規模）に蓄積した医療データベース
  - 日本版センチネル計画(2011～)
    - 米国のセンチネル計画（2008, HHS, FDA, 1億人データ）の日本版（10大学病院）
    - 薬剤投与履歴、アウトカム（副作用など）情報
  - NDB（ナショナルデータベース）
    - レセプト(2009年～)・特定健診(2008～)情報
    - レセプト情報（80億件以上、7000万件）
  - National Clinical Database：症例データベース（医師のみ：400万件）その他、各疾患領域で広がっている
  - 分譲可能な従来型コホート(JPHC,J-MCCなど後述)

# 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>の医療はもはや成り立たない
  - **Personalized (Precision) medicine**
  - 個別化医療を実現するために<層別化パターン>を集める。
    - どこまでの粒度で層別化すればよいか。そのためにビッグデータ収集
- **Clinical researchのパラダイム転換**
  - 臨床研究を科学にする従来の範型RCT
  - RCTの呪縛からの解放
  - 「標本」統計・「推測」統計学ではない臨床知識獲得法
  - **Real World Data: ビッグデータ知識獲得(BD2K)**
- **Occasionalでreactiveな医療**
  - Life-long healthcare の見地からProactiveでLife-longな医療
  - オミックスバイオマーカで疾病発症までの過程を詳細認識
  - **Pre-disease段階を検出→先制医療薬へのシフト**

# 医療ビッグデータを 巡る状況(1)

High Throughput 技術発展による  
ゲノム・オミックス医療情報の急増

# ゲノム・オミックス・ビッグデータ を取り巻く状況

- 我が国では
  - まだゲノム・オミックス・ビッグデータの”波”は臨床現場には押寄せてはいない。
- 米国では、
  - 2010年のWisconsin小児病院の腸疾患小児の例から5年経て、多くの有名病院でClinical Seqが日常臨床化して、ゲノム/オミックス・ビッグデータの時代は始まっている。
- しかし、我が国でも
  - 幾つかのバイオバンクのゲノムコホートのデータ蓄積はすでに始まり、数年のうちに医療ビッグデータ時代が押し寄せるだろう。
  - それを向かえる対応を準備する必要がある

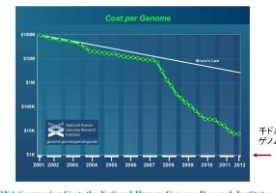


# 医療ビッグデータ時代の到来

クリニカル・シーケンシングの普及など

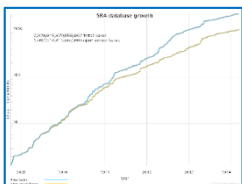
次世代シーケンサの臨床応用 全ゲノム解析 (WGS) 100Gb

米国では数十の著名病院で実施



DNA Sequencing Cost: the National Human Genome Research Institute

2000兆塩基 (2 Pb)  
が登録 (NCBI:SRA)



ゲノム・オミックス情報の蓄積

従来型医療情報  
との統合

クリニカル  
フェノタイピング

臨床情報  
形式化

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



ゲノム医療の実践

医療ビッグデータ



MayoClinicでは  
10万人患者WGS

# 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

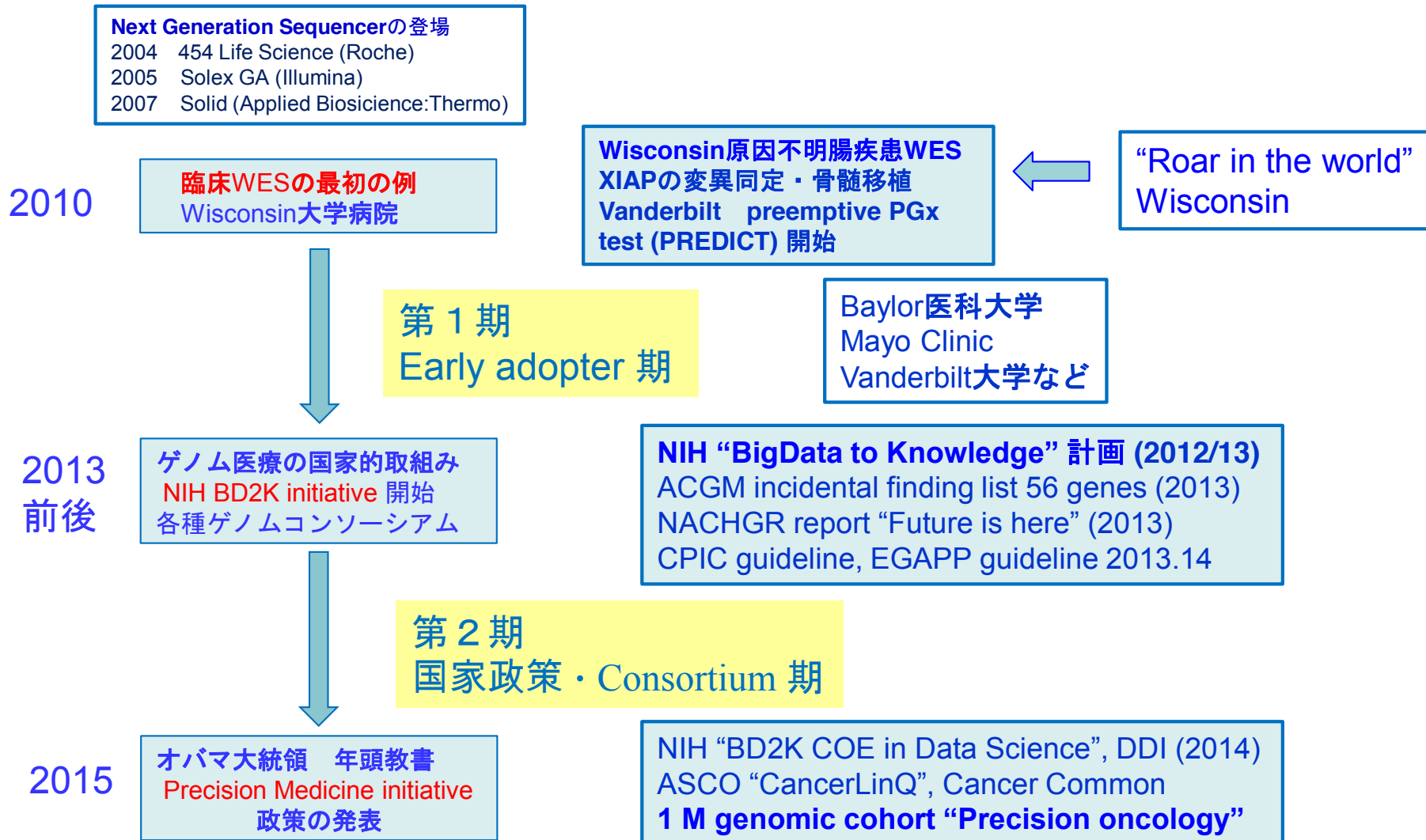
## NHGRI Working Groupのリスト

- 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
  - 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"> <li>• CYP2C19 testing for antiplatelet rx post percutaneous coronary intervention</li> <li>• Personalized decision support for CVD risk management incorporating genetic risk info</li> </ul>
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"> <li>• Screening for actionable mutations in malignant gliomas and glioblastomas for biomarker based RCTs</li> <li>• Targeted rx (such as RET inhibitor) of metastatic solid tumors based on tumor mutation status</li> </ul>
Morehouse	• Exome sequencing of 1200 early onset severe African American hypertension cases and 1200 controls
Duke	<ul style="list-style-type: none"> <li>• 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</li> <li>• SLC01B1*5 genotyping and statin adherence</li> <li>• Effect of genetic risk info on anxiety and adherence in T2DM</li> </ul>

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"> <li>• Selection for gastric bypass surgery vs other wt loss means based on genetic variants predictive of long-term benefit from surgery</li> <li>• IL28B variants and response to hepatitis C treatment</li> <li>• KRAS and BRAF mutational analysis in thyroid cancer patients</li> </ul>
Ohio State	<ul style="list-style-type: none"> <li>• Personalized genomic med study of CHF and HTN pts randomized to genetic counseling vs usual care</li> <li>• CYP2C19 testing in interventional cardiovascular procedures for clopidogrel</li> </ul>
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"> <li>• PGx driven selection/dosing of antidepressants</li> <li>• CYP2C19 genotyping for antiplatelet rx post PCI</li> </ul>
Inter-Mountain	Tumor-based screening for Lynch syndrome

# 米国のGenome/Omics医療の歴史



# ゲノム配列解読の最初の臨床応用 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

臍帯血による骨髄移植を実施 (2010年6月)  
2010年7月半ばには、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。



Medical College of Wisconsin, Human & Molecular Genetics Center, Director  
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

- 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
  - 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
    - ゲノムシーケンスプログラムに登録
  - 32 全ゲノム, 550 全エキソーム (2015年4月まで)
  - アメリカ病理学会 (CAP)およびClinical Laboratory Improvement Amendments に  
よって承認されたLabo(CLIA). 最初は外注。
  - 解析：in-houseのBioinformaticianで
    - 最初, 候補とした変異について調べ
    - つぎにincidental findingも含めて精査
    - ACGM勧告 (56遺伝子)

Wisconsin  
小児病院



Wisconsin 医科大学 (MCW)



Froedtert 病院





# その他の臨床ゲノムシーケンス実施病院

- Baylor医科大学病院

- 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/変異分析ともに
- CAP/CLIA認証の検査室を病院内に立ち上げる。  
そこでの配列解読
- 臨床分子遺伝学者によって解析・結果報告



- そのほか

- Washington大学、Partnerヘルスケア、Ohio大学、Geisinger clinic など多数続く、前スライド表

# メイヨクリニック

## 個別化医療 (Individualized Medicine)

- Center for Individualize Medicine (2011)発足

- 5つのTranslational Programを準備

- 臨床部門 CIM clinicsを拡張 (2012)

- Clinical Sequencing
- 難治性がん
  - 例 胆管がん
- 原因不明遺伝病
  - “診断オデッセイ”

Translational programs	Infrastructural programs
Pharmacogenomics	Biomedical Informatics
Biomarker Discovery	Biobank and Biorepository
Clinomics	Bioethics
Epigenomics	Medical Genome Facility
Microbiome	Information Technology



- 全患者に全ゲノム配列解析(WGS), エキソーム配列解析(WES)

- 10万人患者 (診療圏) の全ゲノム・データベース構築

- eMERGE コンソーシアム

- 初期からのメンバー
  - phase I (2007-11)から参加、
  - phase II (2011-15)では

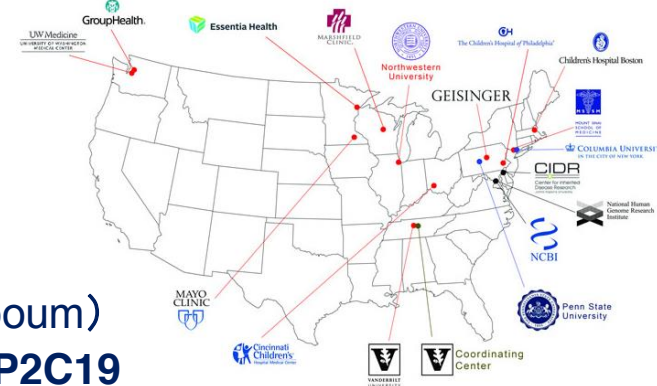
臨床現場におけるゲノム医療の実現

先制的ゲノム薬理学 (PGx) 検査 (Weinshilboum)

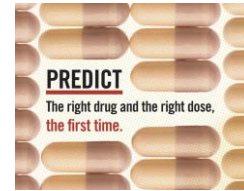
同時期にOhio大学、Vanderbilt大学 CYP2C19

ゲノム医療臨床支援 (CDS) が電子カルテEMRに装備

- RIGHT protocol: Right Drug, Right Dose, Right Timed Using



# 薬剤代謝酵素多型性のゲノム医療実践 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](#)

**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.

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・オミックス医療の基軸概念

## 3世代区分

### 第1 (生得的) ゲノム医療

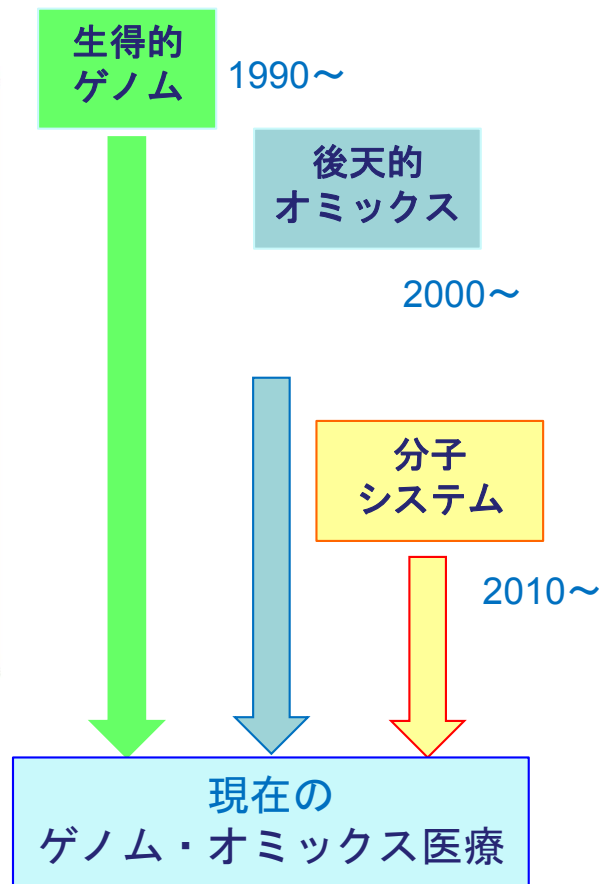
- 生得的ゲノム情報の変異・多型性

### 第2 (後天的) オミックス医療

- 後天的体細胞の網羅的分子プロファイル

### 第3 システム分子(分子システム)医学

- 細胞分子ネットワークの歪み/構造変化



H.Tanaka:

”Genome and Omics Medicine

—Principles, Clinical Implementation and BigData Approach—

Springer 2016

# 網羅的分子医学 第1パラダイム

## ゲノム医療 (1990~現在)

Manolio (2009)

- 「生得的な(germ-line)」ゲノム変異や多型性に基づいた個別化医療

### 生得的ゲノム

- 全細胞で生涯を通じて同一99.5%は共通

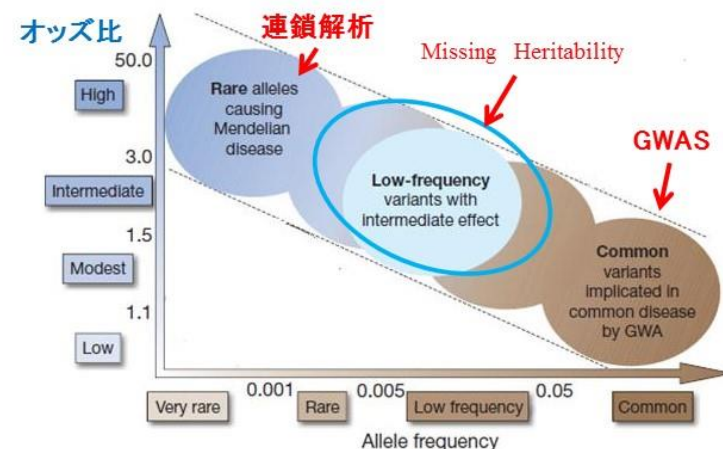
- 疾患原因遺伝子(Disease causative gene)

### 家系調査/Linkage解析

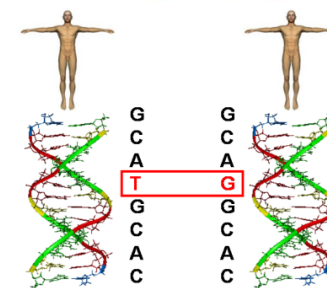
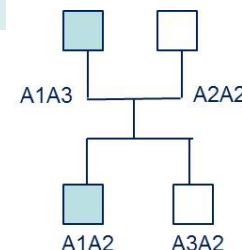
- 1980年代から: ハンチントン病
  - CAGリピート, HTT遺伝子, 99%浸透率
- その他にデシャンヌ型筋ジストロフィー, 嚢胞性線維症
- 当時400程度のDNAマーカ ヒトゲノム解読計画へ

- 疾患感受性遺伝子(Disease susceptibility gene)

- 多型性: 一塩基多型(SNP), 3000万位,
  - そのほかにマイクロサテライト、CNVなど
- 全ゲノム関連解析 (GWAS) 患者対照分析
- 薬剤代謝酵素の多型性 (PGx)



単一分子へ起因する方法の限界



1塩基の変異が疾患や薬剤への応答性に関与する場合がある

# 網羅的分子医学 第2パラダイム オミックス医療 (2000~)

- 「後天的・疾病依存的に変化する網羅的分子病態像」

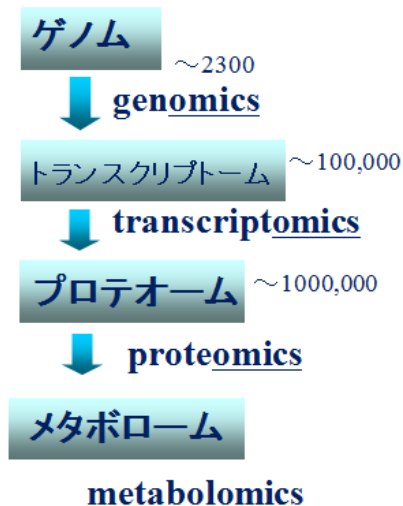
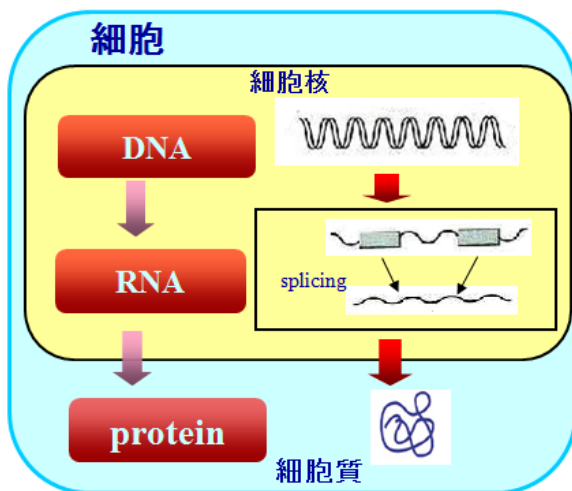
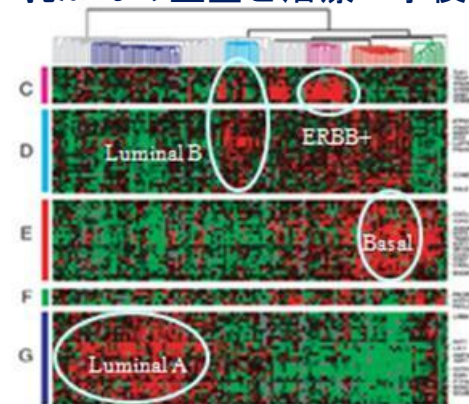
後天的体細胞ゲノム変化+ゲノム以外の疾患オミックスに基づく医療

- 疾患 オミックスプロファイルによる分子的病態像
  - 後天的ゲノム変異・遺伝子発現プロファイル・プロテオームなど、病態進行の段階に依存

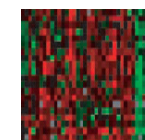
- 疾患の進行度の評価 → 予測医療・先制医療

- Ongoing state of disease progressionを表現
- 臨床症状や病理変化より早く変化 Predictive Omics

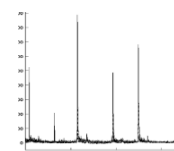
乳がんの亜型と治療・予後



Next generation sequencer



DNA microarray



mass spectro



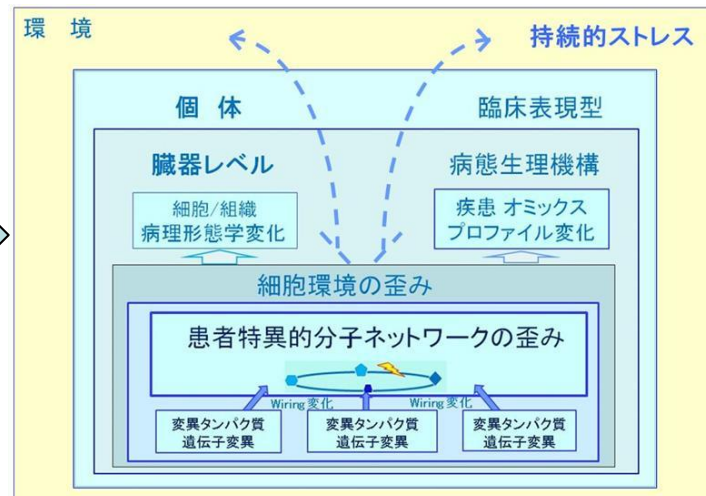
# 網羅的分子医学 第3パラダイム システム分子（分子システム）医学（2010～）

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  
(分子システム) 構造変化」  
疾病の理解における  
第3のパラダイム



# ゲノム・オミックス医療の臨床実装の現状

## 実施されているゲノム・オミックス医療

1. 病因未知の遺伝疾患のWGS/WESによる原因遺伝子変異の同定(undiagnosed disease)  
Wisconsin大学病院、Baylor大学病院
2. 難治性のがんのWGS/WESによるDriver 遺伝子変異の同定  
Mayo Clinic, MD Anderson
3. 遺伝性がんの生得的な原因遺伝子  
BRCA1/2などの診断
4. 薬剤代謝酵素の多型性診断と電子カルテへの実装 (DNAアレイ)  
Vanderbilt大学病院・Mayo Clinic

第1世代が殆んどである



# 米国のゲノムオミックス医療の 推進・評価組織

当初一番乗りを目指した病院別プロジェクト⇒  
共通の課題を解決する国家ProjectやConsortiumへ

- **NACHGR** :
  - National Advisory Council on Human Genome Research : NHGRIが主導
  - **NACHGR working group**
  - "early adopter"を集めて個々の経験を共有し、まとめてゲノム医療の課題を検討する
- **EGAPP** :
  - CDCP (Center for Disease Control and Prevention) が設立(2005)
  - Evaluation of Genomic Applications in Practice and Prevention
  - 遺伝子診断法の評価、臨床応用の適切性判断
- **CPIC: 薬理ゲノム学臨床実装コンソーシアム**
  - **Clinical Pharmacogenomics implementation Consortium,**
  - 薬理ゲノム研究ネットワーク (PGRN) のメンバ, PharmGKBのスタッフで構成,
  - 薬理ゲノムガイドンス(CPIC guideline) を臨床家、臨床検査室に提示し、PGx試験が広く臨床で用いられることを目的
- **ACGM** (American College of Medical Genetics and Genomics)
  - Incidental findingの推奨リスト (56遺伝子)

# 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**
  - 「データ・情報学に関するNIH長官諮問委員会」WG
  - 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に対応して始まる
  - 研究費の配分 2013年に開始。計画実施予算は2014年から
  - データ科学のための副長官 (Associate Director of Data Sciences) を任命 **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。
- <http://bd2k.nih.gov>

# NIH BD2K計画の実施

- 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



# 関連プロジェクト *eMERGE*

- 前史 : Nat. Center for Biomedical Computing : i2b2 形式の医療データ
- Phase I (2007-2011)
  - 電子カルテから臨床phenotypingを行う
  - EMR : 臨床phenotypingとBiorepositoryに基づくGWASが可能か (EMR-based GWAS) 利用に関する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)

- 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
  - 小児病院とMount Sinai/Gesinger



- CSER consortiumとliason

- “Clinical Sequencing Exploratory Research” コンソーシアム NHGRIにより予算化
- 臨床的転帰を改善するための全ゲノム/エキソーム解析

# 臨床データの表現型形式化 (Phenotyping) の問題

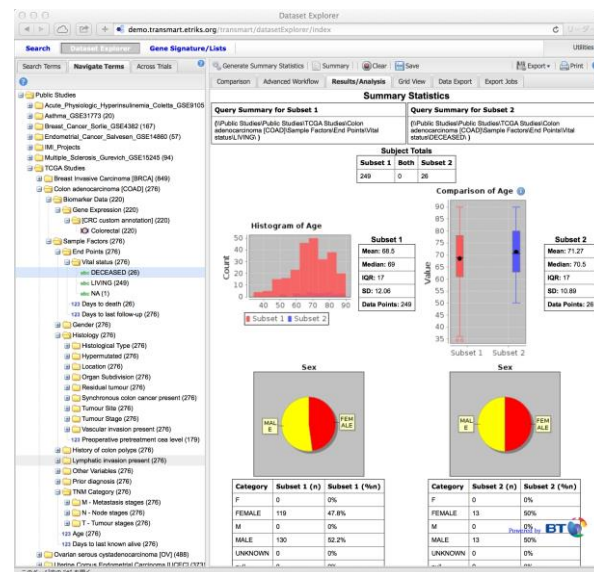
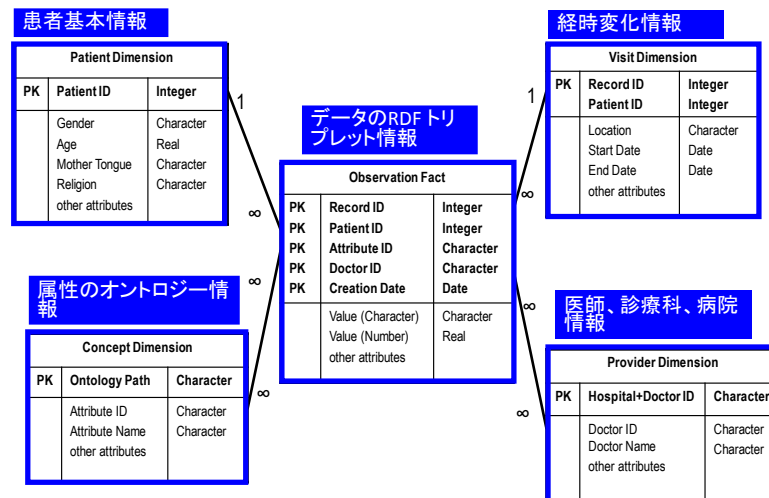
## i2b2 (Informatics for Integrating Biology and the Bedside)

- 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 (三つ組み RDF : resource description framework) 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
- 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

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. <b>Note:</b> With clinical trials, this node is typically named Treatment Groups.
Sample Factors	Patient information, such as demographics and medical history.



# バイオ・医療におけるBig Dataの 構成要素

- 網羅的分子情報 (Genome/Omics)
  - ゲノム・オミックス情報、multi-omics
- 臨床環境表現型 (Phenotyping)
  - eMERGE-I, PheKB, EWAS
- 知識発見システム (Data Science)
  - Data-mining, Knowledge Discovery, 人工知能

医療におけるビックデータ

ゲノム情報 + 臨床・環境情報(EMR等)

知識発見(learning) システム

# 米国のBD2Kの幾つかの動き (1)

## CancerLiQ

- ASCO(米国臨床癌学会)
- The ASCO **CancerLinQ** initiative
  - “Learning Health System”あるいは”**Rapid Learning**”の概念をがん治療において実現する。
  - データがバラバラの状態を改変、集合化
  - 診療の現場(EHR)から大量の診療データを集め分析する
  - 新しい臨床治験へのガイドライン検討
  - 病院の患者によって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、仮説を生成する治療知識を供給する。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

# 米国のBD2Kの幾つかの動き (2)

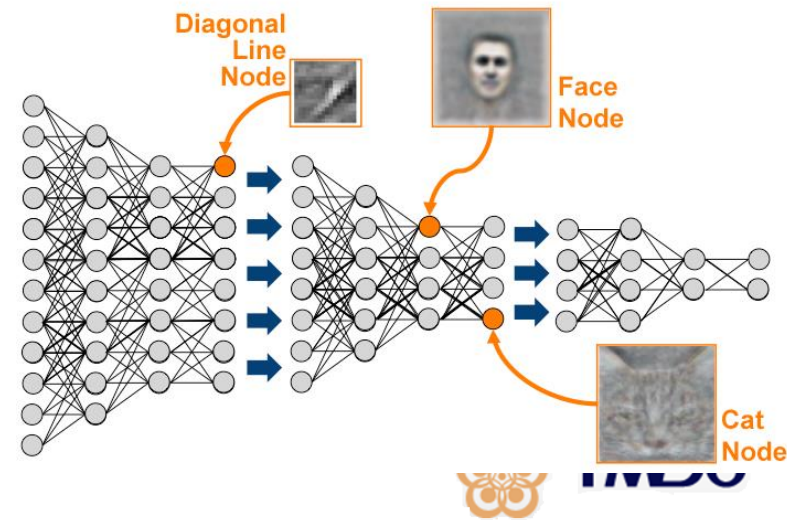
## 米国研究施設とIBM Watson

- **Learning systemの不可欠性: IBM Watson**
  - 自然言語処理、大量データベース探索、確信度付き解答: Deep QAシステム (jeopardy)
  - MITのSTARTと呼ばれるオンライン自然言語QAシステム: 質問をシンプルな質問に分解
  - CMUのOpen Advancement of Question-Answering Initiative (OAQA) システムが骨格
  - 質問解答に最も適切なテキスト資料を特定する知識源拡張アルゴリズム。テキストから知識を自動的に抽出
  - 大規模情報抽出、構文解析、知識推論により大量の情報資料をシステムの一般知識情報源に変換
  - 自然言語理解に応用される統計学的学習理論 (例えば、カーネル法) が基礎
- **Memorial Sloan-Kettering Cancer Center (MSKCC)**
  - The Oncology Expert Adviser software (OEA)
  - IBMワトソンの計算能力および自然言語処理技術と、MSKCCが持っている臨床知見 (分子・ゲノムデータ、がん病歴の膨大なリポジトリなど) を組み合わせ、個々の患者にとって最高の治療方針を決定するのに役立つ、最新の研究に基づいた詳細な診断情報や治療の選択肢を見出す
- **New York Genome Center**
  - がん専門医ががん患者に対してより良い個別ケアを提供できるよう支援するツールとしてゲノム研究専用デザインされたWatsonの試作システム
  - 最初の対象として脳腫瘍のglioblastoma (グリア芽細胞腫)、ゲノム配列と医療情報、医学文献から個別化治療を提案
- **その他のWatsonの応用**
  - Cleveland Clinic
  - 問題解決型学習プログラム、仮説的な臨床シミュレーションの中で、Watsonを対話的に利用して症例を分析。Watsonの技術を利用した共同型の学習教育ツール
  - MDA Anderson
  - 治験に適切な患者を診療情報から選別(clinical trial)

# 米国のBD2Kの幾つかの動き (3)

## そのほか

- 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
  - 匿名化して研究に使えるようにする。”e-trials”
  - Melanomaに対するipilimumab
- Craig Venter “Human Longevity Inc.”
  - 健康・長寿（健康寿命伸長）
  - **ゲノム科学、幹細胞治療**（Haririと共に）初期資本7000億円HiseqX 5sets
  - 一年**40000ゲノム**（幼児/老人,患者・健常者）収集、
  - ゲノムDB構築、臨床情報も収集
  - 腸内細菌も含む 一日5人のヒト全ゲノム
  - がん（Moore’s Cancer Centerと提携）糖尿病、認知症などの成人疾患も
  - 機械学習の専門家が加わる
- Google Xプロジェクト“Baseline”
  - 健康に関する尺度発見
  - Conrad AのもとにDuke大学やStanford大学が協力
  - 現在175名、先制医療的なバイオマーカー探索
- **Deep learning** とゲノムオミックス医療への応用
  - 医療への応用
  - 多層化ニューロネット
  - 多段特徴抽出





# わが国での ゲノム・オミックス医療の臨床実装

(Clinical Implementation of Genomic Medicine)

- 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。
- 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
    - 遺伝子検査より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
- 先天的神経筋疾患
  - 東大病院ゲノム医学センター

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、近年はゲノム情報の収集が加わる
- **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）とそれに対応できる臨床表現型（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£), 2011-16, 25M£
  - 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100万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



# 国内の主なバイオバンク・ゲノムコホートの状況

我が国における主なバイオバンク・ゲノムコホートを対象者、規模、目的で大別

大規模／多目的

## 東北メディカル・メガバンク

東北大学・岩手医科大学  
2013年～/被災住民8万人+3世代7万人

## JPHC(多目的コホート)

国立がん研究センター  
1990,93,2011(NEXT)～地域住民/10万人

## J-MICC(多施設共同コホート)

愛知がんセンター等、多施設共同研究  
2005年～/10万人、愛知がんセ・名大

山形分子疫学コホート 2002年/9100人

## 久山コホート

追跡率99%・剖検率80%  
50年の歴史/8000人九州大学

## ながはま0次予防コホート

京都大学 2002年～/9100人

## 鶴岡メタボロームコホート

鶴岡市・慶応大学 2013年～/10000人

## バイオバンク・ジャパン

(BBJ) /東大医科研・理研  
2003年～/20万人(第3期～)  
47疾患・12医療機関

## ナショナルセンターバンク (NCBN) 2012～

- ・国立がん研究センター
- ・国立循環器病研究センター
- ・国立精神・神経医療研究センター
- ・国立長寿医療研究センター
- ・国立国際医療研究センター
- ・国立成育医療研究センター

## 難病バンク

2009年～医薬基盤研究所(40種以上分譲)

小規模／特定目的(特定疾患)

健常者

患者

# Biobankの目的

## 組合せ特異的遺伝子環境相互作用

Idiosyncratic Effect of Combination of GxE factors

- 遺伝的素因と環境の相互作用
- 相互作用の特異的組合せ効果
  - ハワイの白人、日系人と結腸がん発生
  - **相対リスクの乗算ではない。Idiosyncratic Effect**

		CYP1A2 Phenotype ≤ Median		CYP1A2 Phenotype > Median	
		Likes rare/medium meat	Likes well-done meat	Likes rare/medium meat	Likes well done meat
Non-Smoker	NAT2 Slow	1	1.9	0.9	1.2
	NAT2 Rapid	0.9	0.8	0.8	1.3
Ever-Smoker	NAT2 Slow	1	0.9	1.3	0.6
	NAT2 Rapid	1.2	1.3	0.9	<b>8.8</b>

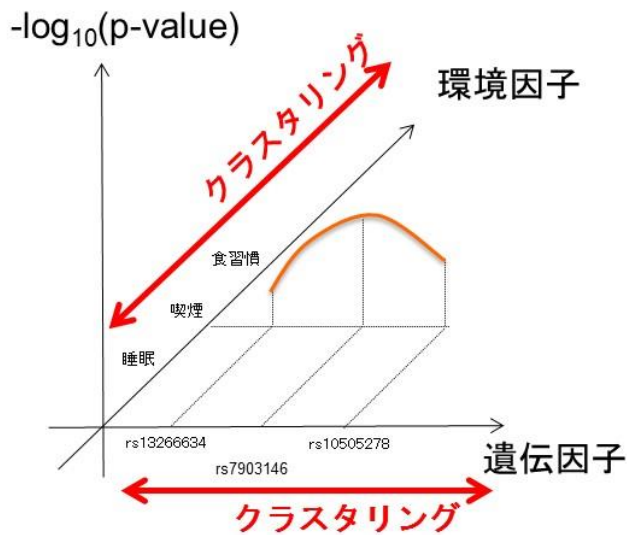
個別化予防

HCA(ヘテロサイクリックアミン, 肉を高温で焼いた時に生成される発癌物質)HCAを減らすためには、油漬け, 2.電子レンジによる下処理, 3.頻繁に肉を裏返す事が必要

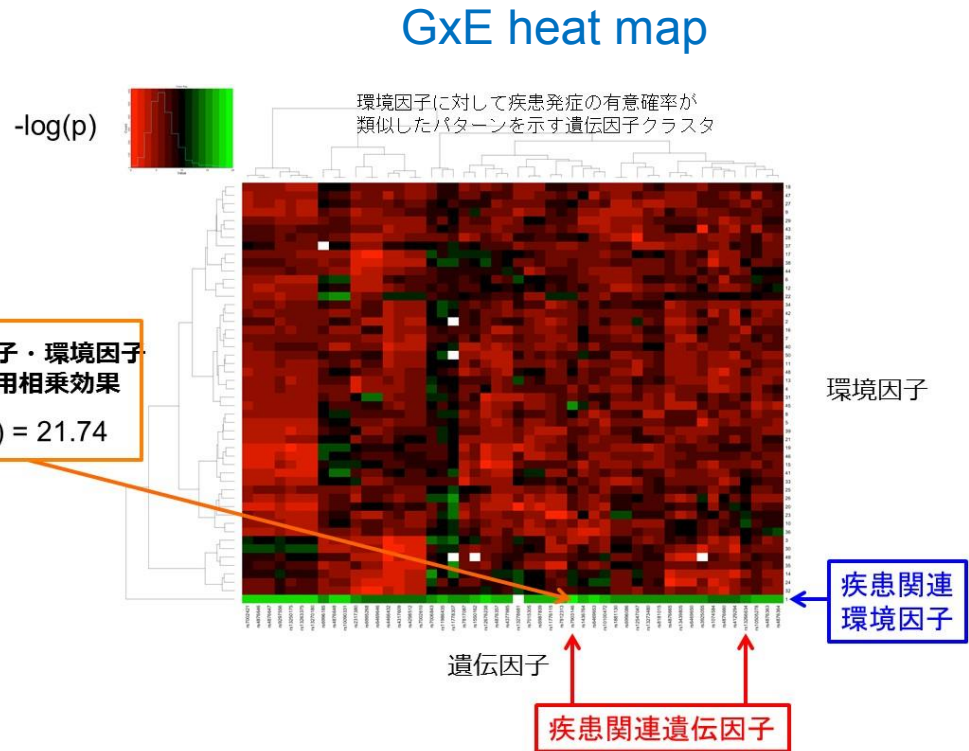
L. Le Marchand, JH. Hankin, LR. Wilkens, et al Combined Effects of Well-done Red Meat, Smoking, and Rapid N-Acetyltransferase 2 and CYP1A2 Phenotypes in Increasing Colorectal Cancer Risk, Cancer Epidemiol. Biomarkers Prev 2001;10:1259-1266

# 東北MMBの第2期の課題

## 疾患の発症に組合せ的に関連する 遺伝因子・環境因子相互作用の同定



遺伝因子・環境因子相互作用の同定



affected		environment	
		0	1
genotype	0 (aa)		
	1 (aA)		
	2 (AA)		

unaffected		environment	
		0	1
genotype	0 (aa)		
	1 (aA)		
	2 (AA)		

# 医療ビッグデータを 巡る状況 (2)

日常的センシング  
mHealth, wearable sensing

# 生涯的Perspectiveへの 健康・疾病概念のパラダイム転換

- <若い人中心－急性期疾患中心>の疾病概念から
- 慢性期疾患中心－「生涯にわたる」健康疾病概念へ  
Life-long (course) healthcareの概念
- 医療施設中心（施設医療）→  
患者中心・日常生活圏・患者参加型（participatory medicine）

Reactive (対応的)で  
Occasional (機会的)な  
医療・ケア



Proactive (予見的)で  
Life-long (生涯的)な  
医療・ケア

生涯型医療・先制医療への転換

# 近未来のライフコースの予測 に基づいた健康医療産業の戦略

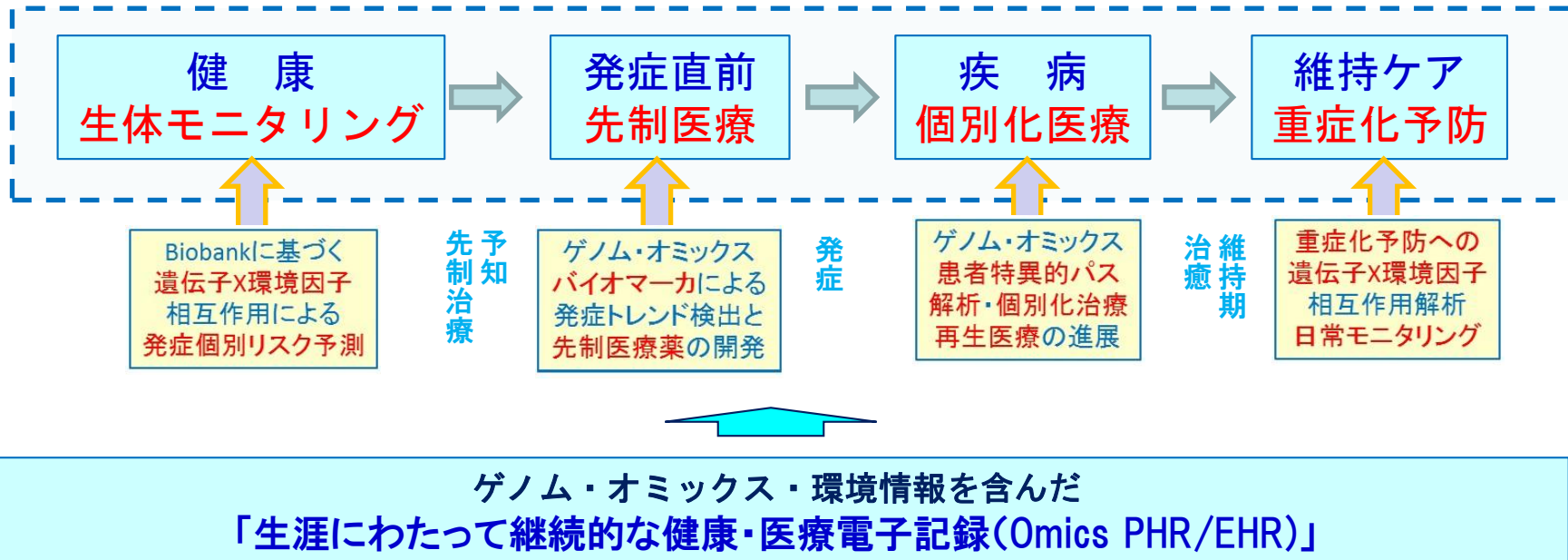
- 慢性疾患の増加
  - 超高齢化による慢性疾患（一般のがん、脳卒中、認知症）の重要性の増加。この順で発症頻度の年齢分布ピークが継起する。
- 健康モニタリング
  - 日常的生体モニタリングによる健康状態観測による発症前管理。疾病へのトレンドの予測
- 発症リスク予測
  - ゲノム・オミックス医療の進展による発症リスクの（発症前）個別化予測の確度の上昇
  - <遺伝的素因と環境・生活習慣要因>との相互作用の解明
- オミックス・バイオマーカの進展
  - バイオマーカによる発症トレンド認識の上昇・先制医療の実現
- 疾病の個別化治療
  - ゲノム・オミックス医療の進展による発症罹患後の個別化医療・予後予測医療の実施
  - <患者特異的ネットワーク調節不全>の同定を基礎にシステム医学
- オミックス・バイオマーカによる3次予防
  - 回復・維持期における再発・重症化予防オミックス日常モニタリング

生涯にわたって継続的に電子記録するOmics EHR/PHR

# 近未来のライフコースと医療

これまでの認識をまとめると健康医療ライフコースは以下のようなになる

## life-long healthcare



# 継続的モニタリングの種類

- 生涯継続的な生理量モニタリング
  - Quantified Self (mHealth)
    - モバイル・ウェアラブル生理モニタリング
    - 参加型医療・proactive professional consumer
- パンオミックス・バイオマーカの継続的モニタ
  - 疾患発症を予知するオミックスバイオマーカを継続的に計測
  - SnyderらのiPOP (integrated Personal Omics Profiling)概念
  - 疾患予知オミックスは先制医療・先制医療薬へ
  - 液性バイオマーカ (Liquid Biomarker)
  - システム分子 (分子システム) 医学的なマーカ
    - Dynamic Network Biomarker (L.Chen)



# 生体生理量センサリング

- Quantified Self

- 米国での運動、Wearable Computerと生体センシングを結合して自己の健康・行動をモニターする。世界に広がる

- 東北大学 - 東芝COI

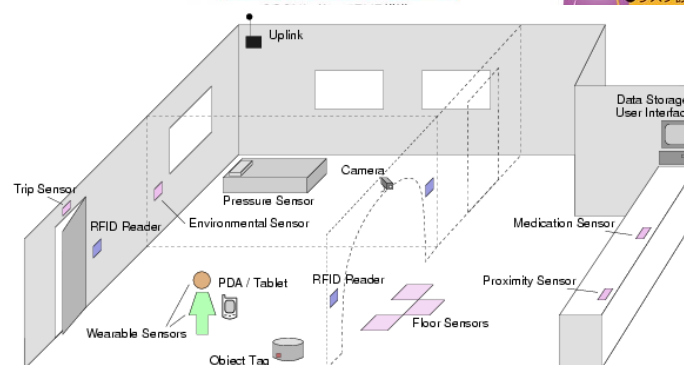
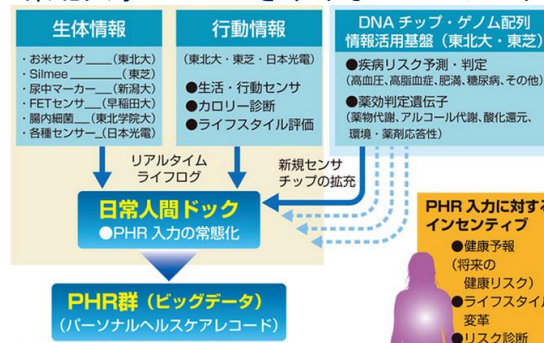
- 「さりげないセンシングと日常人間ドックで実現する理想自己」
- そのほか Wearable sensor
  - コンティニューアなど幾つかの日常モニタリング, など各種のmHealthのプロジェクト

- 生涯型電子カルテ (EHR/PHR)

- アレルギーなどのリスク情報を追加
- データ信託銀行と産業界の利用

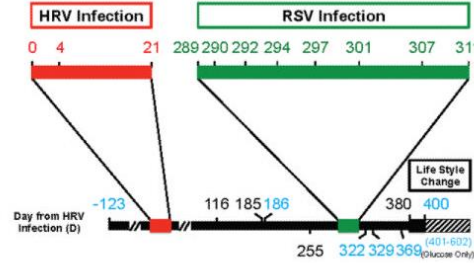
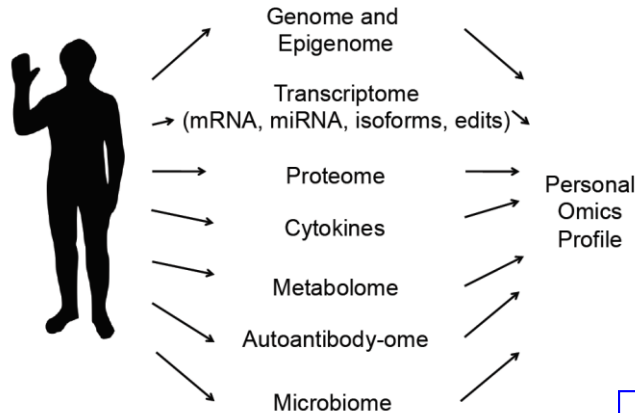


ECG; EEG; Skin Conductivity  
東北大学COI 「さりげないセンシング」

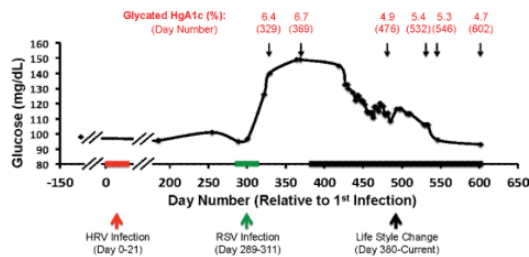


# iPOP (integrated Personal Omics Profiling)

## Personal "Omics" Profiling (POP)



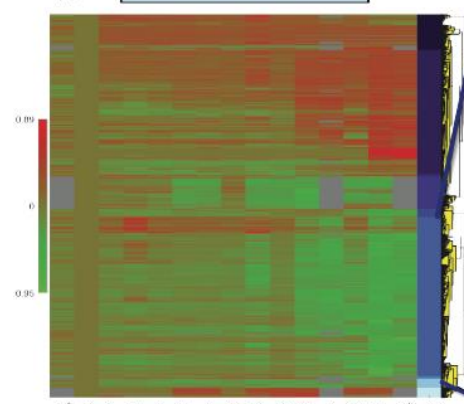
統合個別化オミックス  
 プロファイルの時系列  
 分析 (Fourier分析) に  
 より統合オミックスの  
 乱れを検知



## Integrated Omics clustering



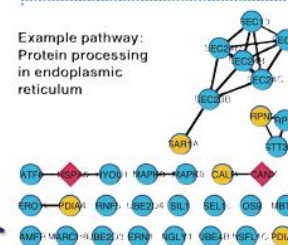
### (I) Autocorrelated data clusters



Full Reactome (FI) known pathway map for cluster:



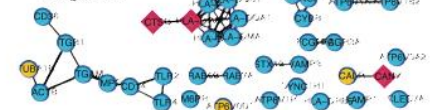
Example pathway: Protein processing in endoplasmic reticulum



Dynamic expression pattern observed in:



Example pathway: Phagosome



Example pathway: Lysosome



Example pathway: Insulin



# バイオマーカー 第2世代へ

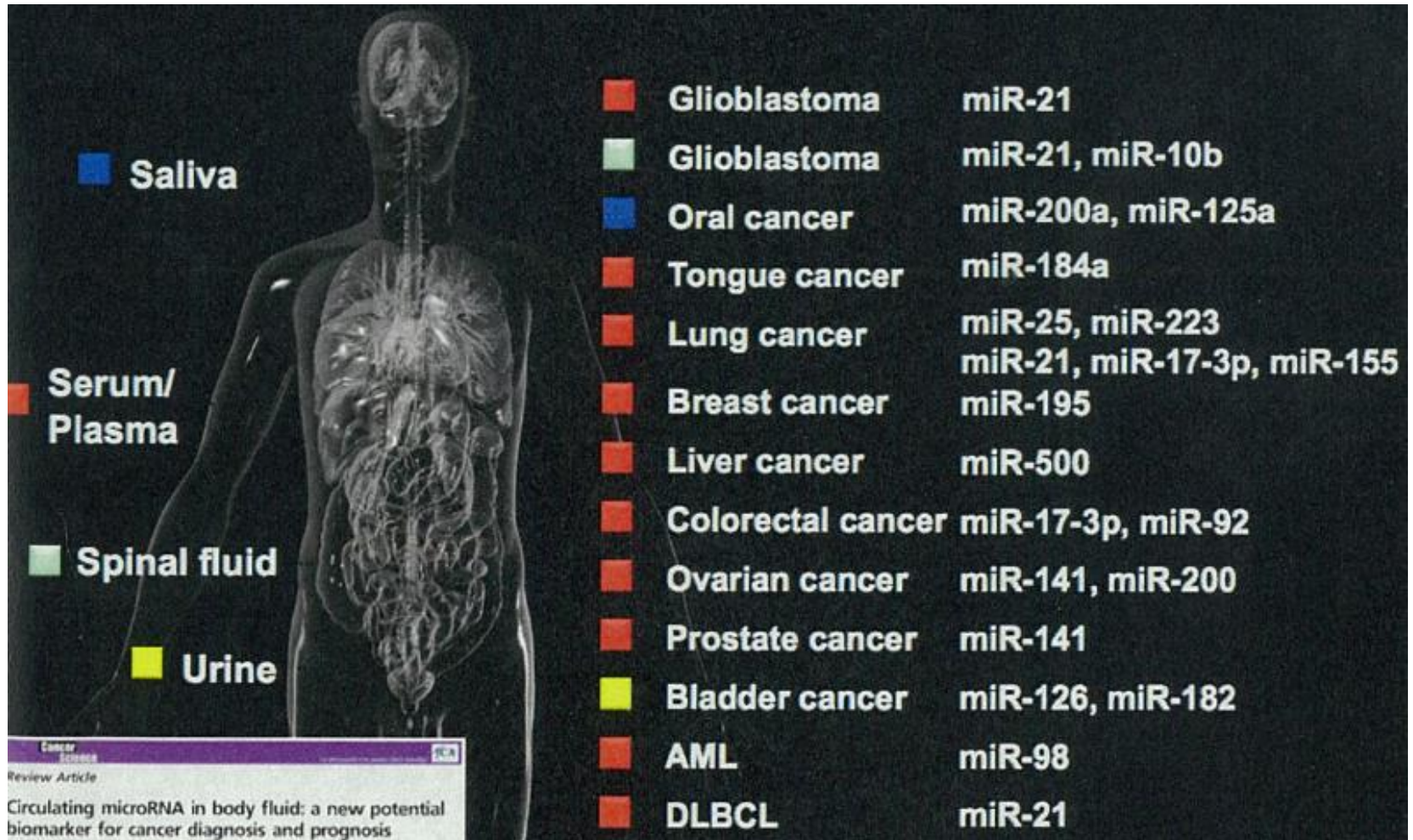
## Predictive Omics Biomarker

### 液体細胞診 (Liquid Biopsy)

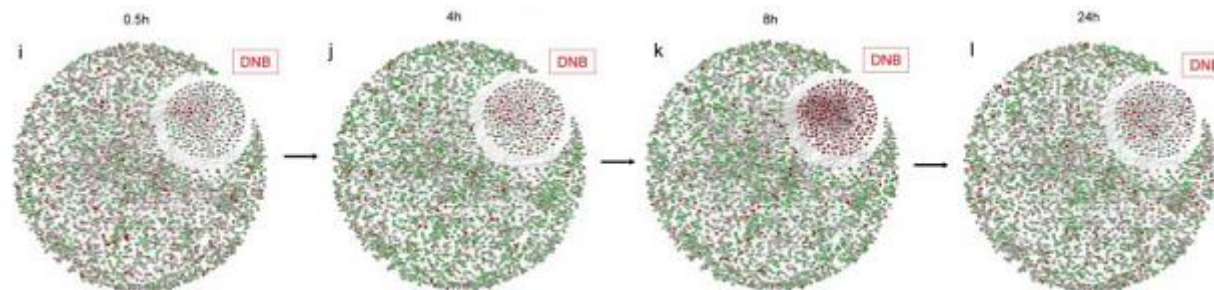
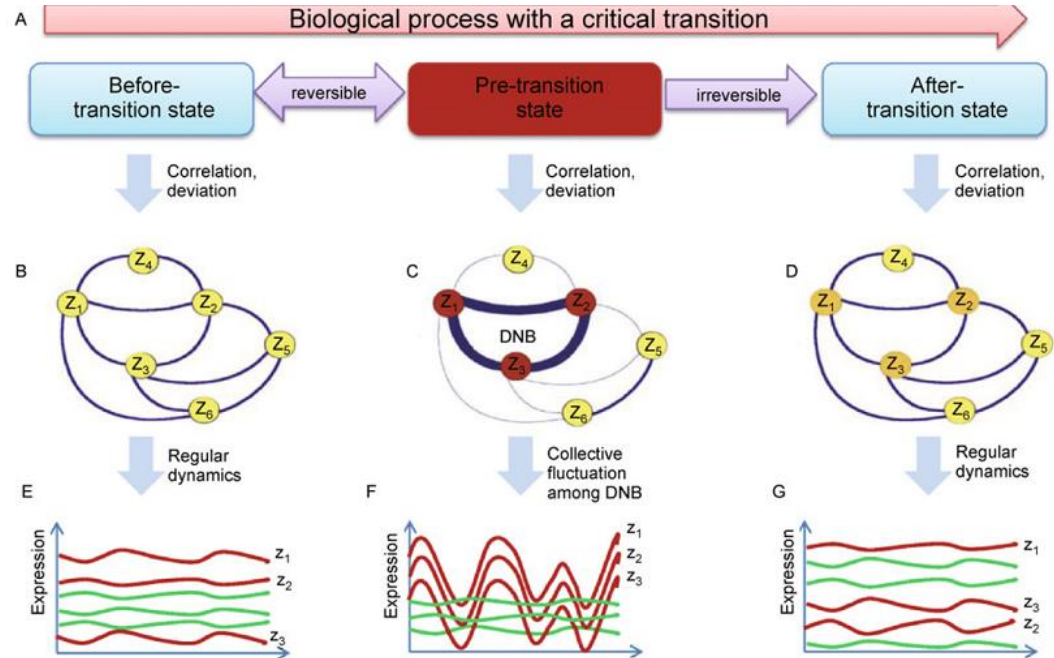
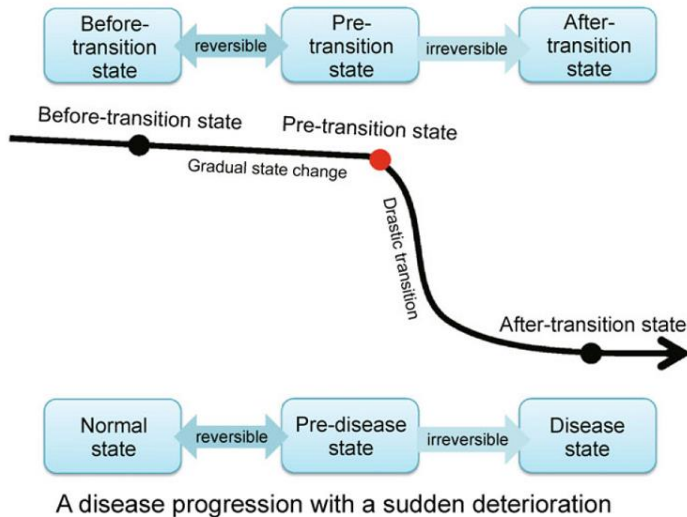
- microRNA, エキソソーム研究
  - 細胞外RNA (exRNA)
  - 唾液検査：慶応先端生命研 東京医大と膵臓がん検出84%
  - 国立がん研究センター：NEDO 5年プロジェクト (79億円)
    - 血中miRNA網羅解析、miRNAチップ、Biobankを利用して臨床関連性を抽出
- 液体細胞診 Liquid Biopsy
  - 循環腫瘍細胞 (CTC)
  - 循環miRNA
  - 循環DNA (無細胞) : ctDNA
  - 循環細胞外小胞 (Exosome)
  - がんメタボローム
- がんのEarly Detection Research Network (EDRN)
  - NIH, exRNAの研究に予算
  - NCIが主導、がんの早期発見、stage Iバイオマーカー



# ヒト体液内に循環しているmiRNA 非侵襲診断バイオマーカー



# 第3世代へ ネットワークバイオマーカー Chenの疾患発症DNB



# 生涯型生体データモニタリング

- センサー群による健康無意識管理
- 先制医療のための基礎データ
  - 健康から疾患発症の過程データ
  - 液性オミックスバイオマーカの可用性
  - PLR(Personal Life Repository)
  - 「生涯データ信託銀行」と産業界の利用
- 参加型医療のための健康医療ビッグデータ

# Precision Medicine

## 「高精度医療」



# Obama大統領と Precision Medicine Initiative

- 2015年，一般年頭教書で発表
- 個別化医療、高精度医療（precision medicine）の推進
- 100万人のコホート研究 GxE発症相互作用
- 250億円（215M\$）の予算
  - 130M\$ : NIH, 100万人コホートの組織化
  - 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

# Precision MedicineとPersonalized Medicineはどう違うか

経緯：すでにPrecision medicine は2010年前後から使われていた  
“Towards Precision Medicine” (US Nat. Research Council,2011)  
表現型から疾患体系からゲノム・オミックス機序からの分類への変革

趣旨：基本は、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拡張点

### (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

### (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

医療ビッグデータ・高精度医療が  
変革する  
医療・創薬のパラダイム転換

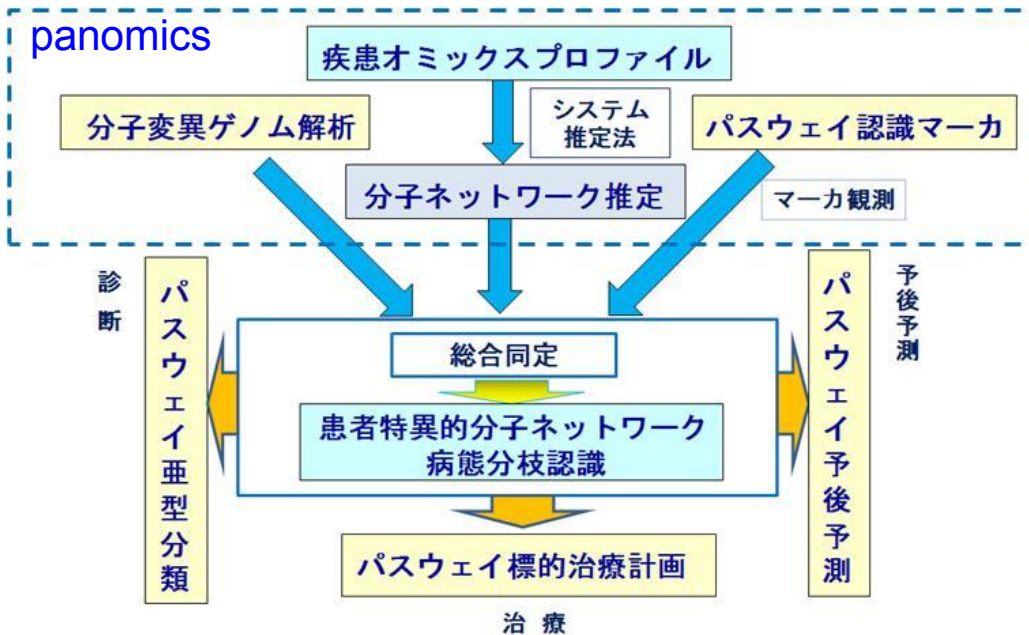
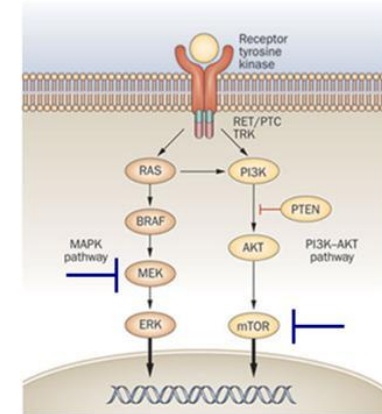
# 個別化・層別化医療

## 高精度医療の実現形態としての 第3世代医療

- Precision Medicineの目標
  - 短期的課題 Precision Oncology
  - 長期的課題 100万人コホート健康医療基盤
- Precision Medicineの実現方法については何も述べていない。CollinsのNEJM論文も。これを補完する

# 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

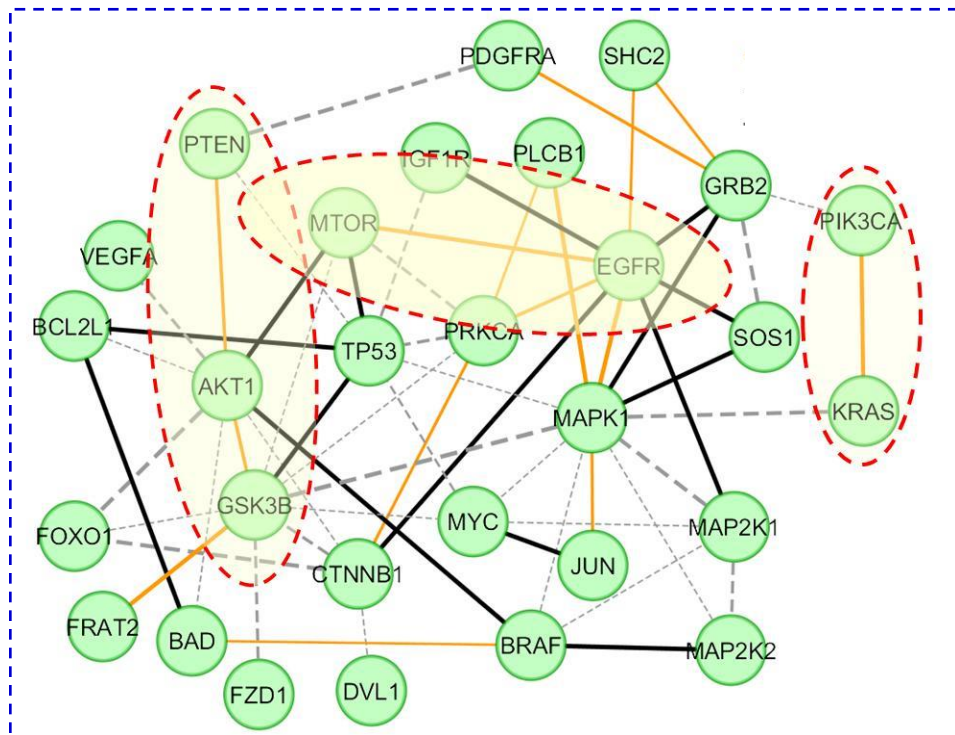
個人のpanomics1例だけではネットワークは同定できない  
臨床表現型と合わせて層別化する。類似症例検索より推測

## 疾患ネットワーク同定方法論

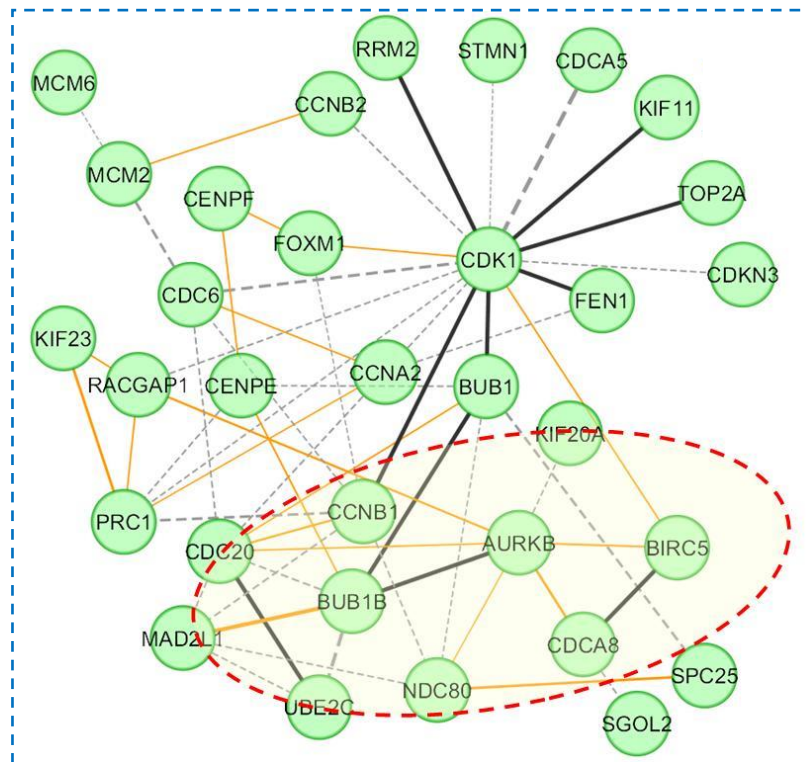
東京医科歯科大学病院 肝胆膵外科にて肝胆膵外科で手術を受けた40症例

— 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

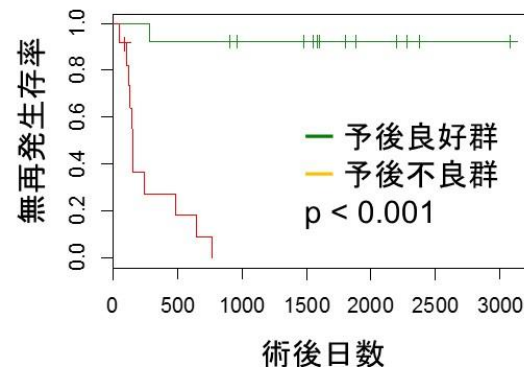


細胞周期分子ネットワーク推定  
differential branches 結果 AurkBを起点としたNW



差異的発現分枝

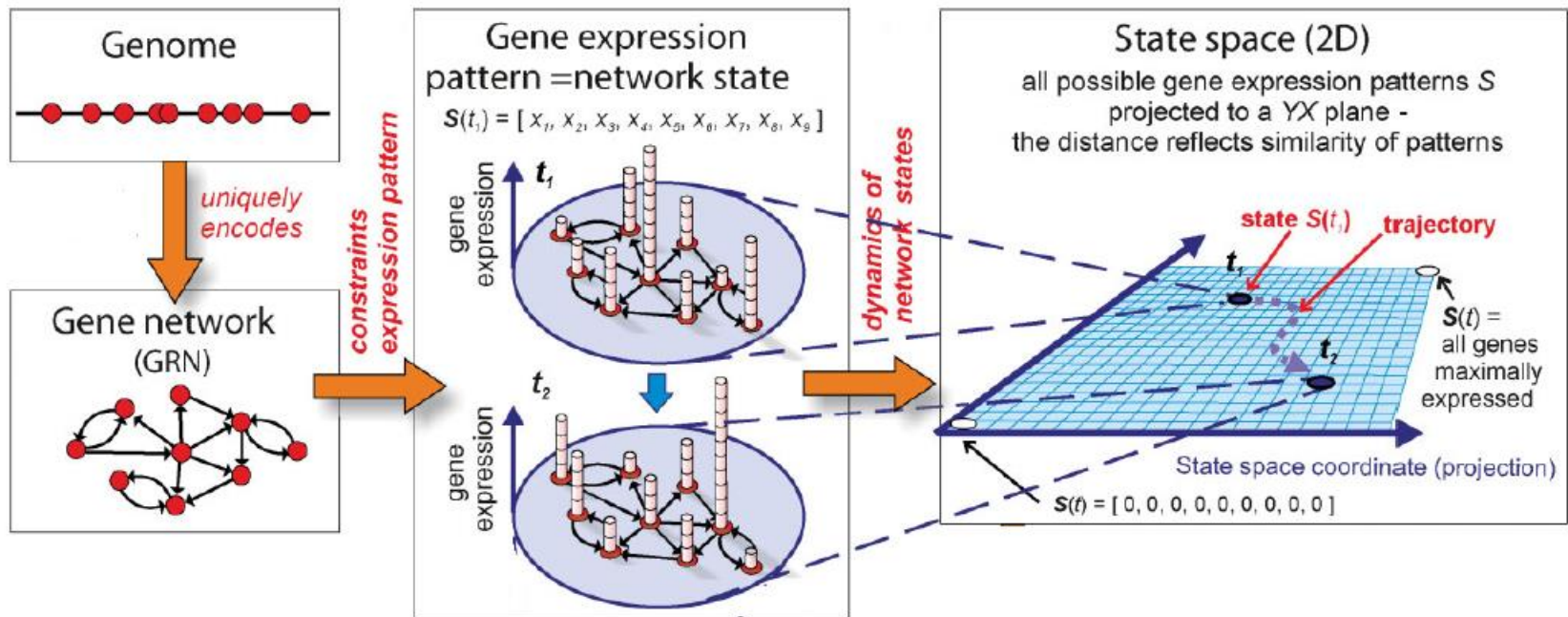
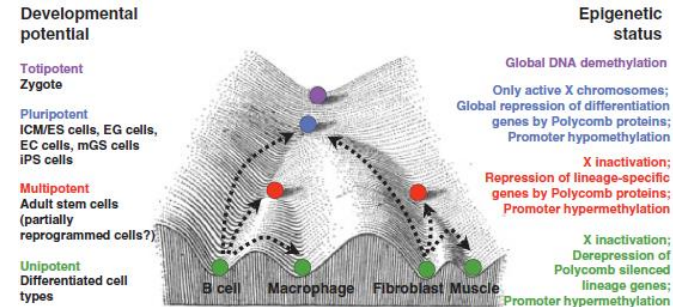
— 予後不良群に特徴的  
— 両者共通  
- - - 既知PPIあり





# 細胞分子ネットワークの状態空間

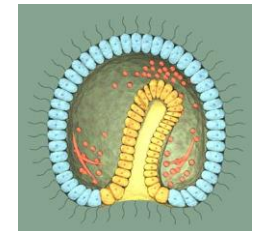
- 生命進化・発生・疾病はすべて分子ネットワークの構造変換である (生命系の基軸)
- 細胞分子ネットワーク (CMN) の状態空間
- 細胞分子ネットワークの結線構造はゲノムにコードされて生得的/固定的である
- CMNの中でどの遺伝子が発現しているかは、細胞型および状態によって異なる



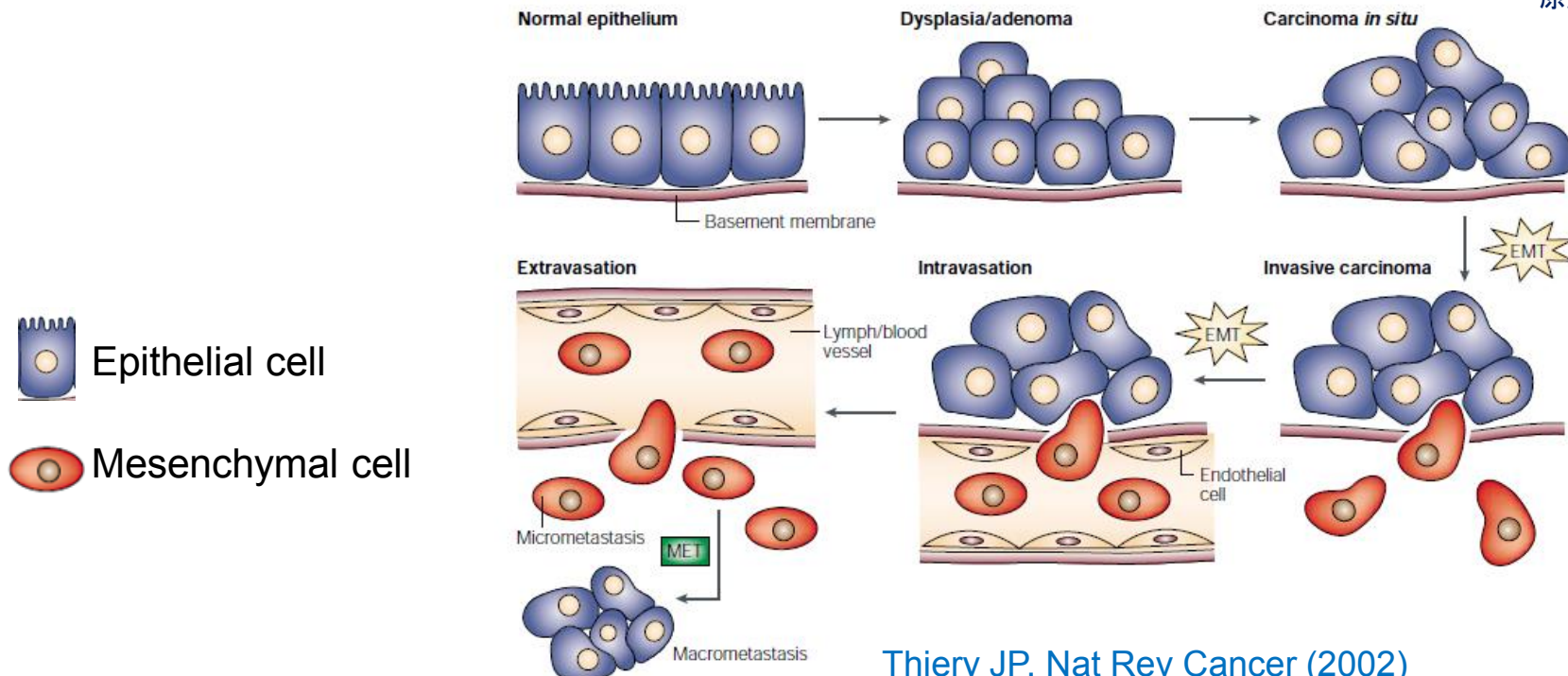


# 上皮間葉転換(EMT)

- Epithelial-mesenchymal transition (EMT)
  - non-motile, polarized epithelial cells, embedded via cell-cell junctions, dissolve their cell-cell junctions and convert into individual, non-polarized, motile and invasive mesenchymal cells
- EMT は生物に取っても重要な発生現象
  - Gastrulation (原腸陥入)、神経堤形成など
- EMT は、細胞分子ネットワーク発現の大域的变化



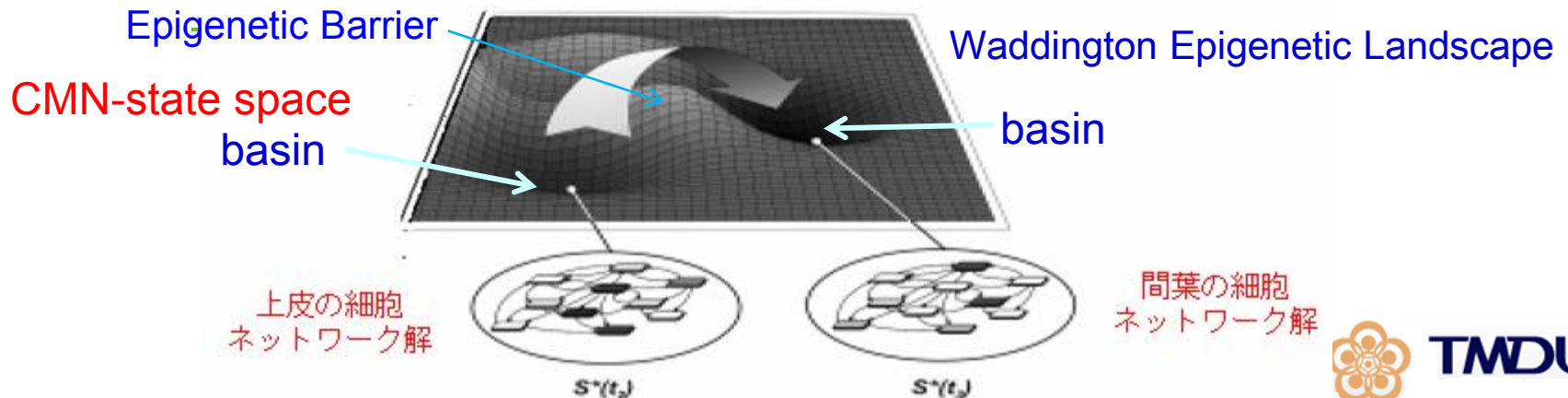
原腸陥入



# がんの浸潤転移

## 細胞分子ネットワーク空間の大域変異

- 細胞分子ネットワーク状態 (CMN状態空間)
  - 細胞分子ネットワークはいくつかの安定解を持つ
  - 各安定解は細胞種類に相当する(約220種類の安定解)、細胞ネットワークが作るCMN空間のエピゲノム地形(準ポテンシャル)「アトラクター状態」これらは多能性幹細胞から分化解である
- EMT (上皮・間葉転移) Epithelial-mesenchymal Transition
  - 細胞分子ネットワークの上皮安定解が間葉安定解へと移行
  - 全ネットワーク状態の転移であるから「相転移」といえる
- がんの転移とは相転移
  - 安定解を不安定化し(ゲノム不安定性)、障壁を越えるためにはゲノムの不安定性が必要である。



# EMTの定量的Waddington エピゲノム地形での軌跡

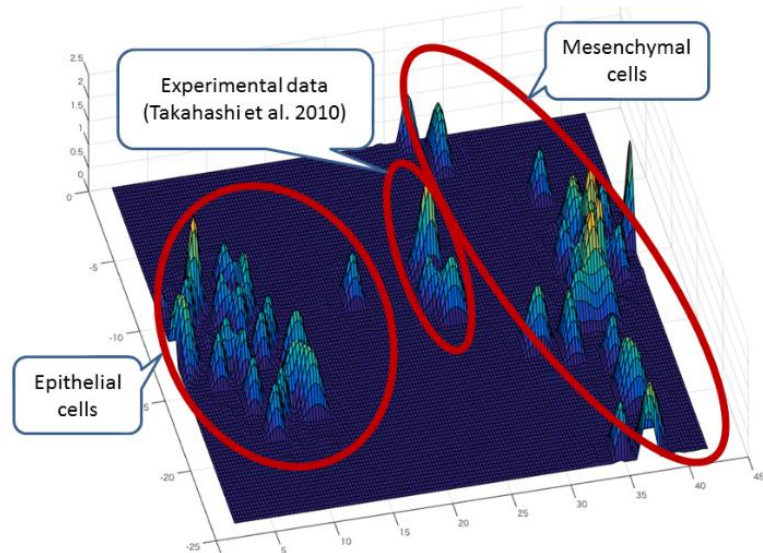
## Article

Network biology approach to epithelial–mesenchymal transition in cancer metastasis: three stage theory

Hiroshi Tanaka<sup>1,2\*</sup> and Soichi Ogishima<sup>2,†</sup>

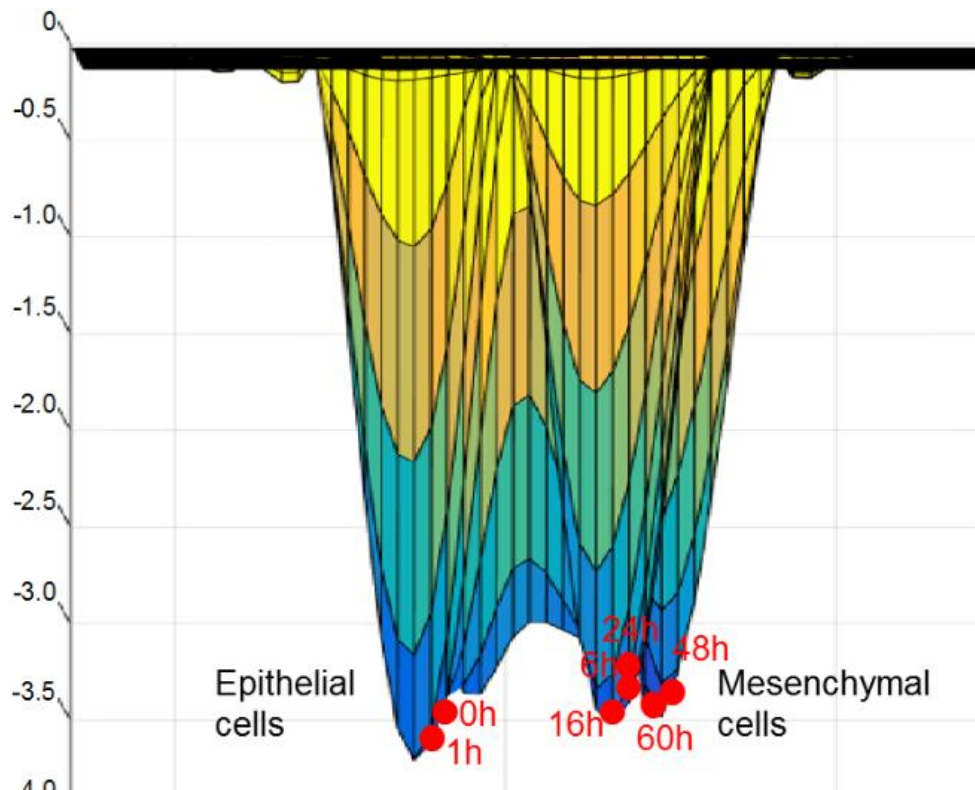
<sup>1</sup> Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan  
<sup>2</sup> Department of Bioclinical Informatics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan  
<sup>†</sup> These authors contributed equally to this work.  
<sup>\*</sup> Correspondence to: Hiroshi Tanaka, E-mail: tanaka@bioinfo.tmd.ac.jp

Tanaka H, et al., J. molecular cell biology 2015

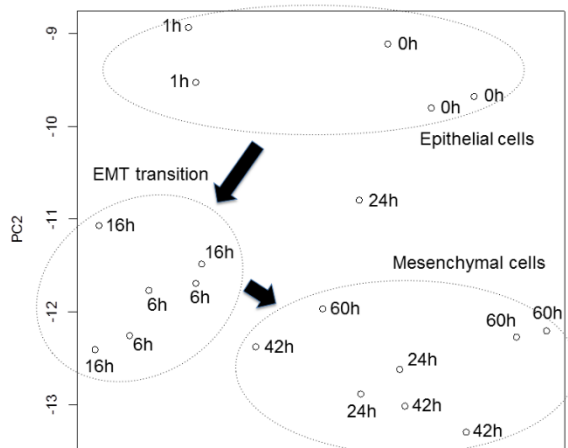


細胞ネットワーク空間の確率分布

## 網膜色素上皮細胞のEMTのqWEL (データはTakahashiら)



## がんの EMT過程



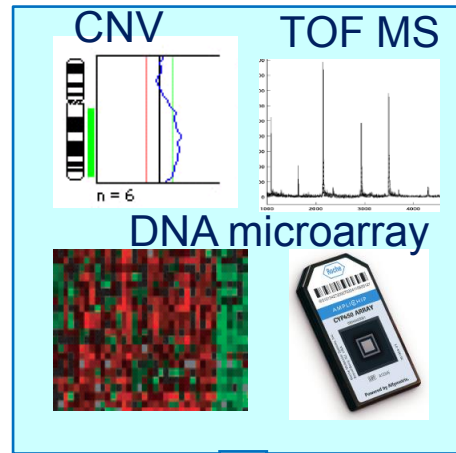


# Precision Medicineとは 分子システム医学のことである

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

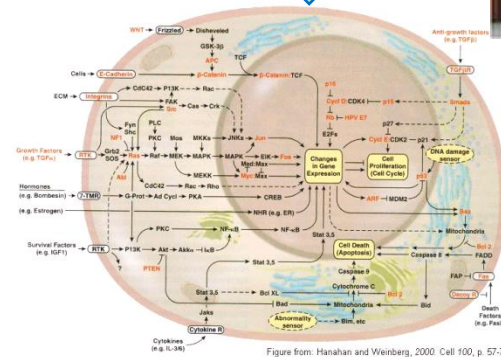


Figure from: Hanahan and Weinberg, 2000. Cell 100, p. 57-70

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



# Precision medicineの将来

- Precision Medicineとはシステム分子医学の実現である
- 疾患発症予測
  - Population型BioBankより、＜遺伝的要因と環境・生活要因(exposome)の相互作用＞を正確に取り入れた発症予測理論・手法が可能になる
- 罹患疾患システム認識
  - 疾患型BiobankのBigData、患者の表現型と分子型を融合した「統合データベース」から層別化パターンが類似症例の検索機能によって分類される
  - Panomicsによる「層別化パターン特異的分子ネットワーク病態」の認識
  - それに基づいた予後予測・薬剤治療選択
- 個別化予防
  - 継続的オミックス/生理変量モニタリングと時系列解析による＜先制医療段階＞の発見と予知—個別化予防の実施

# 医療ビッグデータ 創薬のパラダイム変換

～RCTからReal World Dataへ～

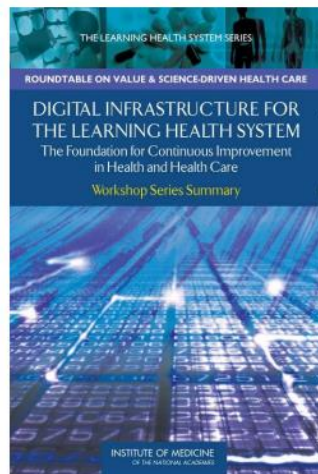
# 新しい米国の医療情報システムの概念 「学習する医療システム」

## Learning Health System/Rapid Learning

- 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*





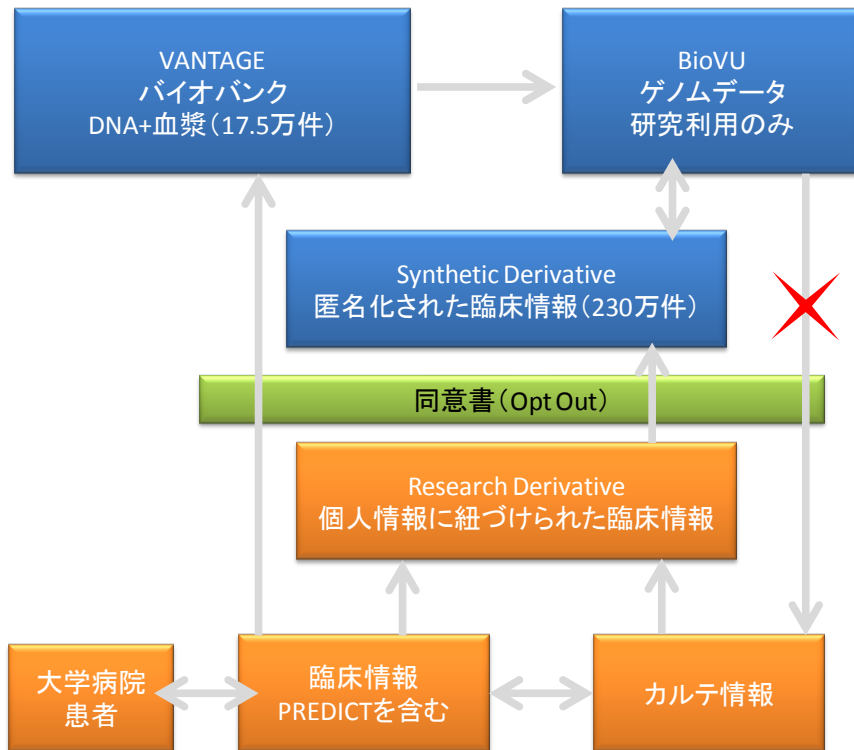
# 「学習する医療システム」 Learning Health System

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

# 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。

# LHS の効果と実装

- 大規模LHSの効用 17年⇒17ヶ月
  - 治験の対象者を調べる
  - 感染症の伝染状況の把握
  - 副作用の把握
  - 生理的特性のある患者の投与量変更
- 医療施設やネットワークでの実践
  - Kaiser-Permanente の患者を使った  
Vioxxの副作用調査（200万人・年データ）
  - その他にMayo Clinic, Intermountain Health, Duke, Cleveland Clinic, NCI caBIG (Cancer Biomedical Informatics Grid),

# 新しい臨床研究のパラダイム

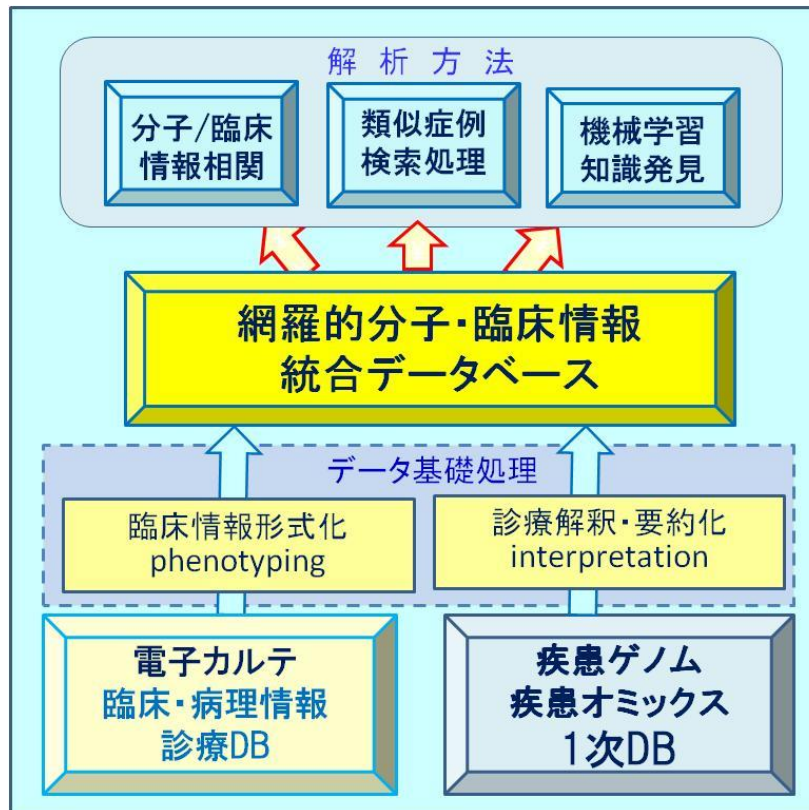
## Real World BD2K

- 大半のRCTは医療の外に人工的な環境
- 「標本」からの「推測」の概念
- 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」
- しかし、新たなパラダイムを構築するまで時間・研究が必要
- 我が国の戦略 段階的移行 BioBankの利用段階



# 統合臨床オミックスデータベースの 基本構造

統合臨床オミックス・データベース  
(integrated Clinical Omics Database)



医療施設にしろゲノムコホートにしろ、LHSの中軸に統合データベースが必要である

## 1. データ基礎処理部

電子カルテから必要な情報をphenotypingして所定の形式に  
分子情報はゲノムはvariant call、オミックス情報はsignature情報を中心に格納する

## 2. 統合データベース本体

どのようなデータ形式か検討の必要  
RDF化やi2b2方式など

## 3. データ解析部

層別化パターンの同定のための解析が第一  
分子・臨床情報相関解析、類似症例検索、機械学習の各システムを開発する。

さらに、これらの統合データベースをVirtualに統合して、医療・創薬の情報基盤とする

# 症例別の分子/病理/臨床情報の画面展開

個別症例  
詳細情報

icOD Integrated Clinical Omics Database  
HOME 臨床オミックス解析 症例一覧 マイクロアレイ解析ワークフロー

アカウント情報 ログアウト

疾患  
すべて  
肝臓腫瘍  
大腸癌  
膵臓腫瘍  
食道癌

検索:  
キーワード:  検索

絞込み:  
疾患:  すべて  
カテゴリ: (選択してください)  
項目:   
値:   
条件:   
時期:

検索キーワード欄に単語を入力すると、文字列データ内にその単語が含まれる症例が抽出されます。



icOD Integrated Clinical Omics Database  
HOME 臨床オミックス解析 症例一覧 マイクロアレイ解析ワークフロー

アカウント情報 ログアウト

<<prev 238件中 : 1件 - 25件 next>> 25件

1: 0101032 Hepatocellular carcinoma (肝臓腫瘍)  
治療 肝S5重区域切除+S8部分切除  
病理 T=T1,N=N0,M=M0

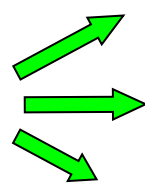
2: 0101062 Hepatocellular carcinoma (肝臓腫瘍)  
治療 肝S8重区域切除  
病理 T=T3,N=N0,M=M0

3: 01017840 Hepatocellular carcinoma (肝臓腫瘍)  
治療 肝右葉切除 (右門脈癌栓術後)

4: 01018257 Hepatocellular carcinoma (肝臓腫瘍)

組織所見

病理診断名: Hepatocellular carcinoma  
組織学的分化度: mod  
組織構造: trabecular>pseu  
発育様式 (eg): eg  
被膜形成 (fc): +  
被膜浸潤 (fc-inf):  
門脈侵襲 (vp):  
肝静脈浸潤 (vv):  
vp/vv:  
肝動脈浸潤 (va):  
胆管侵襲 (b):  
切除断端 (SM):  
pStage:  
非癌部所見:



二ヶ月戻る  
サマリ  
全データ  
主病名  
基本情報  
症状  
背景  
病歴  
生活歴  
検査所見  
血液  
生検  
画像診断  
Angiography  
US (超音波)  
CT  
PET  
MRI  
内視鏡検査  
上部  
下部  
治療  
内科的治療  
外科的治療  
放射線治療  
予後  
病理  
肉眼所見  
組織所見  
分子情報

1ヶ月  
3ヶ月  
半年  
1年  
3年  
10年

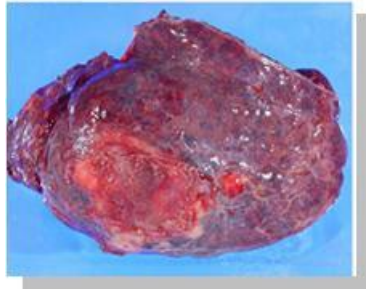
主病名  
基本情報  
症状  
背景  
病歴  
生活歴  
検査所見  
血液  
生検  
画像診断

01018257

CT

検査部位: 肝・胆・脾・膵

右肺下葉に索状影が見られる→陈旧性炎症/無気肺と思われる  
肝: S5+S8切除後。局所にLOAがあり→術後変化と思われる。  
S6辺縁に径21mm程単純にて低濃度、early phaseで淡い不均一な造影効果を示す、late





# 臨床オミックス解析

## 分子/病理/臨床 階層間 関連解析

2次元3層  
マップ

パラメータ設定

Pathome-  
Genome  
Map

臨床/病理/分子の情報を横断的に解析し  
病態と分子情報の相関関係を説明



# ビッグデータと医療 まとめ

- ゲノム・オミックス医療のビッグデータ
  - 個別化・層別化医療
  - 疾患バイオバンクと統合データベース
  - 層別化特異的パスウェイの同定に基づく医療
  - Precision Medicine ⇒分子システム医療
- 生涯継続的オミックスモニタリング
  - 疾病発症前状態（先制医療状態）の把握
  - 先制医療薬の開発への重点移行
- 遺伝要因と環境生活要因の相互作用
  - 個別化予防・3次予防

Precision Medicine 医学医療の的確性の増大

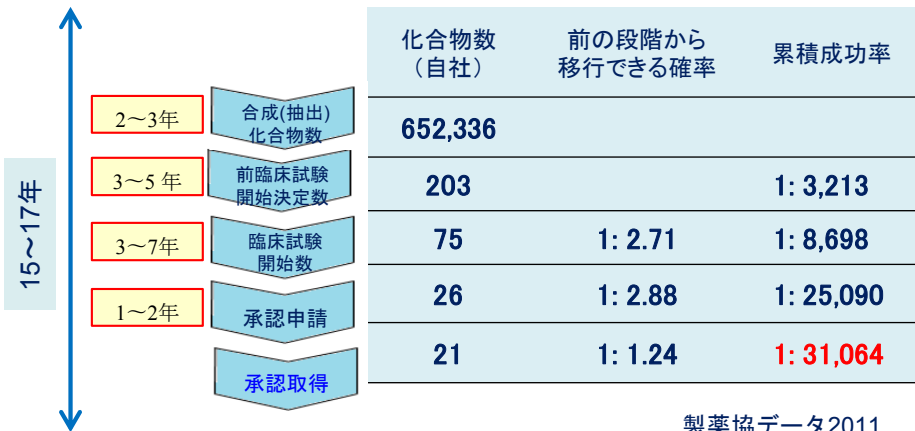
参加型医療：Intelligent Proactive Consumerの増大

# 創薬とビッグデータ データ駆動型創薬へ

とくにDrug Repositioningに関して

# 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



# 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)  
Signature of Differential gene Expression
- DB利用 : 発現刻印をquery, 順位尺度で類似性の高い順に化合物を提示

- 疾病特異的遺伝子発現(Disease-associated SDE)

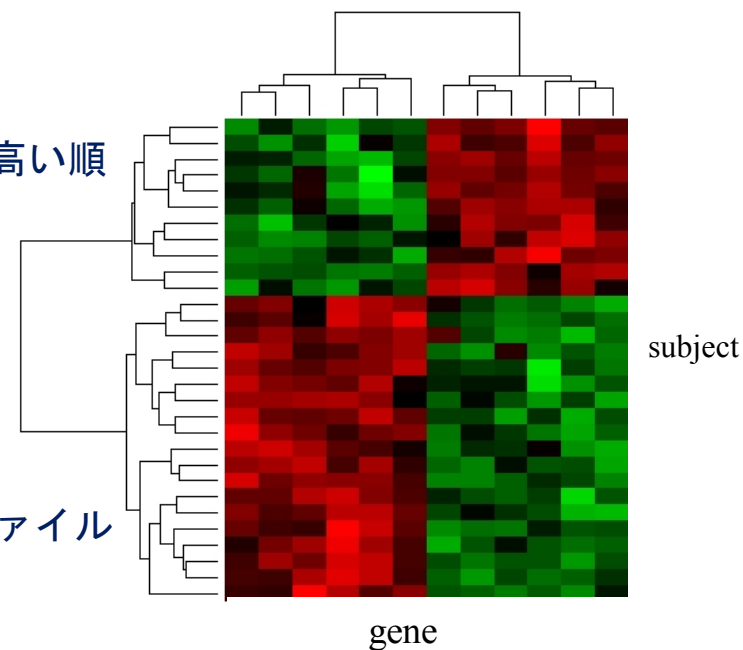
- GEO (gene expression omnibus),

- 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/薬剤特異的変化が基本 (第3世代)。

遺伝子発現プロファイル変化

≈分子ネットワーク変化



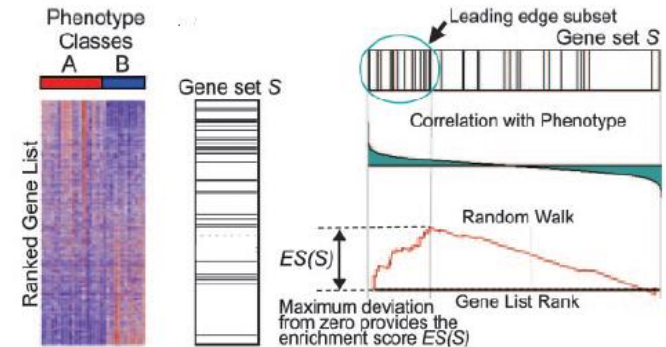
# 遺伝子発現プロファイルによる有効性予測

- 遺伝子発現シグネチャ逆位法 (signature reversion)

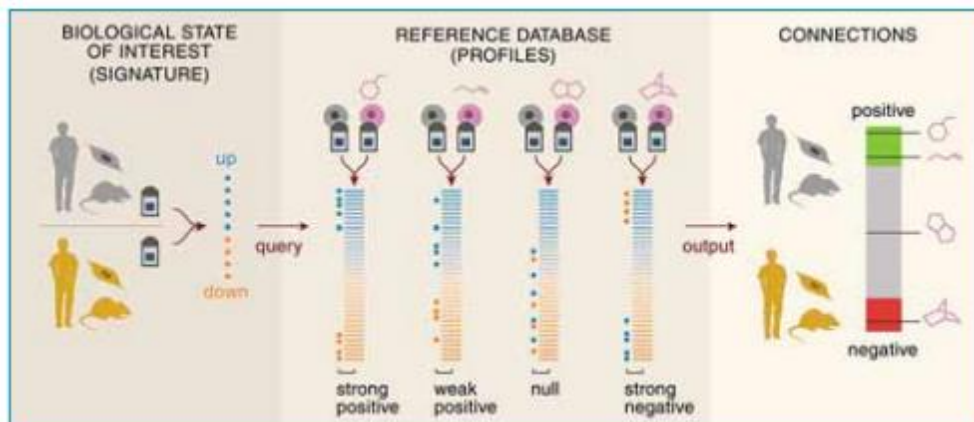
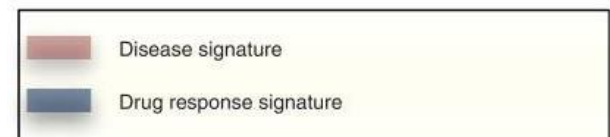
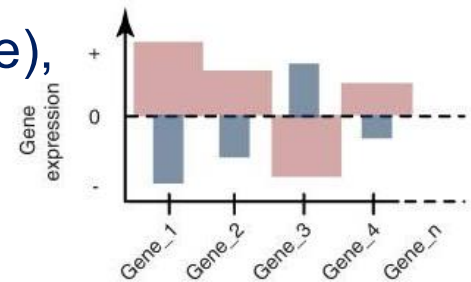
- 薬剤特異的遺伝子発現シグネチャ
- 疾患特異的遺伝子発現シグネチャ
- 有効性予測**：両者が負に相関する
- Non-parametric な相関尺度で評価

- 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

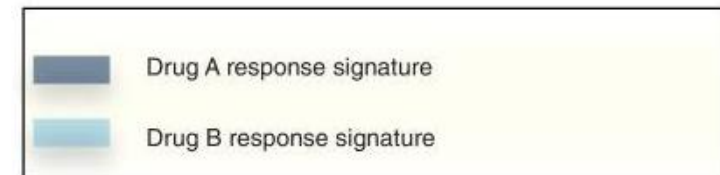
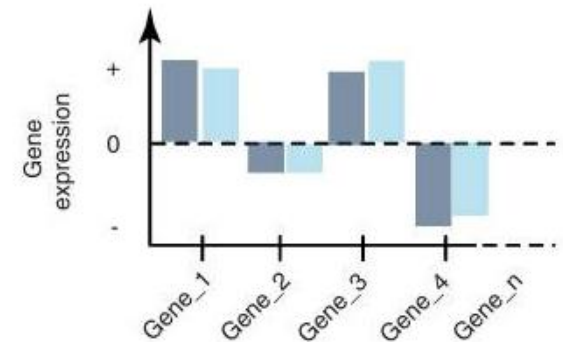


GSEA



# 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－薬剤間
  - 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
- 薬剤－疾患間
  - 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**





# Drug Repositioning

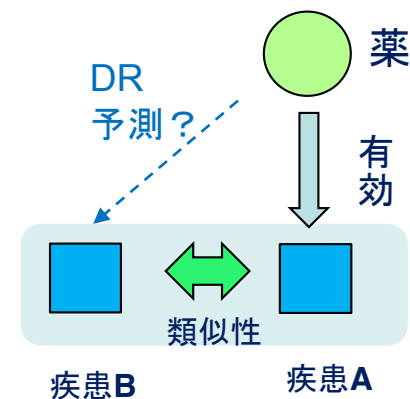
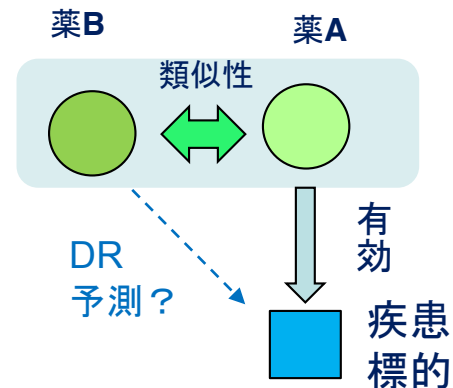
ヒトでの安全性と体内動態が十分に分かっている  
既承認薬の標的分子や作用パスウェイなどを、体系的・論理的・網羅的に解析することにより新しい  
薬理効果を発見し、その薬を別の疾患治療薬として  
開発する創薬戦略

## 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**

# 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



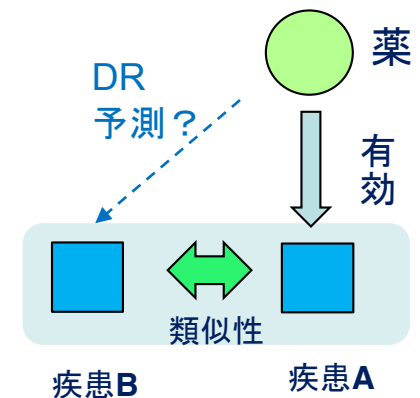
# 疾患ネットワークによるDR

## 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 疾患ネットワークによるDR候補疾患の予測

- 既存の<疾患-薬剤>を決め  
近傍の疾患をDR候補とする

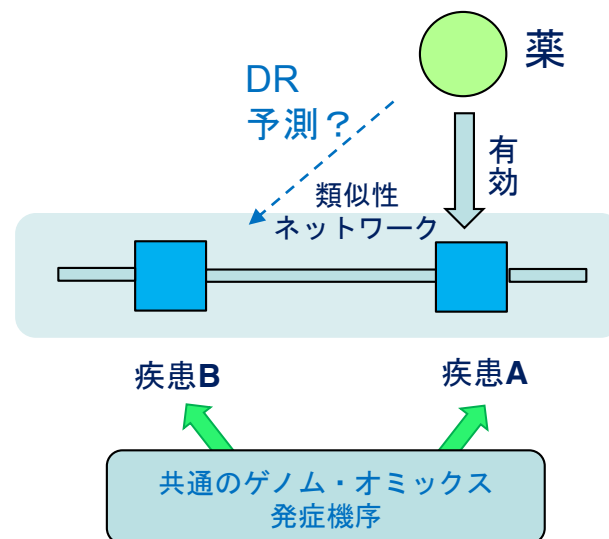


## 第3段階 シグネチャ逆位法による有効性・毒性の推定

- シグネチャ逆位法、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
  - 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

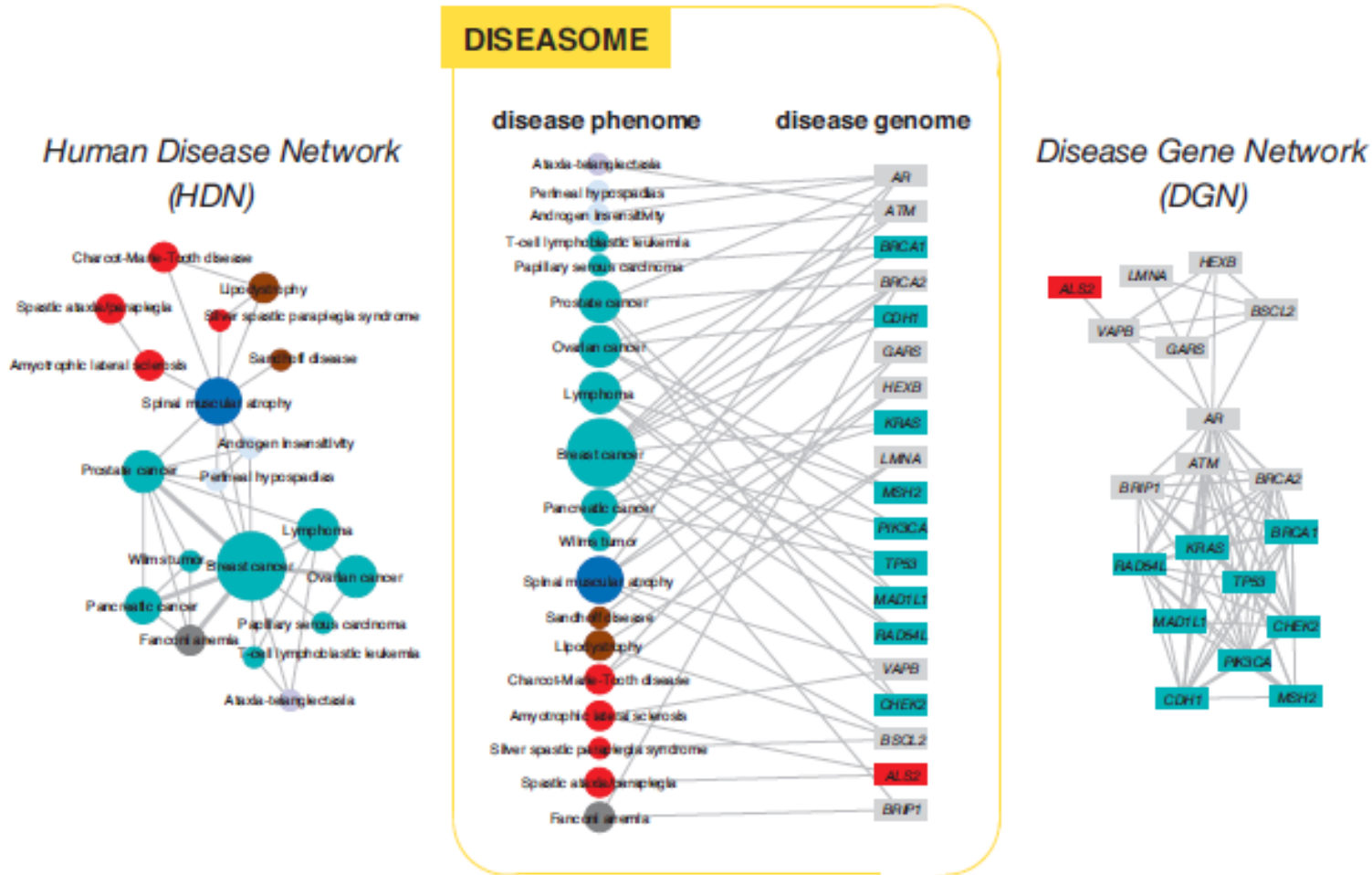
# 第1世代型

## 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

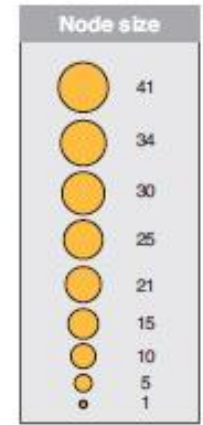
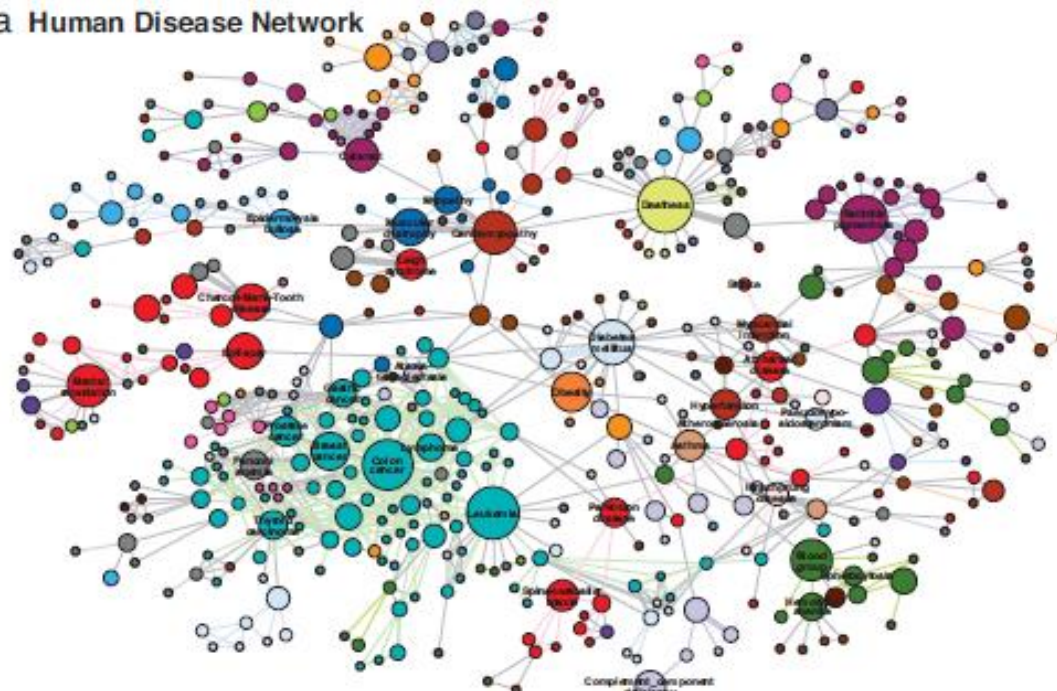


# 疾患 ネットワーク (HDN)

**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

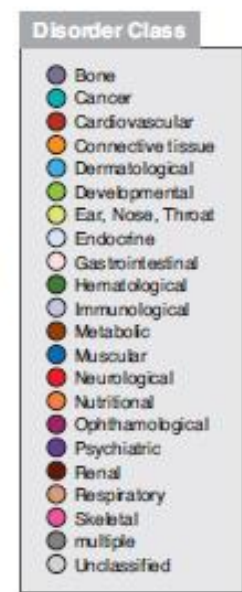
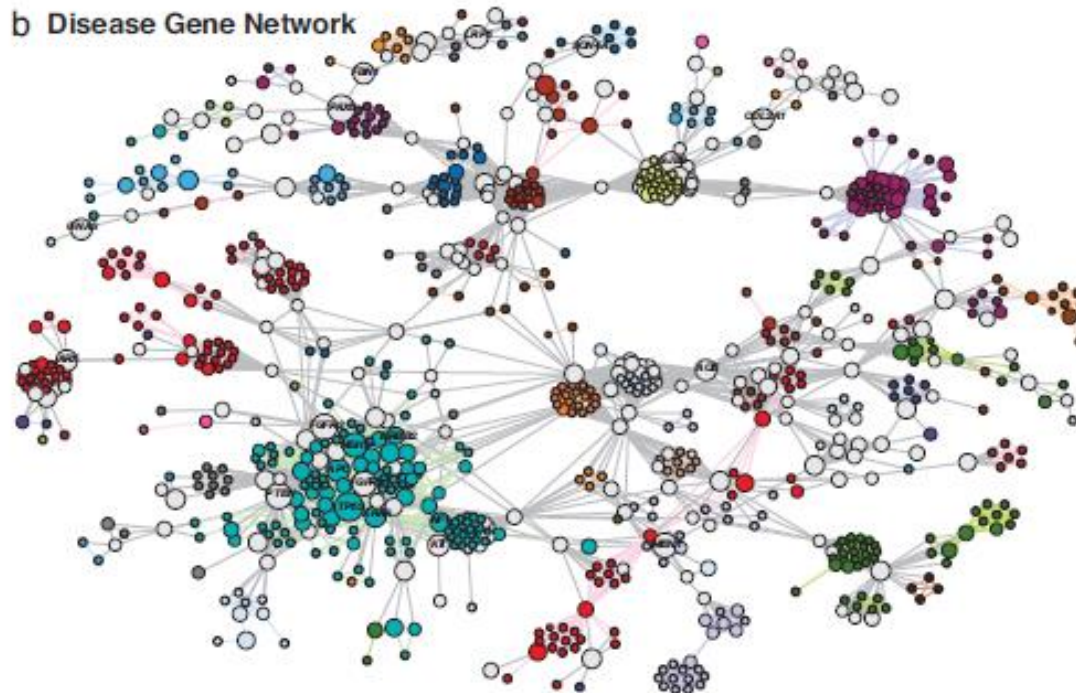


# 疾患遺伝子 ネットワーク (DGN)

**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



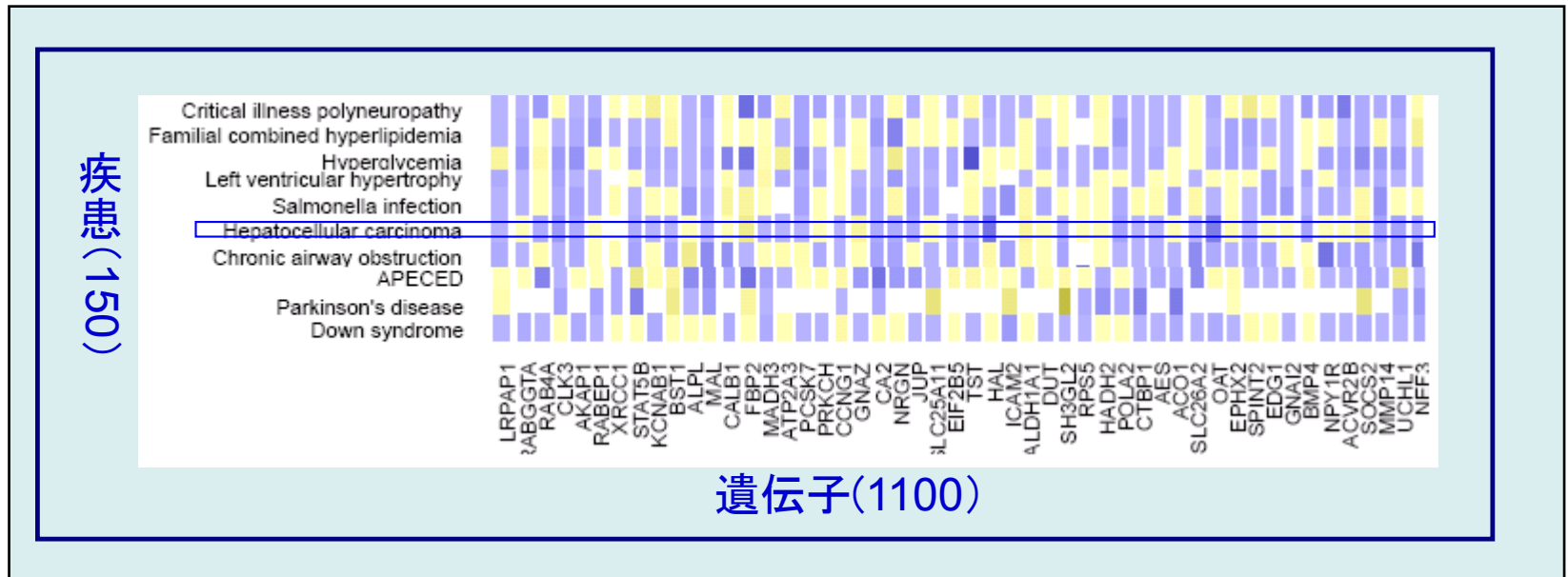
# Diseasomeを巡る状況

- Mendel疾患、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)

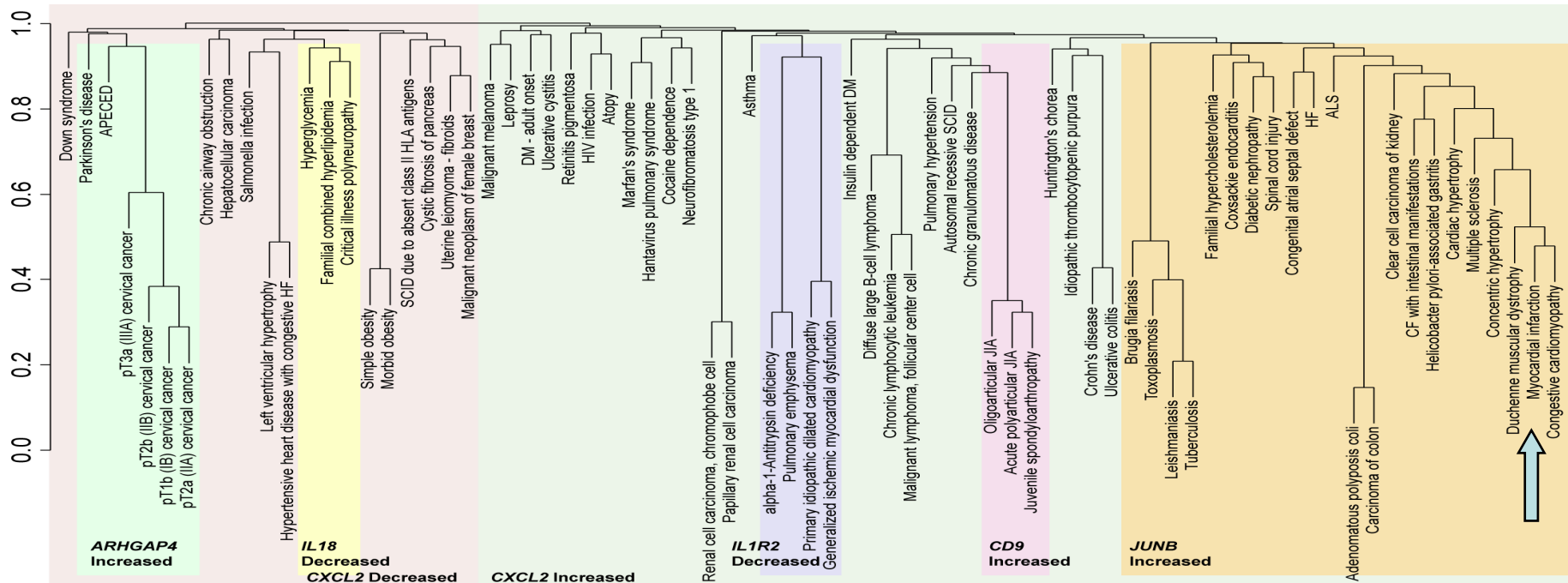
- 遺伝子発現DBのGEO (Gene Expression Omnibus) 利用
  - 約20万のサンプル
- 疾患名は注釈文より用語集UMLSを用いて抽出
- 疾患ごとに多数の遺伝子発現パターンを平均化





# 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)

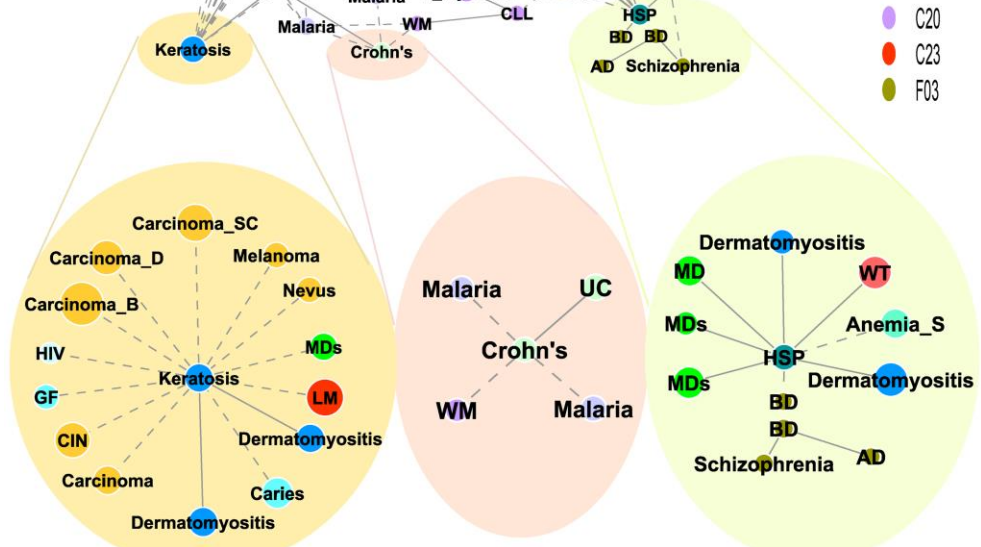
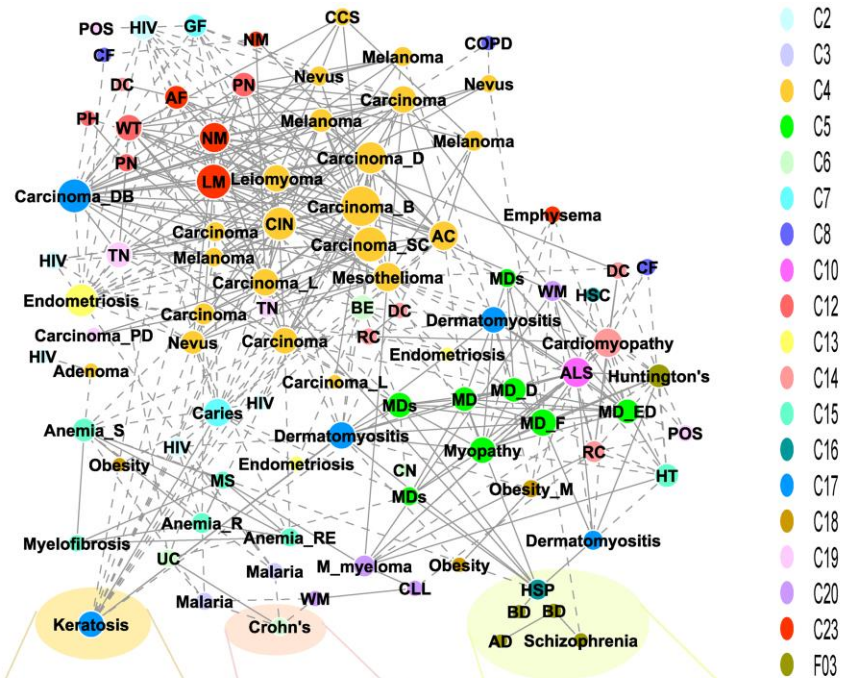
Crohn's disease

⇒ 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)

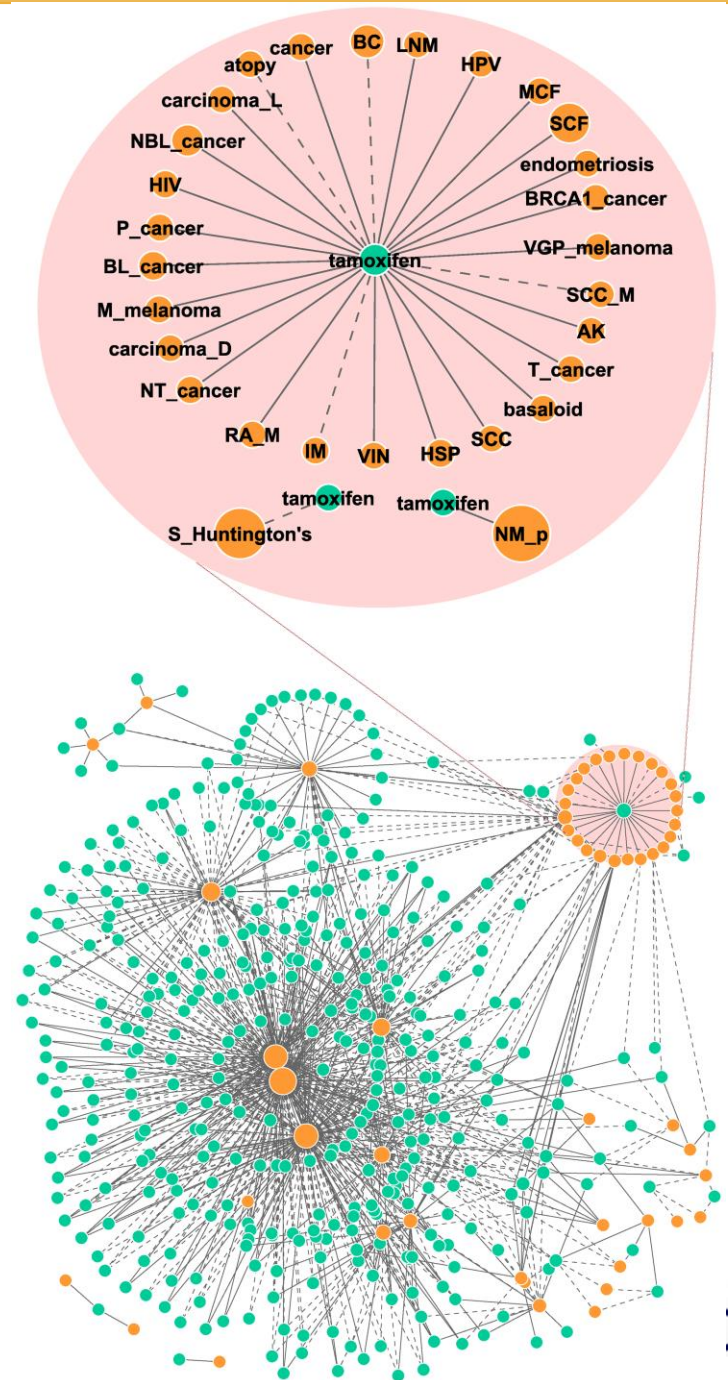
⇒ アトピー,

⇒ マスト細胞分泌抑制、  
アレルギー抑制

(正の値をもっている)

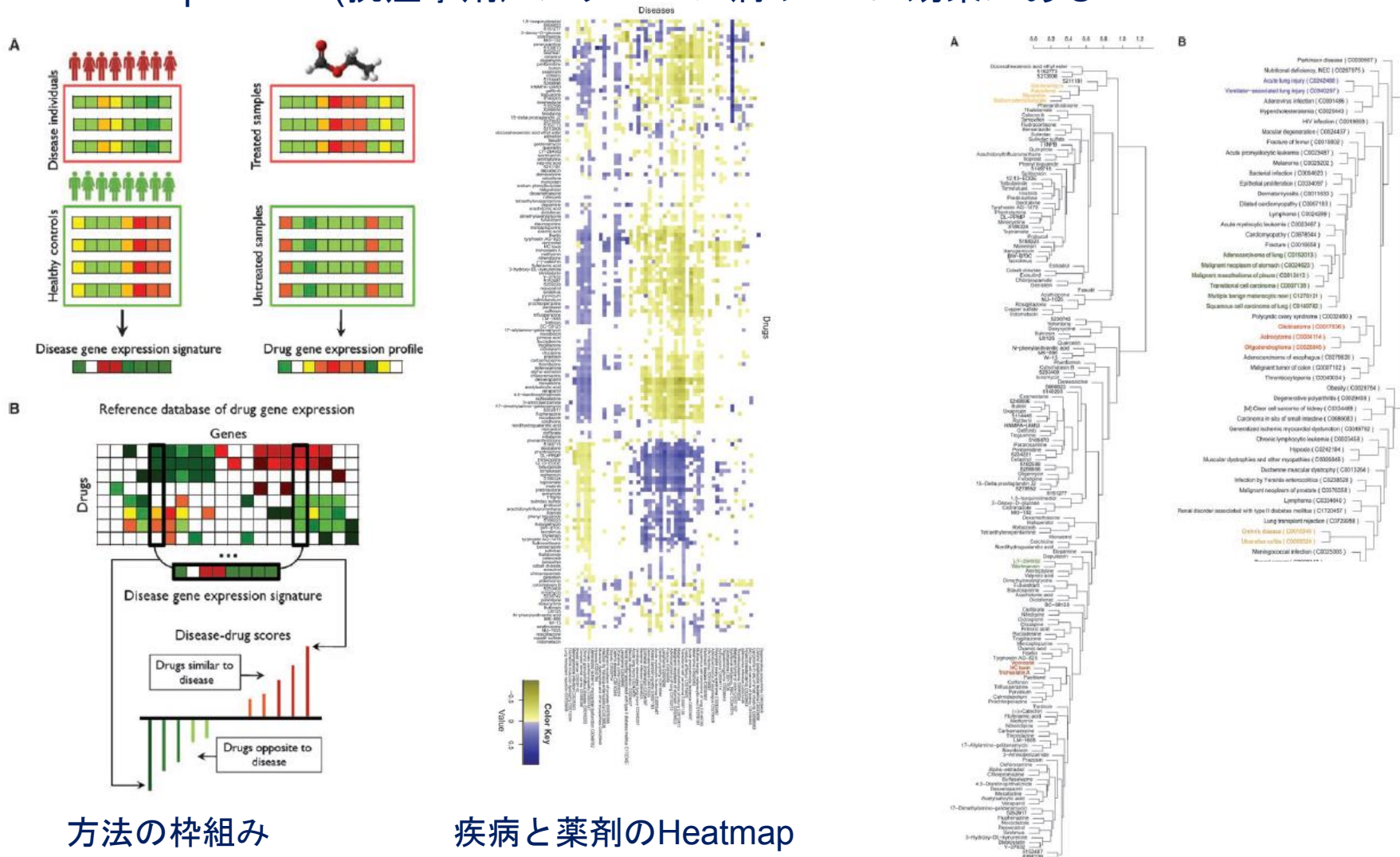
副作用の予測

⇒ carcinogenic



# ButteらのDR方法論

100疾患と164薬剤を選び発現プロファイルを比較。53疾患有意に残る  
Topiramete(抗痙攣剤) がクローン病やIBDに効果がある

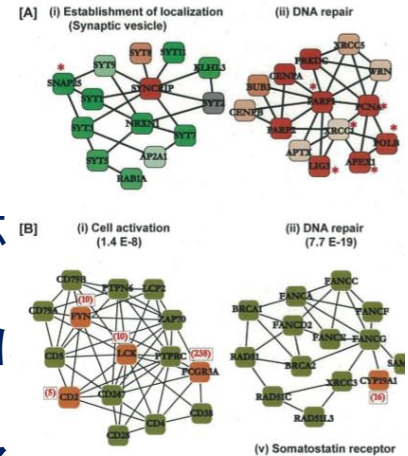


方法の枠組み

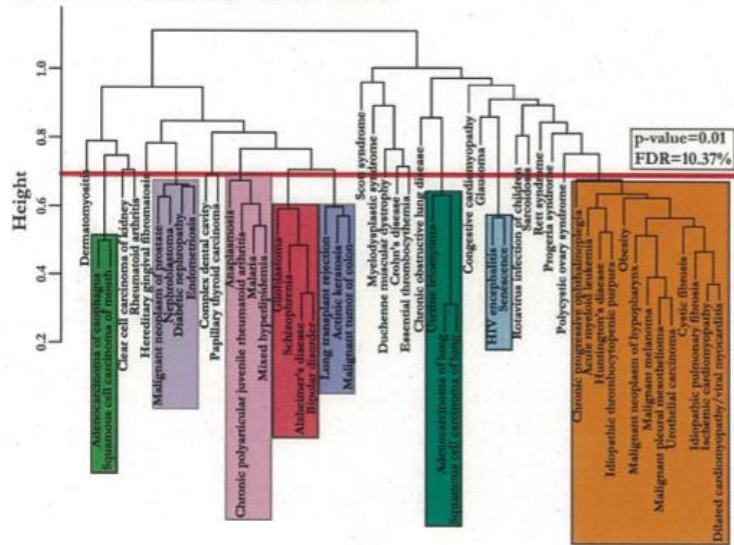
疾病と薬剤のHeatmap

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

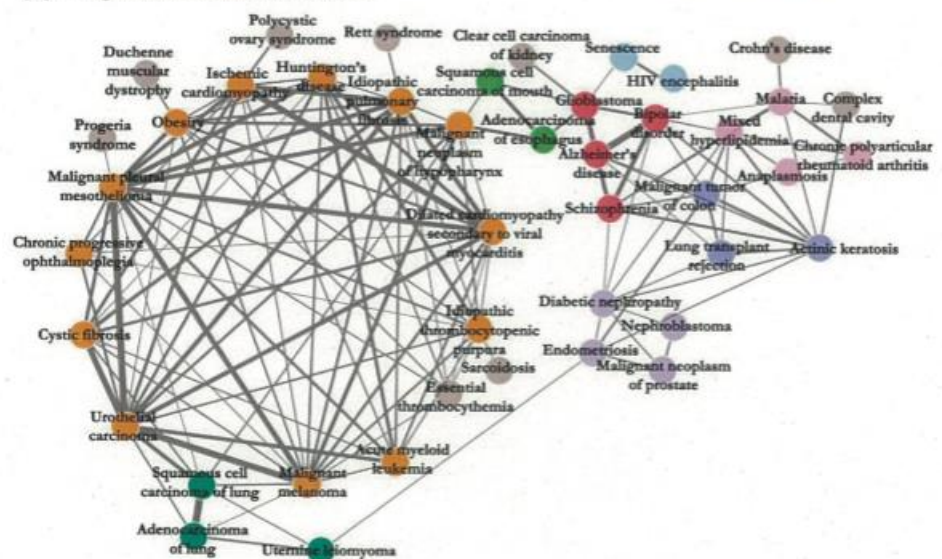
- 遺伝子発現プロファイルを直接使うのではなく 4620Moduleに分解したPPIネットワークでの疾患での平均発現変化をつかう
  - 遺伝子発現プロファイルより疾病によって変化するmoduleを調べる 病気に対するPPIの応マラリアとクローン病
  - moduleの遺伝子の変化を平均して遺伝子の代わりに moduleの発現平均スコアを用いる
- 疾患の大半を占める <共通疾患状態シグネチャ>



[A] Hierarchical relationships between diseases



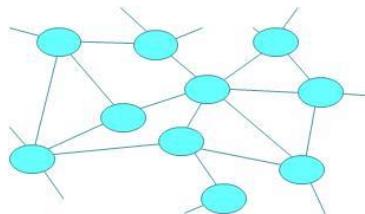
[B] All significant disease correlations



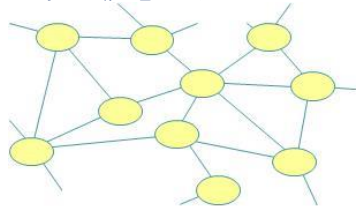


# 疾患ネットワークと薬剤 Projection Map

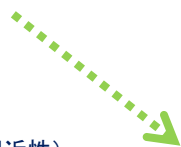
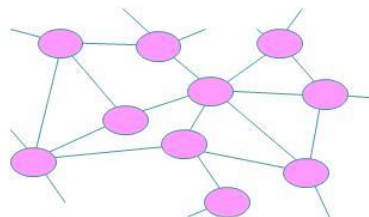
第1世代疾患ネット (原因遺伝子親近性)



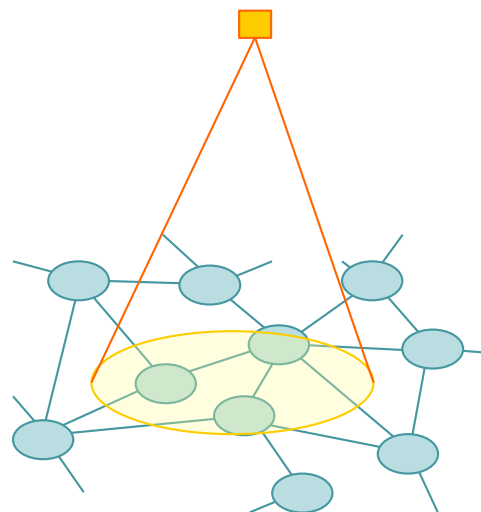
第2世代疾患ネット (OmicsProfile親近性)



第3世代疾患ネット (Pathway親近性)

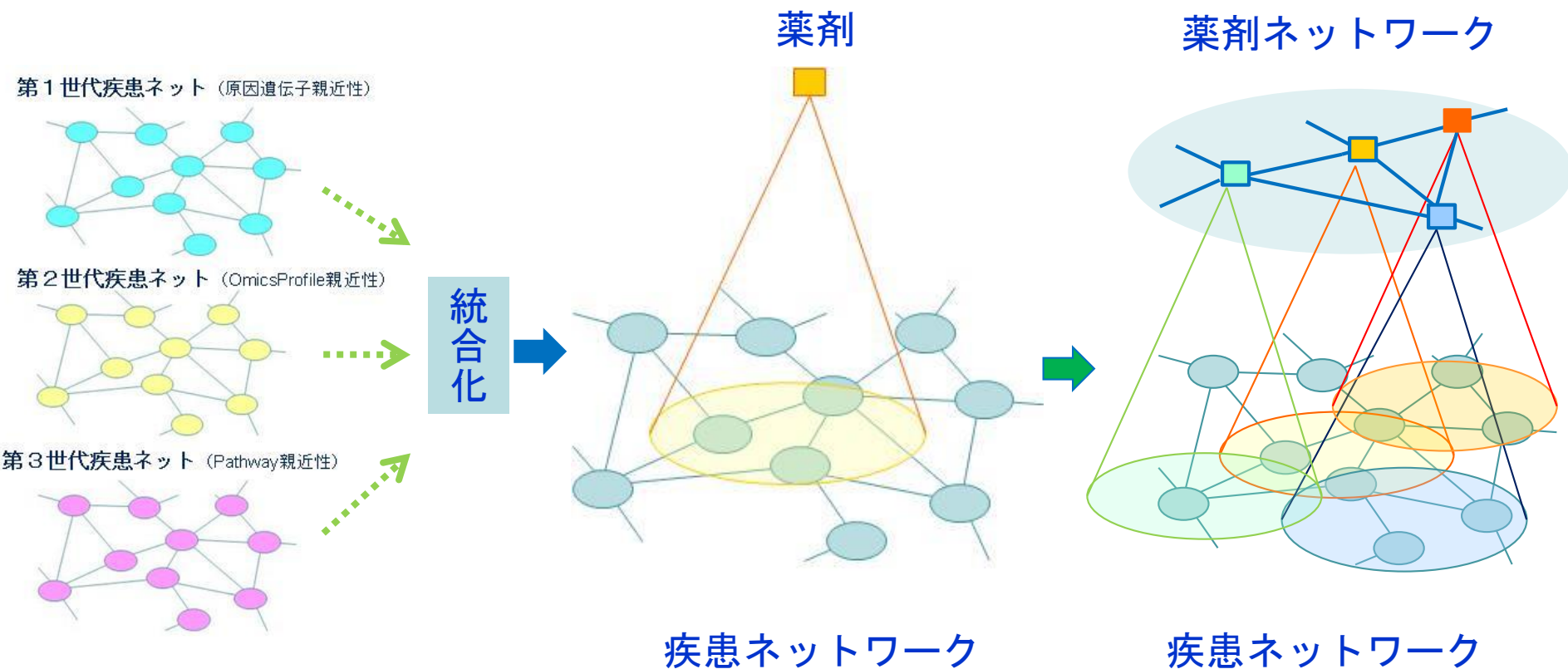


統合化



# 疾患ネットワークとDrug Projection Map

## DR informaticsの構築



# ビッグデータとprecision medicineによる医療と創薬の将来

- 網羅的分子情報を含めた医療ビッグデータの増加による医療の的確性の増大
  - 多様な個別化・層別化パターンの詳細化
  - 診断・治療の的確性の上昇
- 臨床医学知識獲得・創薬のパラダイム変換
  - 医療ビッグデータが機序情報を含むため
  - Real Word Dataに基づいた知識獲得 (BD 2 K)
  - まずはゲノムコホート/Biobankの創薬過程での利用、その後、LHSへ



ご清聴ありがとうございました

