

BioMed Pharm Data Scientistの 新しい役割

東京医科歯科大学 生命情報学

東北大学東北メディカルメガバンク機構

田中 博



医療ビッグデータから知識へ

医療ビッグデータ

1) 網羅的分子情報

- ゲノム・オミックス医療 分子情報
- 個別化医療/予防バイオマーカー
- システム分子医学
 - 患者パスウェイ情報など

新しいタイプの
医療ビッグデータ

2) 臨床診療情報

- 従来型の医療情報
 - 臨床検査、医用画像、処方、処置など

旧来のタイプの
医療ビッグデータ

3) 社会医学情報

- Population医療・医療政策

新しいタイプの医療ビッグデータ

- ゲノム・オミックス医療における網羅的分子情報のビッグデータ
 - **“Small Big Data”**
 - 1個体に関するデータ数が膨大
 - 全ゲノム配列、x30で100Gbpなど
 - 遺伝子発現プロファイル～4万probes
- 従来の医療情報のビッグデータ
 - **“Big Small Data”**
 - 1個体 医療情報・疫学調査 数十項目
 - サンプル数が多数

医療ビッグデータの流れ

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

クリニカル・シーケンシングの普及など

全ゲノム解析 (WGS) 100Gb エクソーム解析 (WES) 6Gb

米国では数十の著名病院で実施

ゲノム・オミックスデータの蓄積

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

ゲノム医療の実践

医療ビッグデータ

ゲノム・オミックス医療の現状

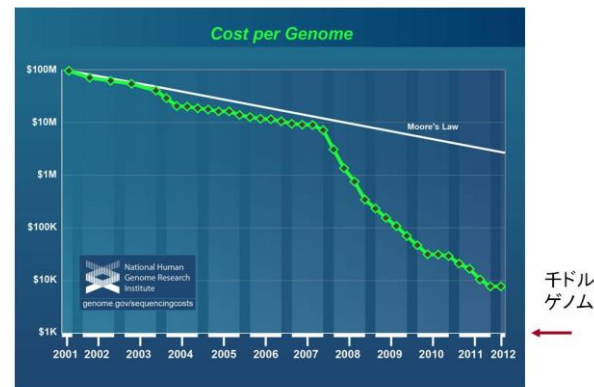
次世代シーケンサのインパクト

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



	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



DNA Sequencing Cost: the National Human Genome Research Institute

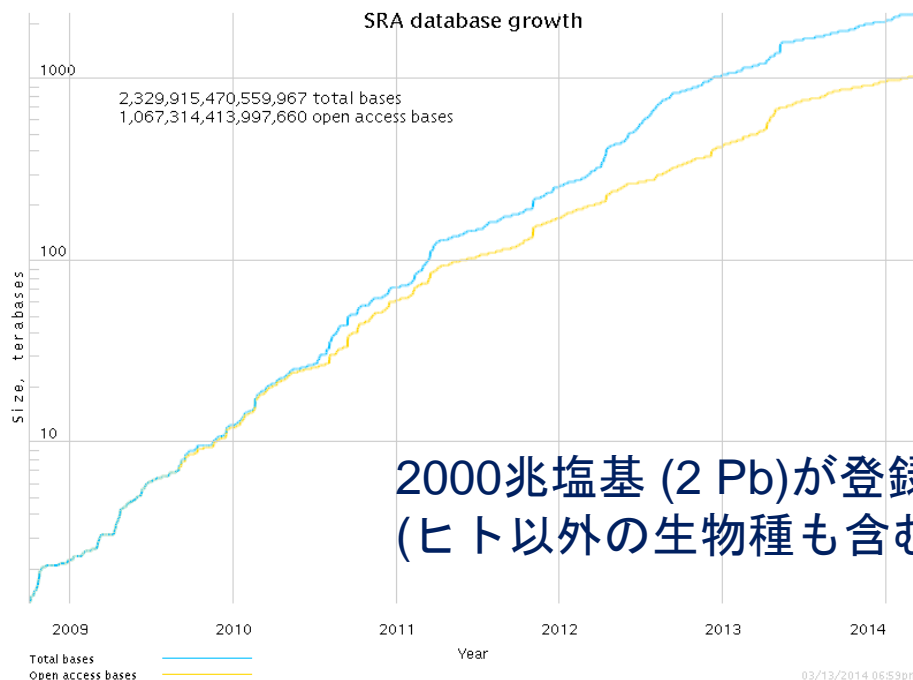
HiSeq X システム 10台構成



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!

次世代シーケンシングの 配列データベース

- 米国NCBIのSequence Read Archive (**SRA**)が代表的
 - <http://www.ncbi.nlm.nih.gov/Traces/sra/>
- 欧州EBI
 - **ENA**(Europe Nucleotide Archive)のなかにSRA
- 我が国
 - **DDBJ**にDDBJ sequence read archive (**DRA**) 日本語のインターフェース



2000兆塩基 (2 Pb)が登録されている
(ヒト以外の生物種も含む)

ゲノム・オミックス医療の概念

ゲノム・オミックス医療の3世代

第1 ゲノム医療

- 生得的ゲノム情報の変異・多型性

第2 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）医学

- 細胞分子ネットワークの歪み/構造変化

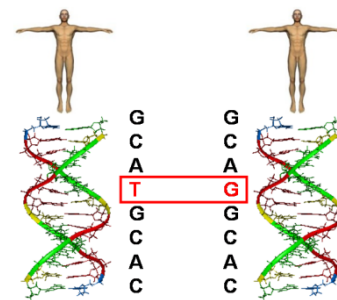
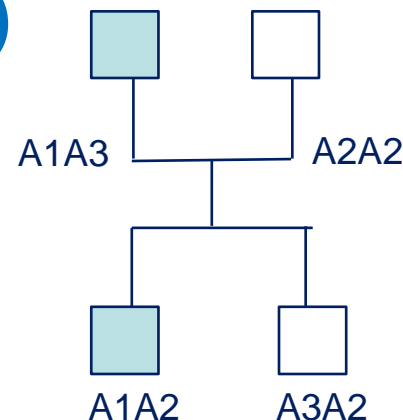
H.Tanaka: Genome and Omics Medicine

—Principles, Clinical Implementation and BigData Approach—, Springer 2015

網羅的分子医学 第1パラダイム

ゲノム医療 (1990~現在)

- 「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 (constitutional genome)は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、Positional Cloning
 - 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マーカー 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
 - HAPMAPプロジェクト、1000ゲノムプロジェクト



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

網羅的分子機序 第1世代 疾患関連遺伝子

- **生得的ゲノム変異・多型による疾患発症の生得的リスクの予測**

- 家系調査/Linkage解析
rare disease: 相対リスク大
しかし、いずれも遺伝難病で
治療手段なし
- 全ゲノム関連解析(GWAS)

common variant 仮説:

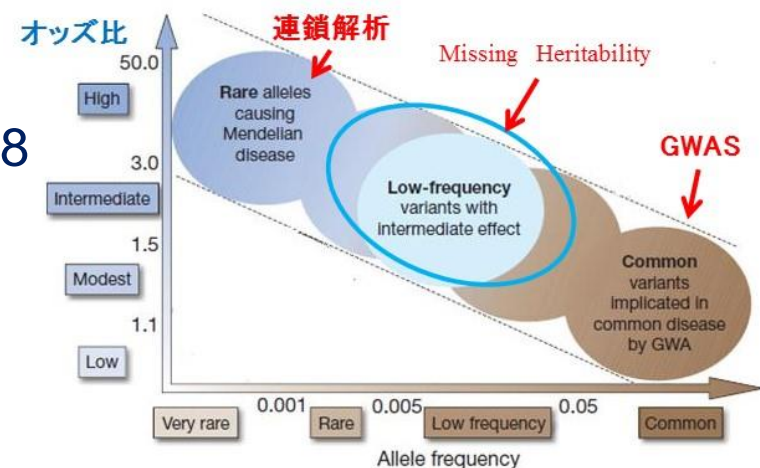
相対リスク小: 1.1-1.5

環境因子: 例, 喫煙肺がん相対リスク4.8

発症予測に効果低い: multiple rare variants 仮説

- 生得的ゲノム変異・多型の疾患予測力?
- 疾患の発症メカニズムなどに寄与

疾患	原因遺伝子・感受性遺伝子	座位
ハンチントン病	HTT(huntingtin) 遺伝子	4p16.3
筋ジストロフィー	DMD(dystrophin) 遺伝子	Xp21.2
嚢胞性線維症	CFTR 遺伝子	7q31.2
家族性大腸ポリポシス	APC 遺伝子	5q21-q22
色素性乾皮症	XPA 遺伝子 (Aグループ)	9q22.3
炎症性大腸炎	CARD1 遺伝子 など	16q21
統合失調症	COMT 遺伝子 など	22q11.21-23
2型糖尿病	CAPN10 遺伝子 / ADIPOQ 遺伝子	2q37.3 / 3q27
晩発性アルツハイマー	APOE 遺伝子 など	19q13.2
高血圧	AGT 遺伝子 / ACE 遺伝子 など	1q42-43 / 17q23
乳がん	BRCA1 / 2 遺伝子 など	17q21 / 13q12
慢性骨髄性白血病	BCR-ABL 融合遺伝子	22q11
腎臓がん	VHL 遺伝子	3p25.3



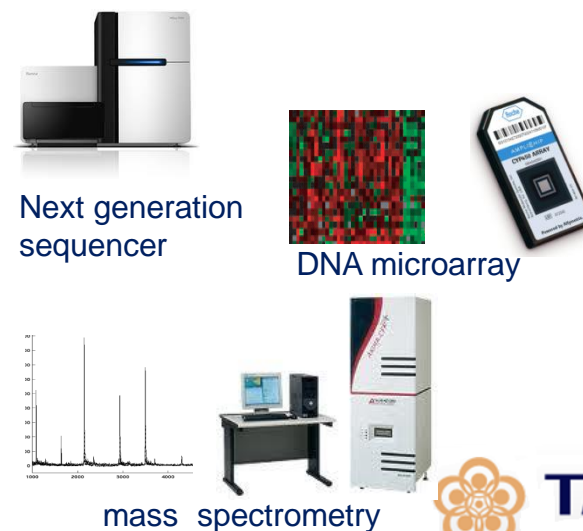
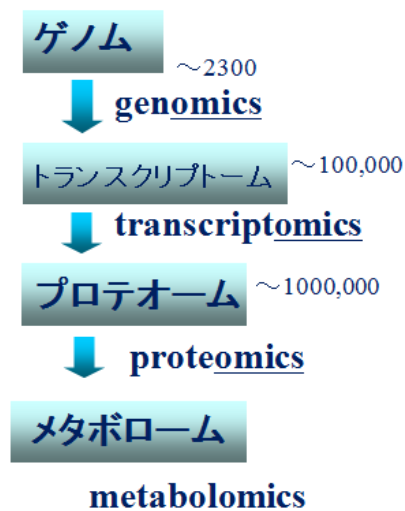
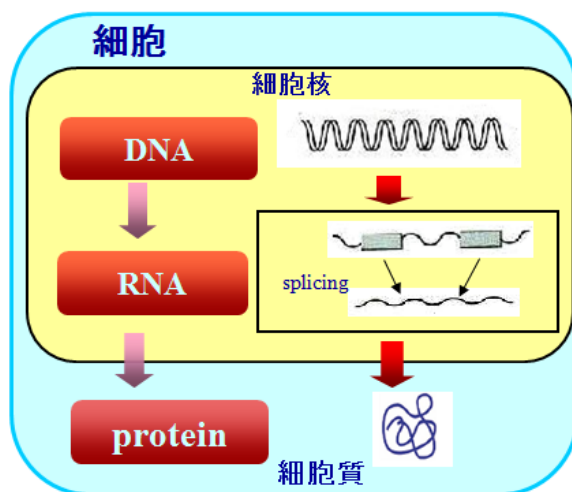
アレル頻度



IMCU

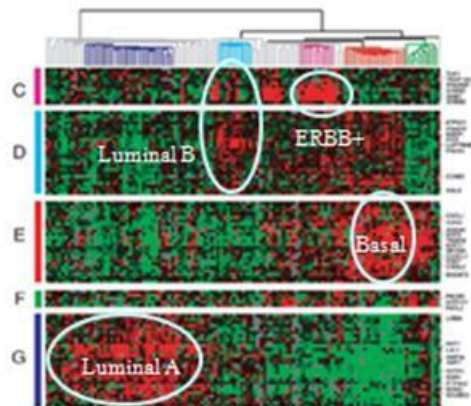
網羅的分子医学 第2パラダイム オミックス医療 (2000~)

- 「**後天的・疾病依存的に変化する体細胞変異**」
後天的体細胞ゲノム変化・オミックス変化に基づく医療
 - 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・メタボロームなど
 - 網羅的分子表現型 "molecular phenome"
 - 疾患組織や病態進行の段階に依存
- **疾患の進行度の評価**
 - Ongoing state of disease progression
 - 臨床症状や病理変化より早く変化→予測医療・先制医療



オミックス医療の実践

- 疾患オミックスプロファイルによる疾患の内在的分類
 - 「疾患オミックスプロファイルの乱れ」 “疾患の個性”
 - 網羅的分子病態像によるサブタイプ分類
 - 遺伝子発現プロファイル・血清プロテオームプロファイルなど
 - 臨床・病的な分類に対してintrinsicな分類（例,乳がん）
- 複数バイオマーカー（シグネチャ）による疾患予後予測
 - 遺伝子発現プロファイル（mRNA）アレイ OncoTypeDなど、RNA-seq
 - エクソソーム miRNA
 - Circulating Tumor Cell, 血中遊離mRNA・exDNA,
- GSEA分析やover-representation法によるパスウェイへ射影



intrinsic分類	ER	PgR	HER2	予後
Luminal A	ER(+) and/or PgR(+)	(-)	(-)	予後良好
Luminal B	ER(+) and/or PgR(+)	(+)*	(+)*	A型より不良
HER2 enriched	(-)	(-)	(+)	予後不良
Basal-like	(-)	(-)	(-)	予後不良

乳がんのオミックスプロファイルと亜型分類と予後・治療

米国における ゲノム・オミックス医療の臨床実装

ゲノム/オミックス医療を 取り巻く現状

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
業務として行われている。

- 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
- Cernerゲノム電子カルテ
 - 用語集CBOを整備
- 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
 - 東京医科歯科大学 iCOD
 - Mofit Cancer Center
 - Oracle HRI (Health Research Informatics) platform
 - UPMC(pittsburgh), MDアンダーソンでもゲノム情報の臨床実践

NHGRI Working Groupのリスト

Investigator	Institution	Major Projects
David Bick/David Dimmock	MC Wisconsin	Using whole genome sequencing to establish diagnosis in patients with currently undiagnosed genetic disorders
Erwin Bottinger	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
Rex Chisholm	Northwestern	Using pharmacogenomics evidence (from GWA genotyping) to guide prescriptions in primary care and assess risk for other conditions such as HFE/hemochromatosis
Charis Eng	Cleveland Clinic	Tumor-based screening for Lynch syndrome, endometrial cancer
Kelly Frazer	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
Gary Gibbons	Morehouse	• Exome sequencing of 1200 early onset severe African American hypertension cases and 1200 controls
Geoff Ginsburg	Duke	<ul style="list-style-type: none"> • Computer-based family history tool for perception of breast, ovarian • SLC01B1

ゲノム・オミックス医療の臨床実装の現状

第1世代が殆んどである

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
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
MayoClinic, 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

いくつかの米国病院のゲノム医療

Wisconsin大学病院

- Nic君での成功例
 - 原因不明の潰瘍、2歳で130回の手術、全エキソーム配列解析
 - 16000の変異よりXIAPの変異を同定、自己免疫の異常化に関係
 - 臍帯血による骨髄移植
- Genome sequencing program
 - 候補選択 (nomination)
 - 従来検査・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:
 - アメリカ病理学会 (CAP)およびClinical Laboratory Improvement Amendment (CLIA)によって証明されたLaboに外注検査
 - 解析: in-houseソフトで
 - 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)



Baylor医科大学

- Whole genome laboratory 設立
 - In-house シーケンシング/変異分析
 - CAP/CLIA認証検査室での配列解読
 - 執行部認証の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, Gesinger clinic など

メイヨクリニック

個別化医療 (Individualized Medicine)

- 全患者に全ゲノム配列解析(WGS)あるいはエキソーム配列解析(WES). **10万人患者 (病院診療圏患者のみ)**
- eMERGE consortiumのメンバー
 - phase I (2007-11)から参加、phase II (2011-15)では
 - **先制薬理ゲノム試験** : Preemptive PGx (Genotyping) testing
臨床の現場における (POC)
先制的PGxデータと臨床診断支援 (CDS) を電子カルテに統合
ゲノム支援治療
 - **RIGHT protocol: Right Drug, Right Dose, Right Timed Using**
 - 本年6月、Weinsilbaum博士を招聘



- 通常診療で困難な症例を
“Center for Individualized Medicine”
で分析・治療指針を提示する

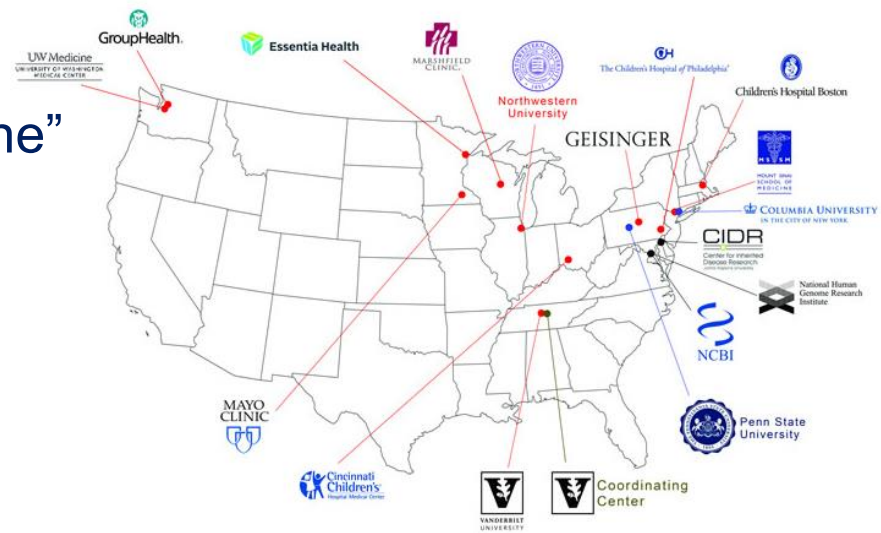
Clinical Sequencing

難治性がん

- 例 胆管がん

原因不明遺伝病

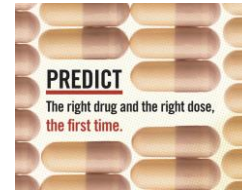
- “診断オデッセイ”



eMERGE計画 参加施設



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

Pharmacogenomic Resource for
Enhanced
Decisions in Care and Treatment



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

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

この患者の場合(*2/*2)プラスゲレル
(商品名エフィエント)に替えるか

分量を2倍にしると警告している

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

- **NACHR** :
 - National Advisory Council on Human Genome Research : NHGRIが主導
 - **NACHR 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,
 - 薬理ゲノムガイダンス(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)

ゲノムオミックス医療 ビッグデータベース構築の 動向

NIH「ビッグデータから知識へ」計画

“Big Data to Knowledge” (BD2K) initiative

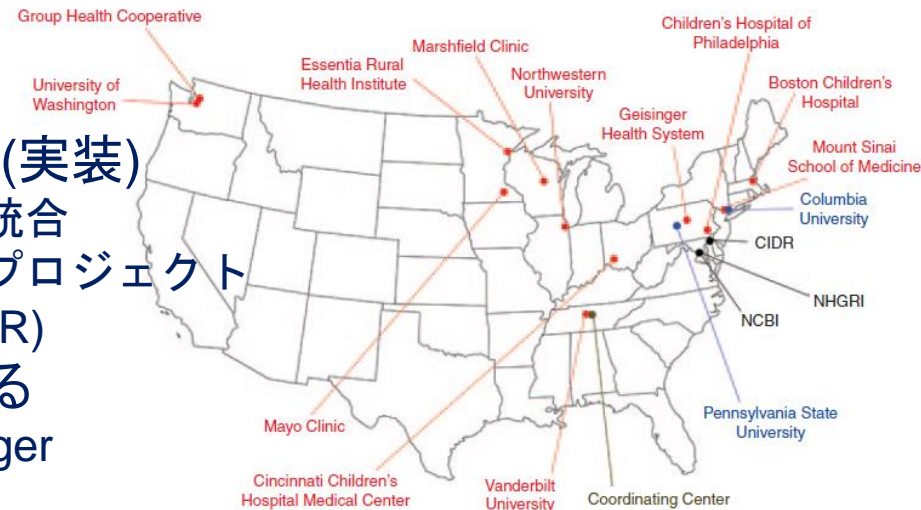
- **BD2K: “Big Data to Knowledge” Initiative 開始 2013**
 - データ科学のための副長官を任命 **Boune, PhD.**
 - 27のNIH研究所に大規模データの蓄積・利用
- 医療におけるデータ科学の全米COE創成
 - Center of Excellence in Data Scienceを全米11か所
- **Data Scientist 人材養成プログラム**
- **データ発見索引 DDI (Data Discovery Index) Consortium**
 - データベースカタログの発展・PubMedのDB版
 - UCSD: BioCADDIEを中心にDDI研究開発
- **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」
 - 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>

関連プロジェクト eMERGE

- 前史 : Nat. Ctr for Biomedical Computing : i2b2 形式の医療データ
- phasel (2007-2011)
 - 電子カルテを通して臨床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

Big DataとLearning system1

- **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
- **Cleveland Clinic**
 - 知的医学教育システム
 - 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は、医学生が最新情報を探索し、さまざまな仮説の提示と検討を行ない、可能性のある解答、診断、治療の選択肢を裏付ける重要なエビデンスを見出すプロセスを学ぶ。
- **MDA Anderson**
 - 治験に適切な患者を診療情報から選別(clinical trial)

Big DataとLearning system2

- **Learning systemの不可欠性:** ASCO(米国臨床癌学会)
- **The ASCO CancerLinQ initiative**
 - focused on building a “learning health system” composed of a knowledge-generating computer network
 - collect and analyze cancer care data from millions of patient visits and expert guidelines
 - 学習システムを構築し、大量の診療データを集め、診療の現場に治療知識を供給する。17万人のがん症例データベースを2015年までに構築。
 - 各がんについて1～2万人の症例を集める
 - 統計学習、ニューロコンピュータを駆使して学習

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・全世界規模で症例の分子(ゲノム) 情報とそれに対応できる臨床症例の収集。疾患ゲノムコホート、臨床治験DBなども
- **個別化予防の情報基盤**
 - **Population型BioBank**：前向きコホートで健常人の分子情報（ゲノム）と環境情報を集めて追跡するゲノム・コホート

- **最近の動向**

- **UK biobank** :
 - 50万人の健常者の健診・血液を集め、その健康医療状況を追跡する
- **BBMRI** (Biobank/Biomole. Res. Infra.) BioVUの構成
 - 250以上の欧州BioBankを統合

- **わが国のBiobank計画**

- 東北メディカルメガバンク、
Biobank Japan, 6NC 疾患コホート等

バイオ・医療におけるBig Data

- 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 (DataScience)
 - Data-mining, Knowledge Discovery, 人工知能

バイオ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

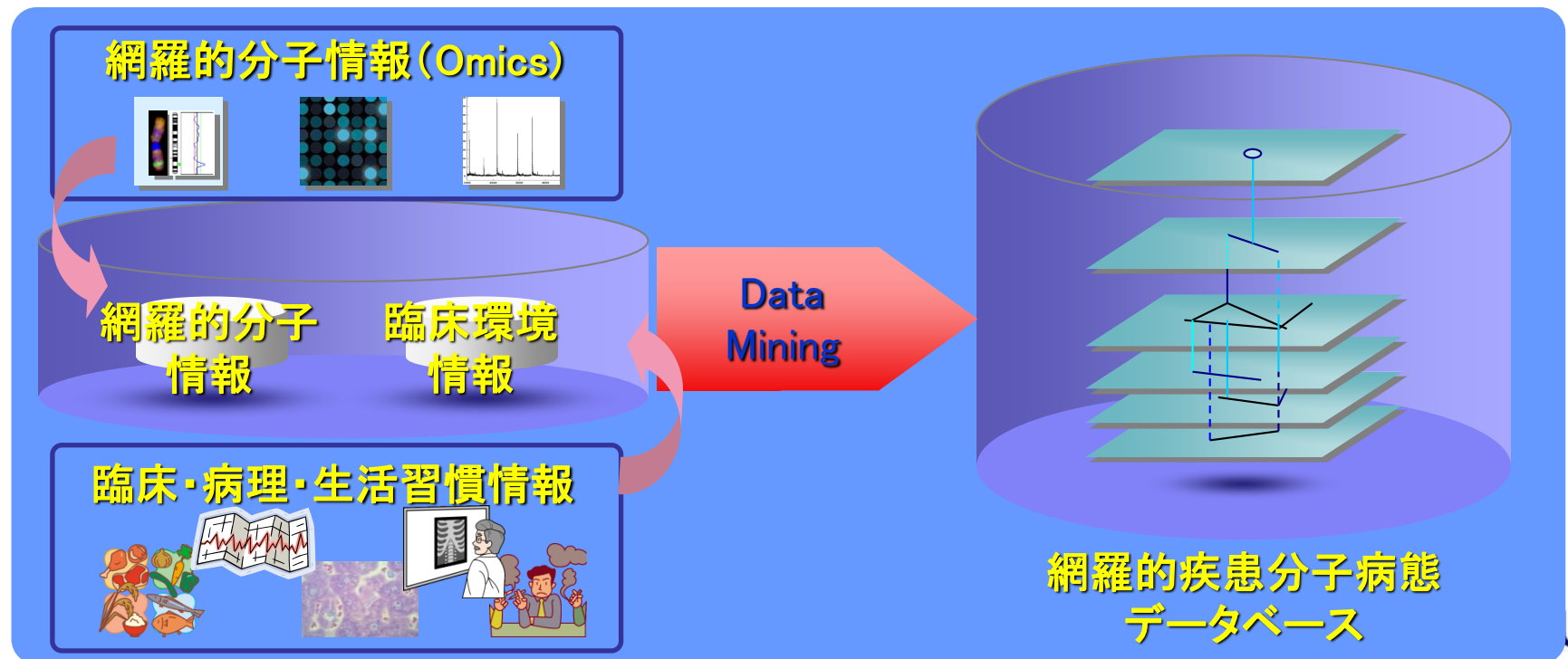
我々の研究室での研究
臨床分子融合データベースと
未来のゲノム・オミックス医療
システム分子医学

統合臨床オミックスデータベース(iCOD)

・「網羅的分子情報に基づく統合的個別化医学(オミックス医学)」の確立のためには網羅的分子情報と疾患病態の関連を明確に体系づけた統一的な病態像を提示し、散在するOmics情報を医学観点から統合することが必要である。

・網羅的分子病態データベース構築方法論の確立とその実証的構築を行う

Shimokawa, K., Mogushi, K., Shoji, S., Hiraishi, A., Mizushima, H., Tanaka, H†. (2010) iCOD: an integrated clinical omics database based on the systems-pathology view of disease. BMC Genomics. 11: S19



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

疾病
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

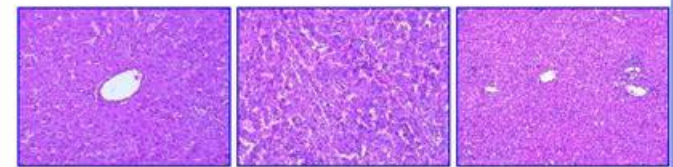
1: 01001032 Hepatocellular carcinoma (肝細胞癌)
治療: 肝S5重区域切除+S8部分切除
病理: T=T1, N=N0, M=M0

2: 01010162 Hepatocellular carcinoma (肝細胞癌)
治療: 肝S8重区域切除
病理: T=T3, N=N0, M=M0

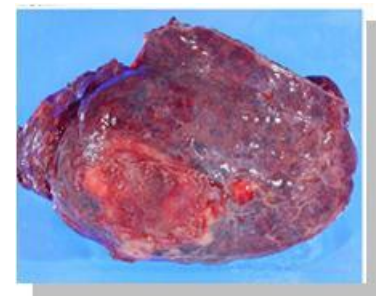
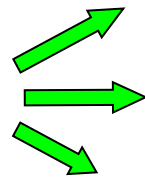
3: 01017640 Hepatocellular carcinoma (肝細胞癌)
治療: 肝右葉切除 (右門脈癌栓術後)

4: 01018257 Hepatocellular carcinoma (肝細胞癌)

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:



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
疾患
生活歴
検査所見
血液
生体
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

01018257

CT

検査部位: 肝・胆・脾・肺

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

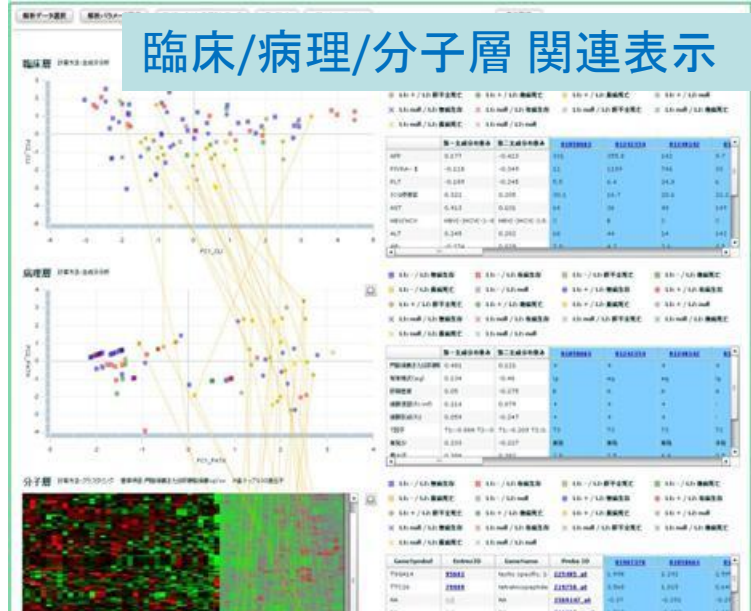
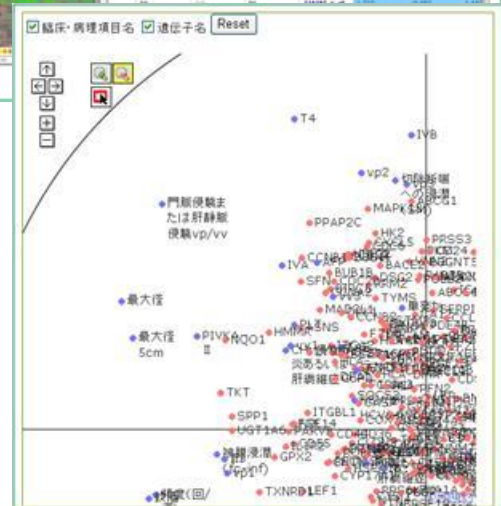
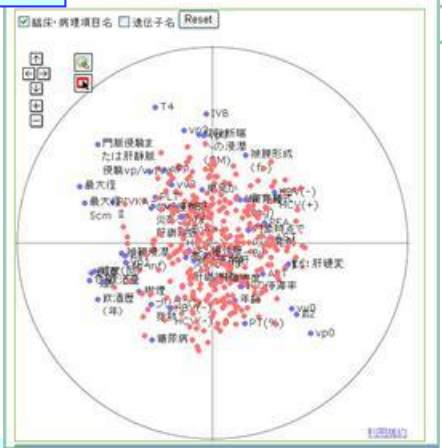
臨床オミックス解析

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

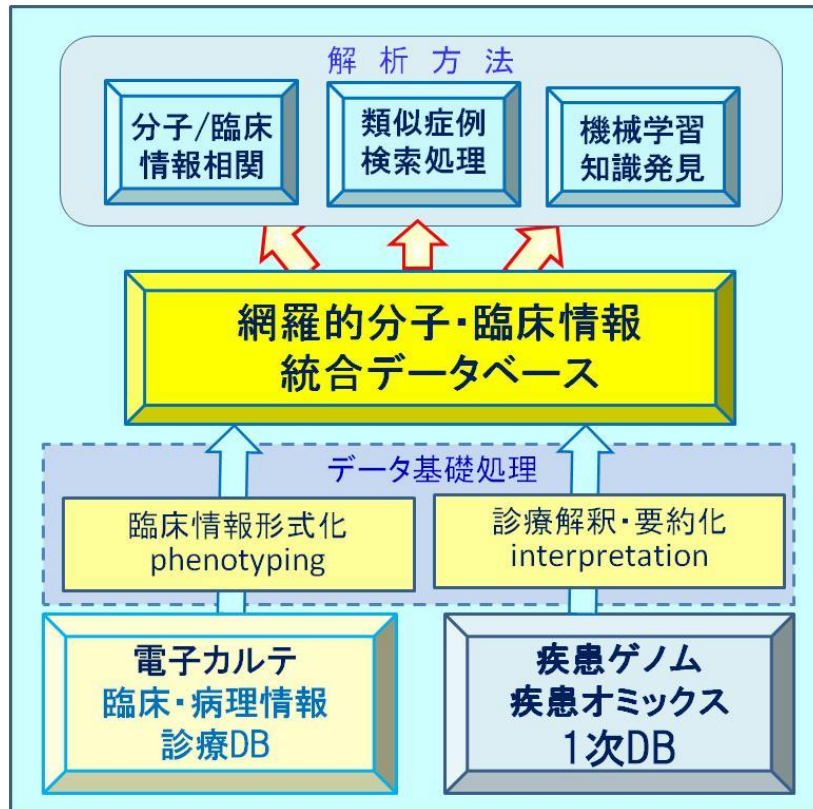
Pathome-
Genome
Map

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解明

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)



電子カルテから入力された臨床・病的診療情報と疾患ゲノム・オミックスから入力された分子情報は

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

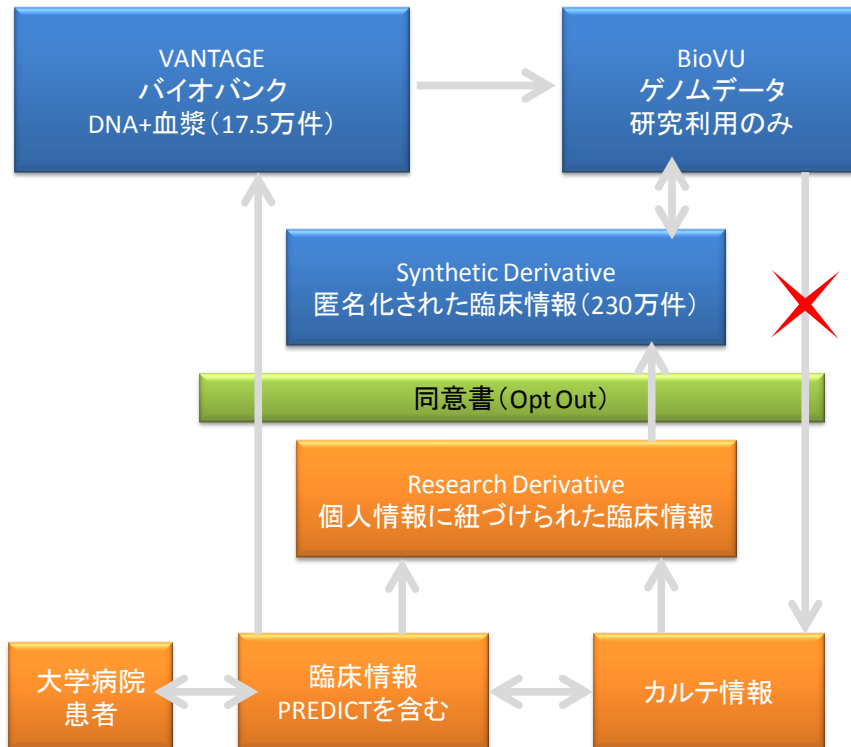
どのようなデータ形式か検討の必要RDF化やi2b2方式など

3. データ解析部

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

Vanderbiltのシステム

- 電子カルテ情報から匿名化を経たSDデータベースと臨床検査から採取した血液サンプルよりなるBioVU
- EHR-driven genomic research (EDGR)のシステム構成



電子カルテ (StarPanel)

REDCap : SDの検索・DLおよび簡便な phenotype検索

Synthetic Derivative : 200万人以上の匿名化臨床情報や投薬情報

- 審査後に利用可能な部分
- 電子カルテ情報を計算機処理で匿名化した臨床情報のDB, Case-Control study
- 研究への参加はopt-out方式

Research Derivatives : (匿名化以前) 個別の患者情報データベース

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 (前出)

バイオバンクと遺伝子解析

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

VANTAGE Core : バイオバンク運営、DNA抽出・ゲノム解析等

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

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**

網羅的分子情報 (omics) の利用

- 薬剤特異的遺伝子発現 (Drug-induced SDE)

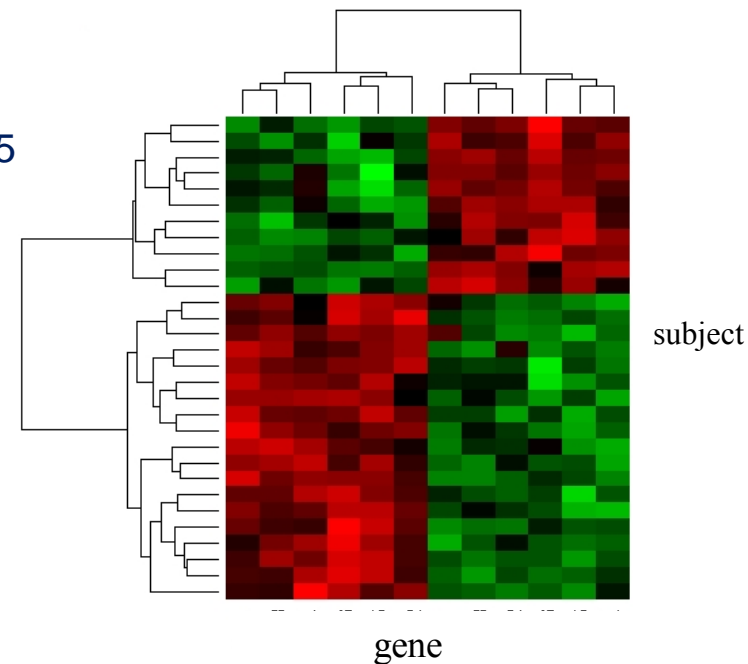
- CMAP : Connectivity Map

- 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) ベース
- 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),

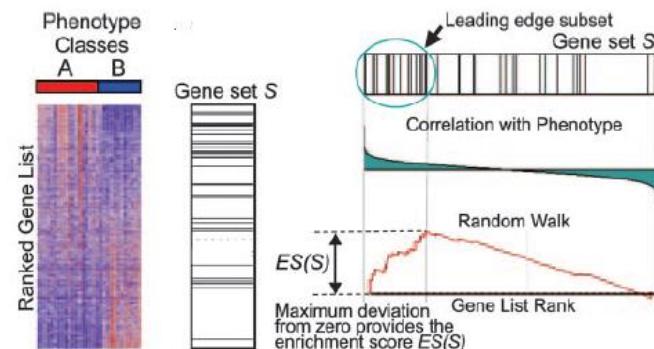
- 疾病罹患時の遺伝子発現プロファイルの変化
- 米国NCBI作成・運用 2万5千実験, 70万プロファイル
- ArrayExpressもEBIが作成、サンプル数同程度



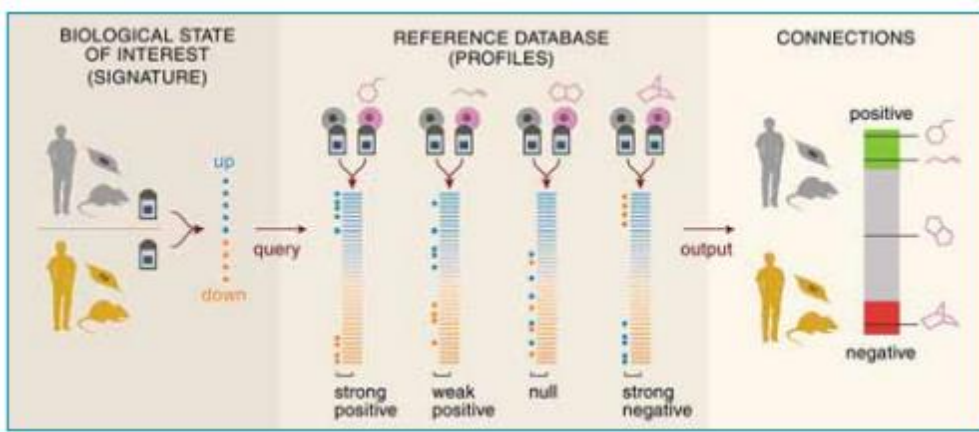
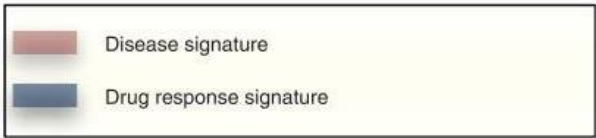
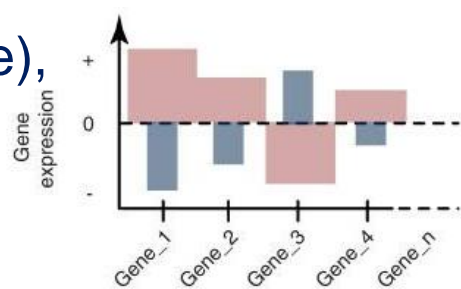
遺伝子発現プロファイルによるDR(1)

- 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
- 両者が十分負に相関する
- 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

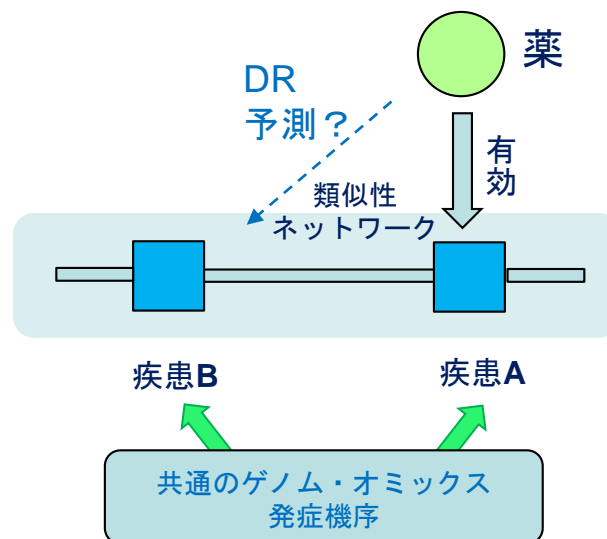


GSEA



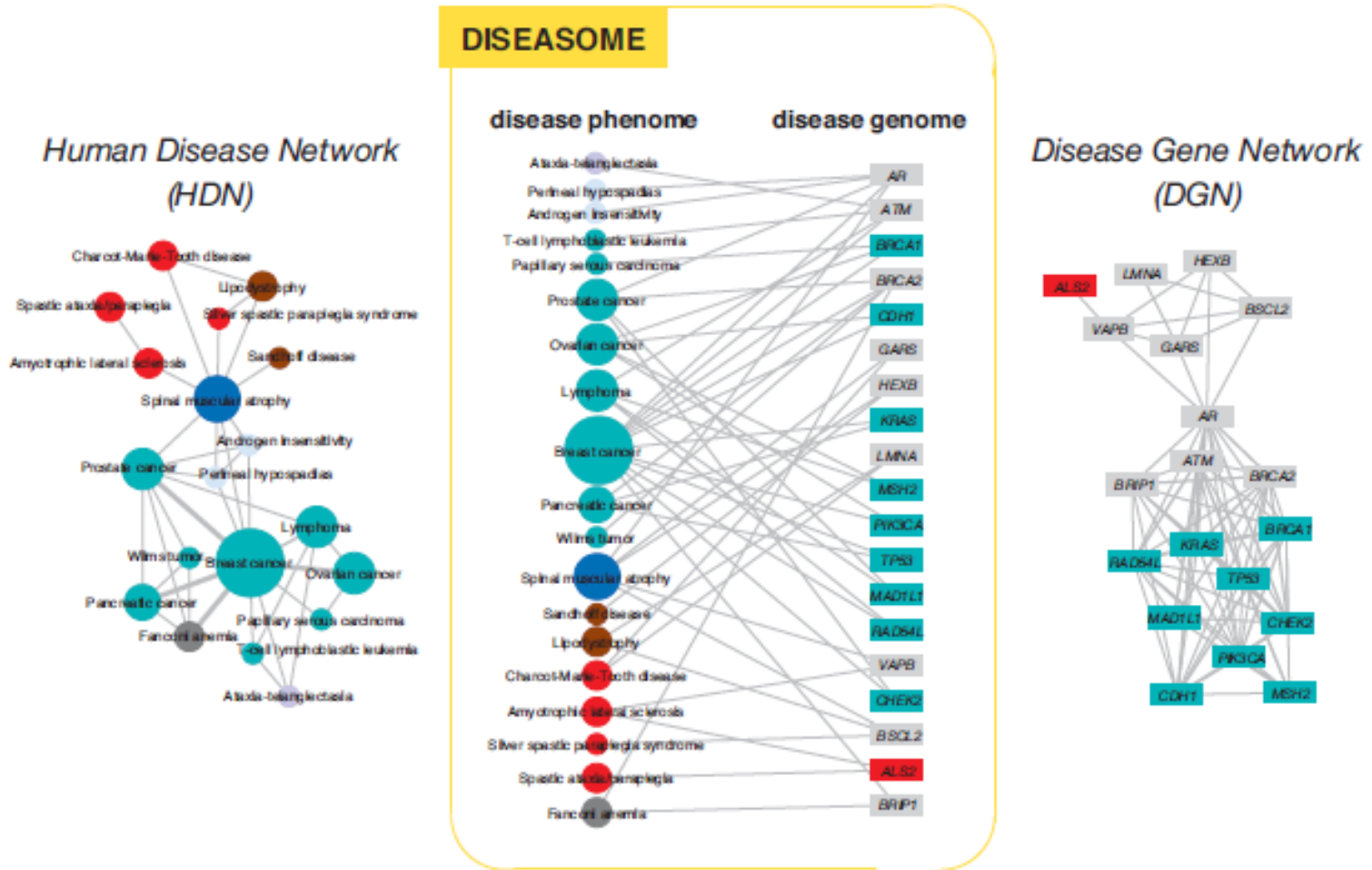
疾患ネットワークに基づいたDR

- 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

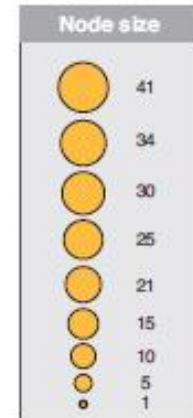
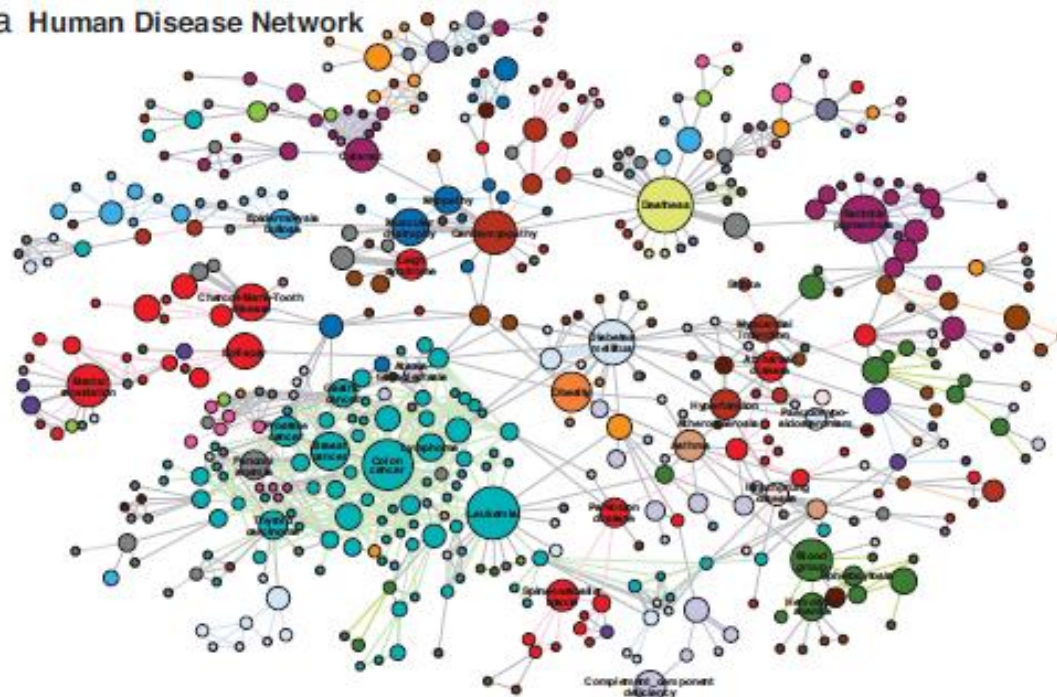


疾患 ネットワーク (HDN)

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

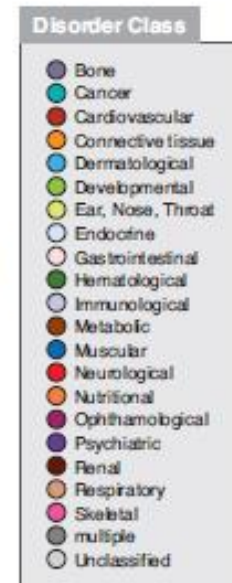
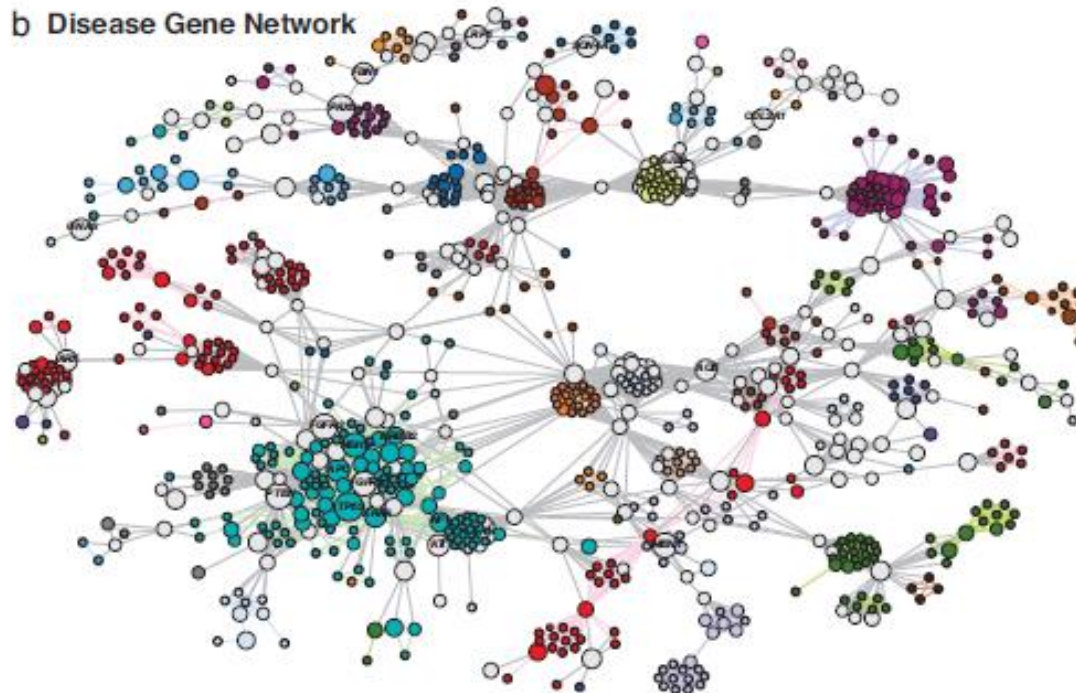


疾患遺伝子 ネットワーク (DGN)

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

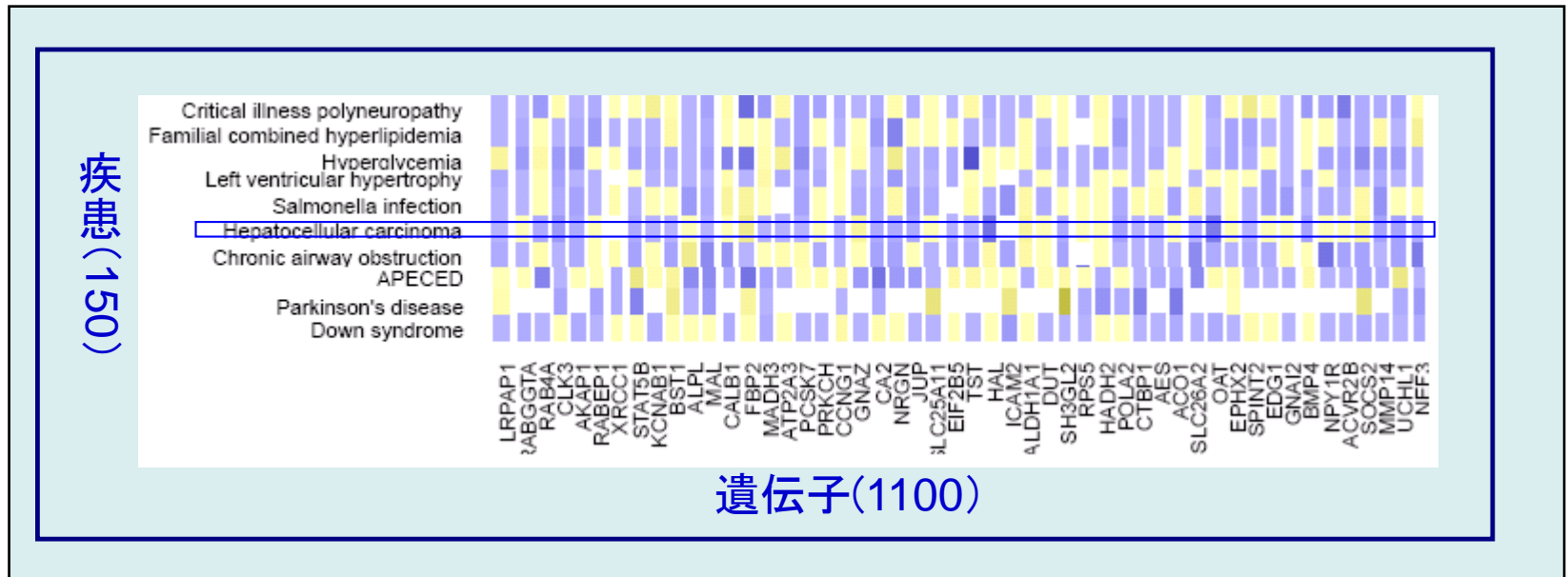
b Disease Gene Network



第2世代型

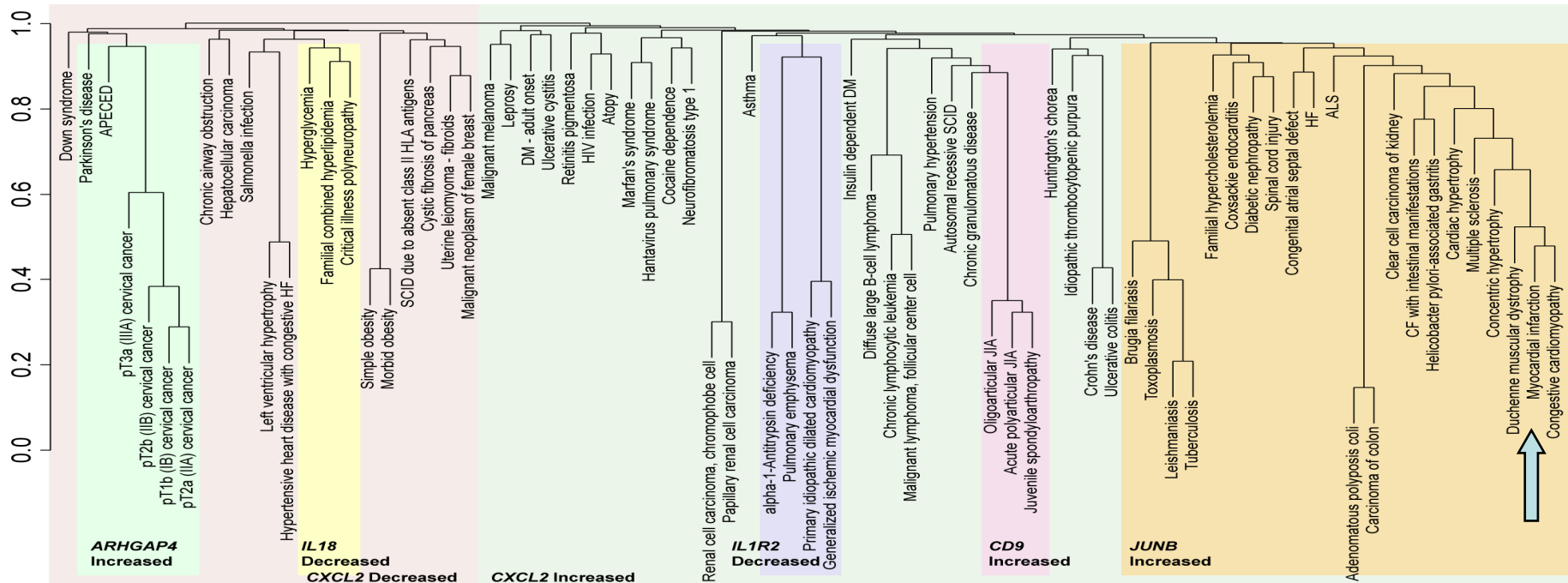
GENOMED (A.Butte et al)

- 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)

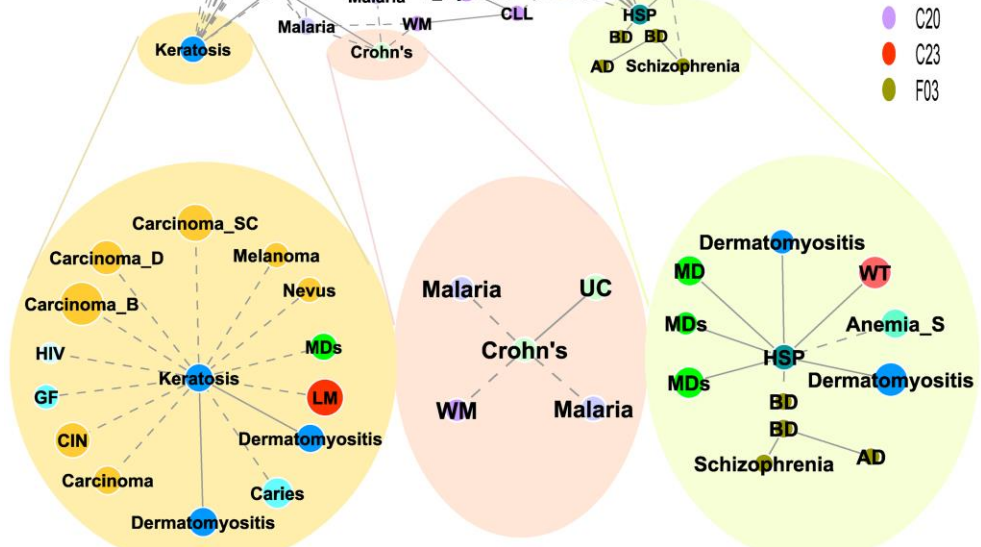
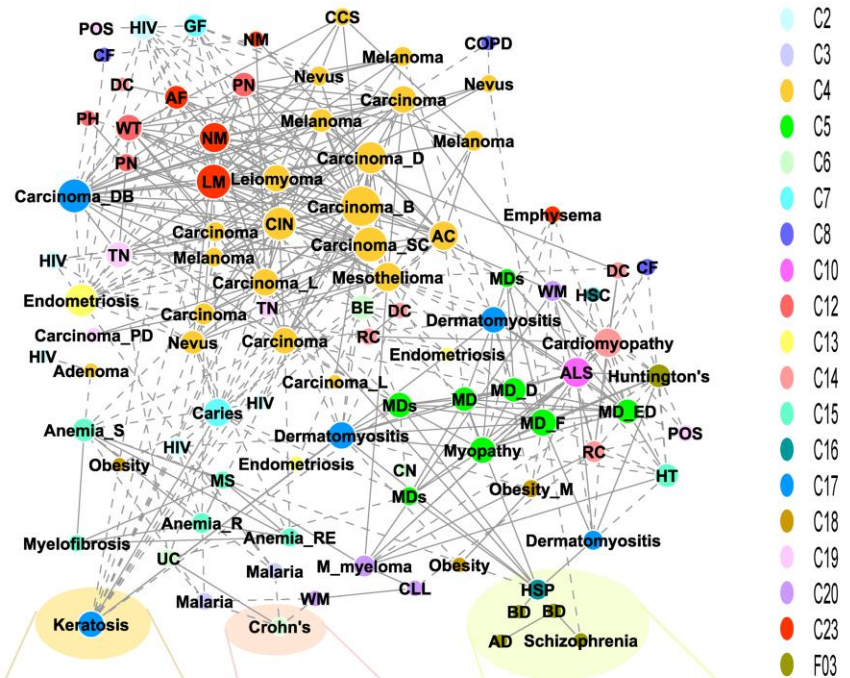
Crohn's disease

⇒ 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

ゲノム・オミックス医療の将来 とシステム分子医学

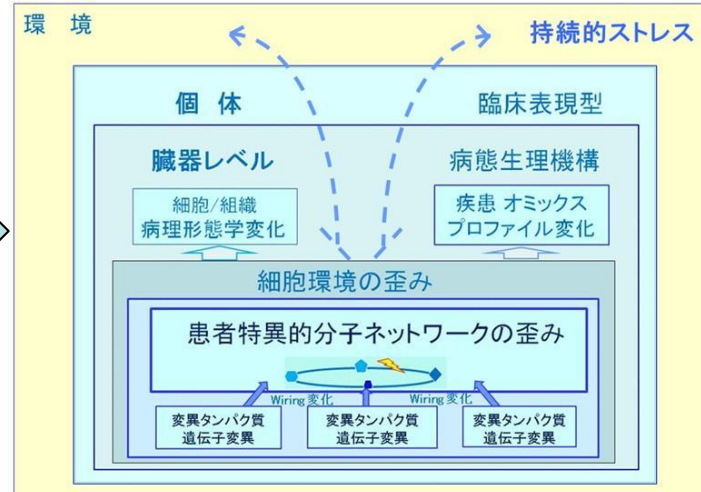
網羅的分子医学 第3パラダイム システム分子医学 (2010~)

システムバイオロジーの疾患への応用
「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



システム分子医学の基本戦略

マルチオミックスからの分子ネットワーク病態認識へ

1. 疾患オミックスプロファイル (disease omics profile) 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

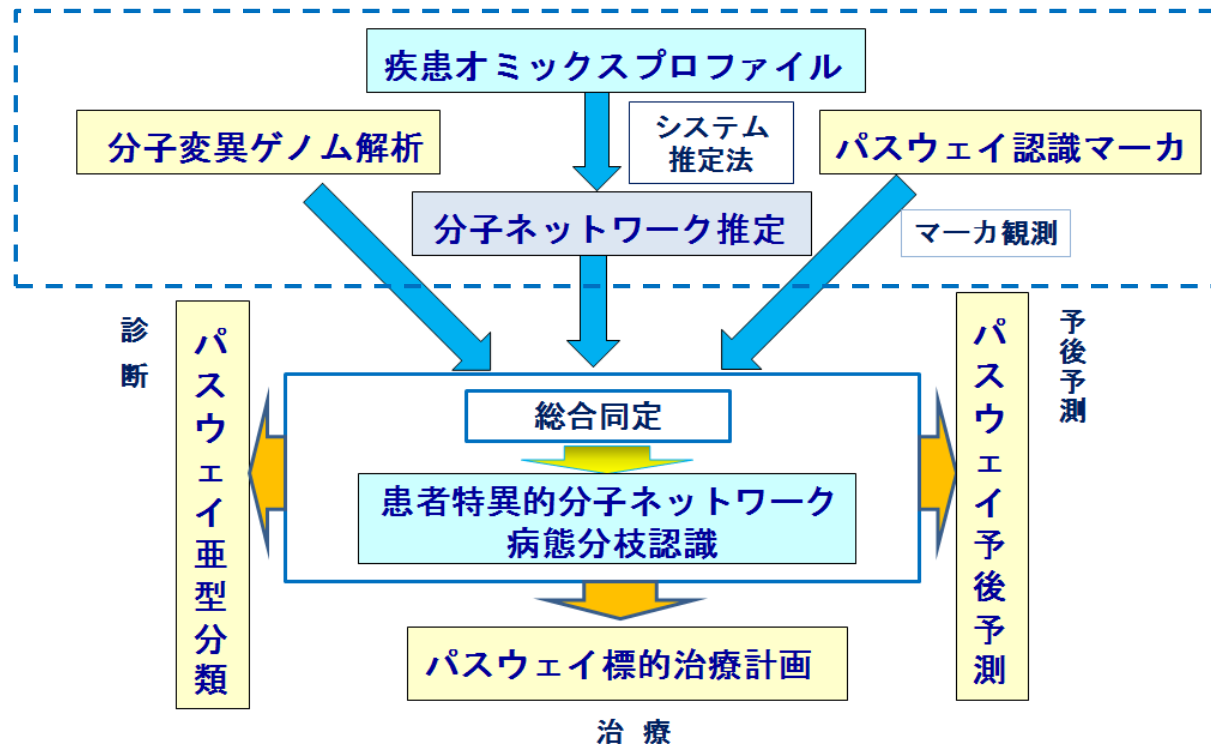
Dysregulated pathway/subnetwork (“disease-perturbed network” Hood) の同定

2. マルチ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分子の変異

プロテオーム (パスウェイバイオマーカ) リン酸化情報



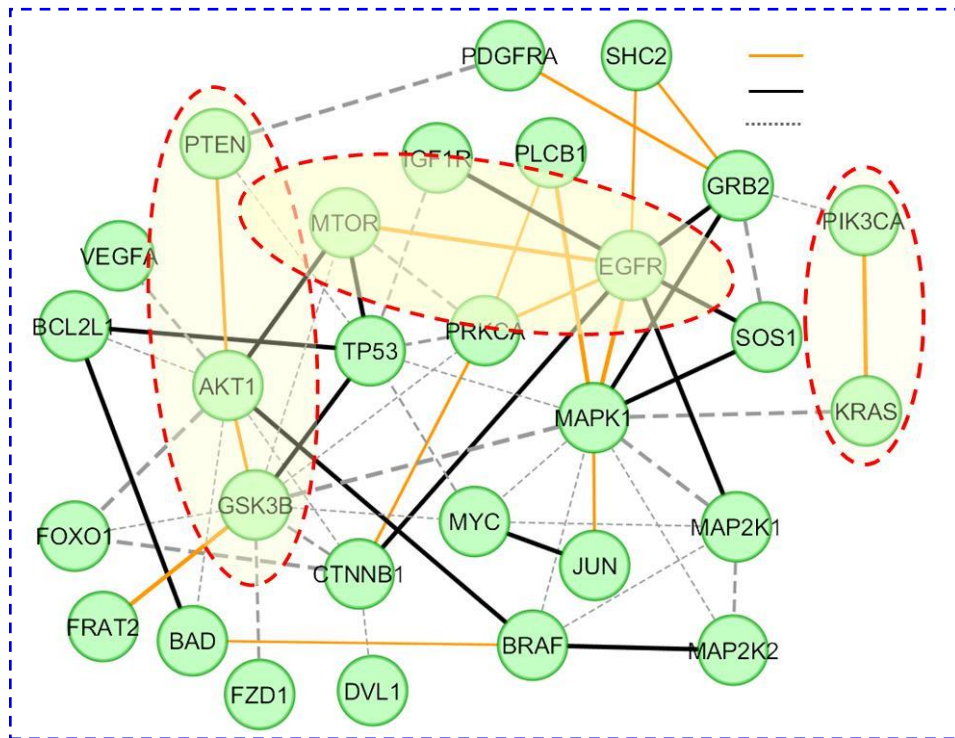
分子ネットワーク歪構造同定

疾患ネットワーク同定方法論

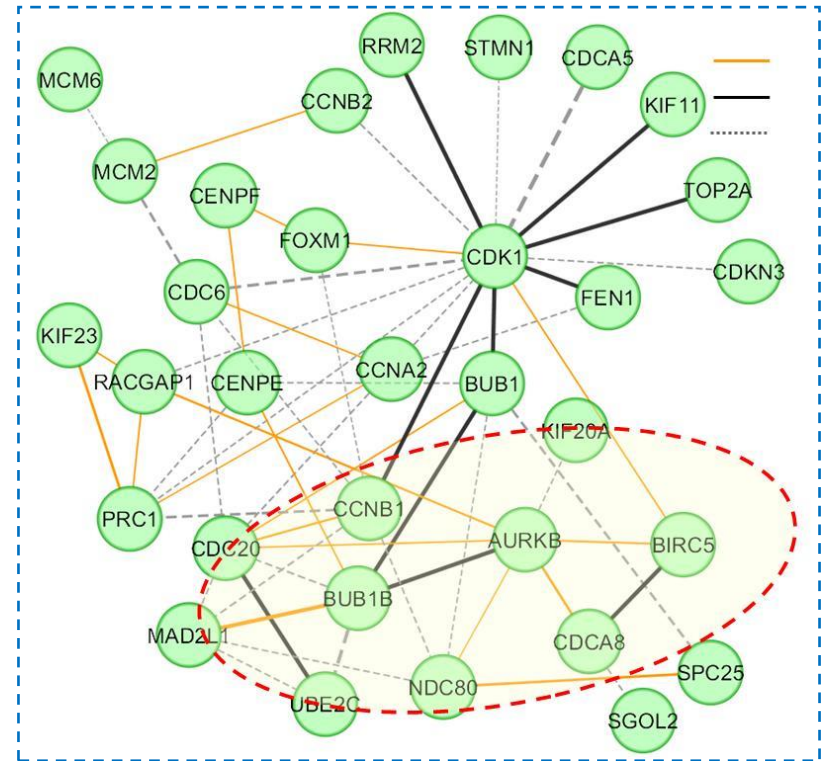
東京医科歯科大学病院 肝胆膵外科にて肝胆膵外科で手術を受けた40症例

— 予後良好群20例 vs. 予後不良群20例

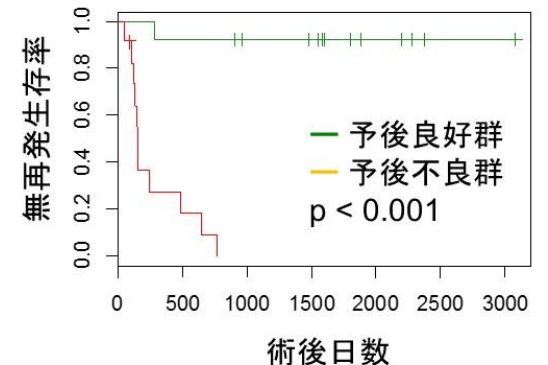
細胞増殖分子ネットワーク推定
differential branches 増殖系の過剰活性



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

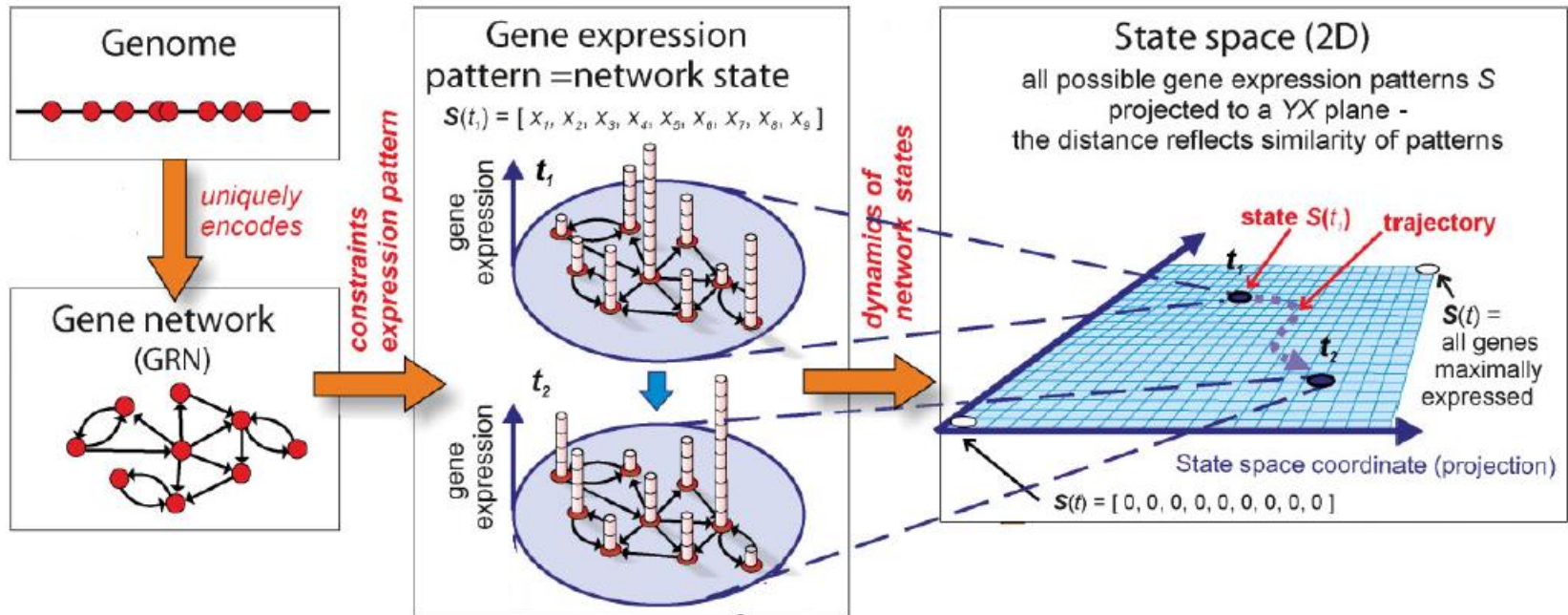
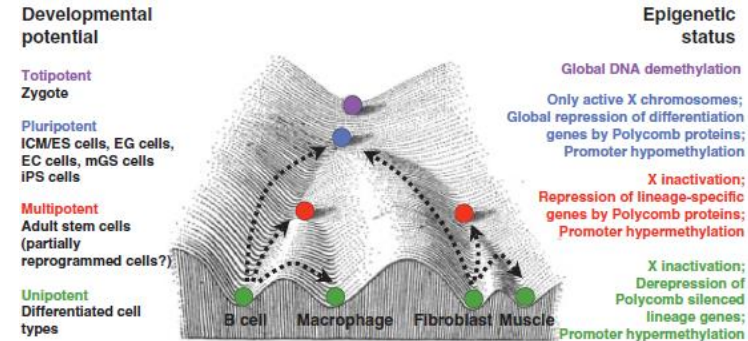


— 予後不良群に特徴的
— 両者共通
..... 既知PPIあり

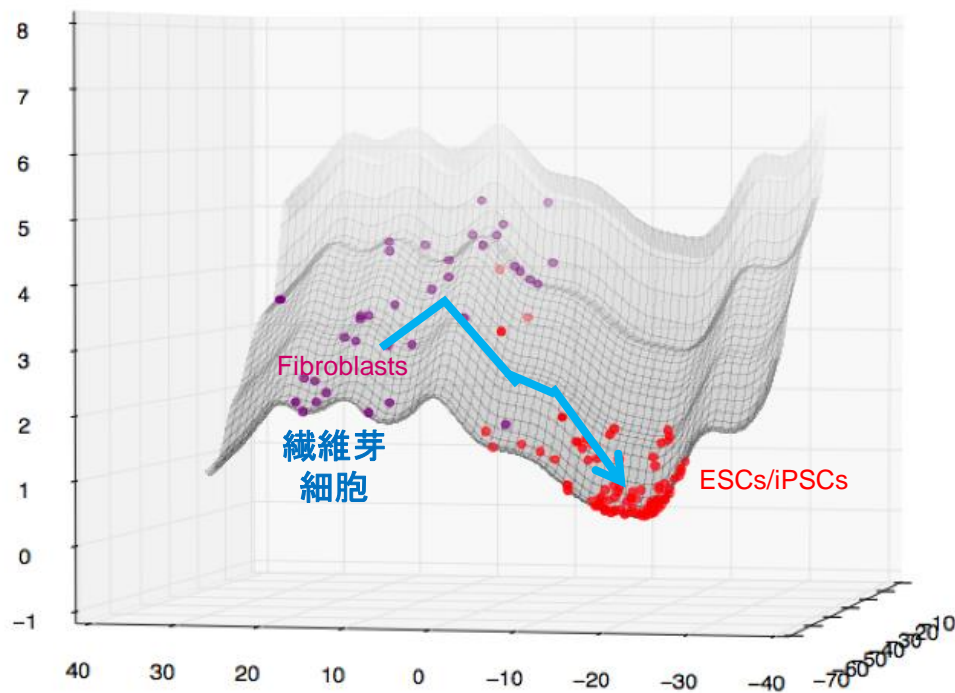
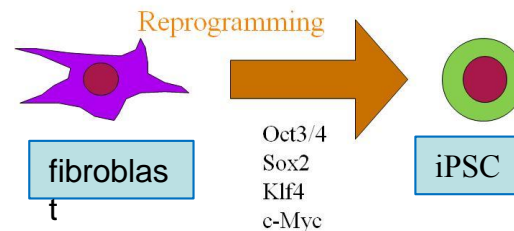
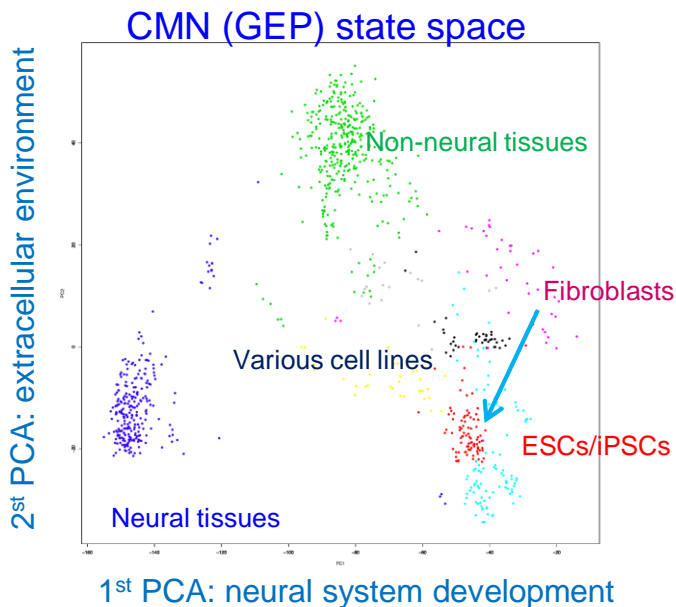


細胞分子ネットワークの状態空間

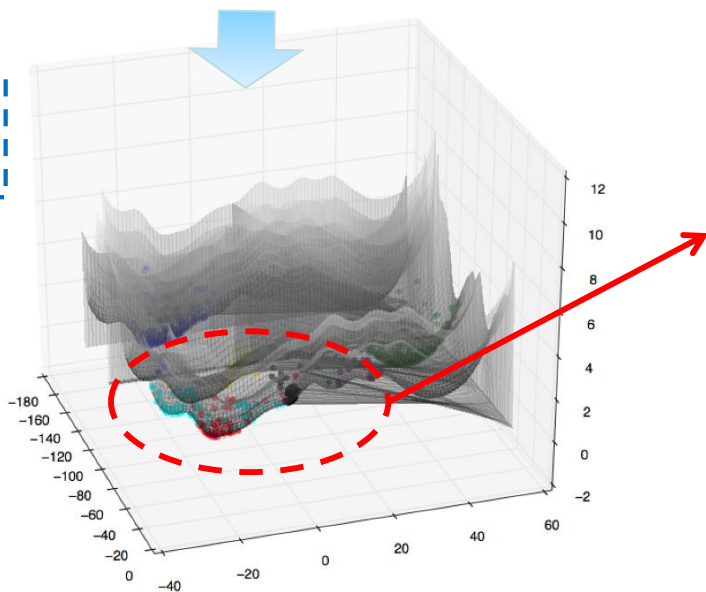
- 細胞分子ネットワーク (CMN) の状態空間を考える
- 細胞分子ネットワークの結線構造はゲノムにコードされて生得的/固定的である
- CMNの中でどの遺伝子が発現しているかは、細胞型および状態によって異なる



iPSC reprogramming プロセス



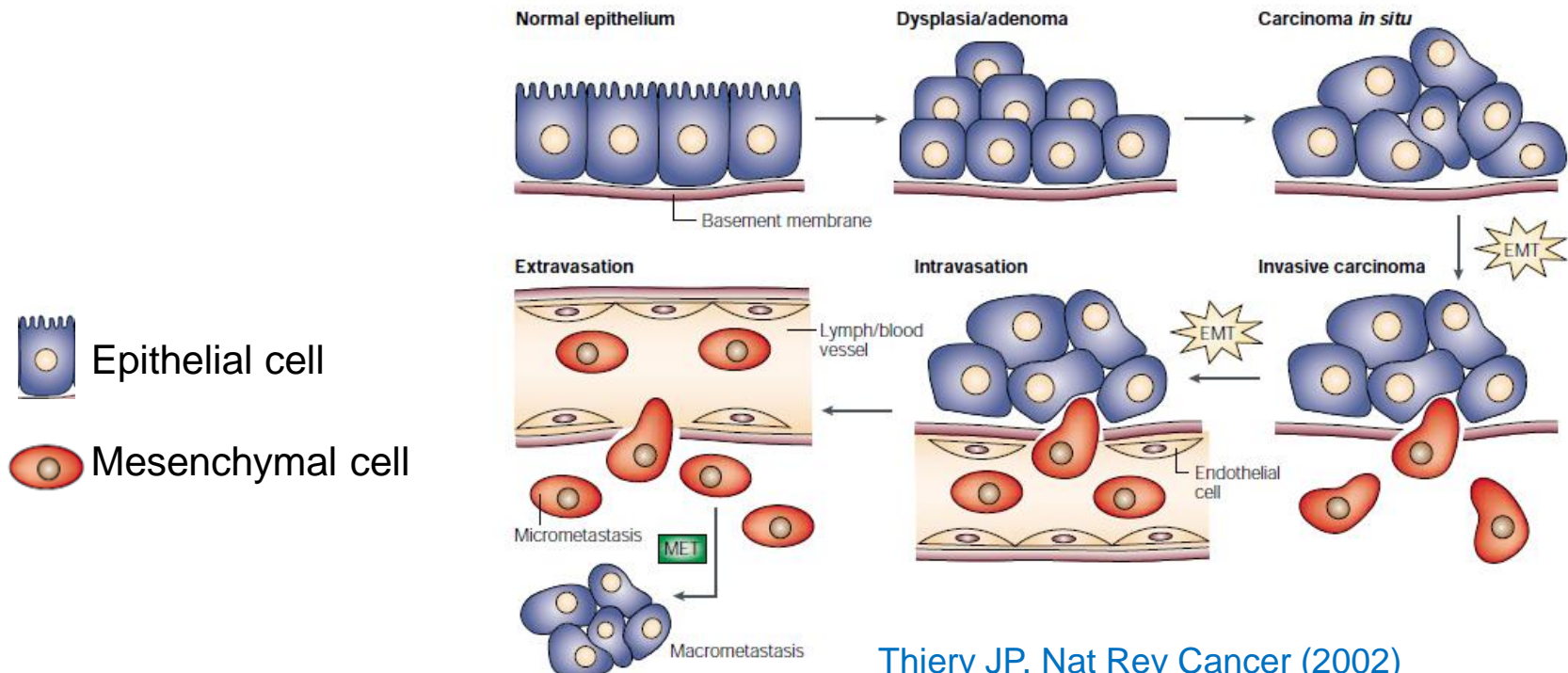
qWEL picture



Oxford press-JBiC prize 2012

上皮間葉転換(EMT)

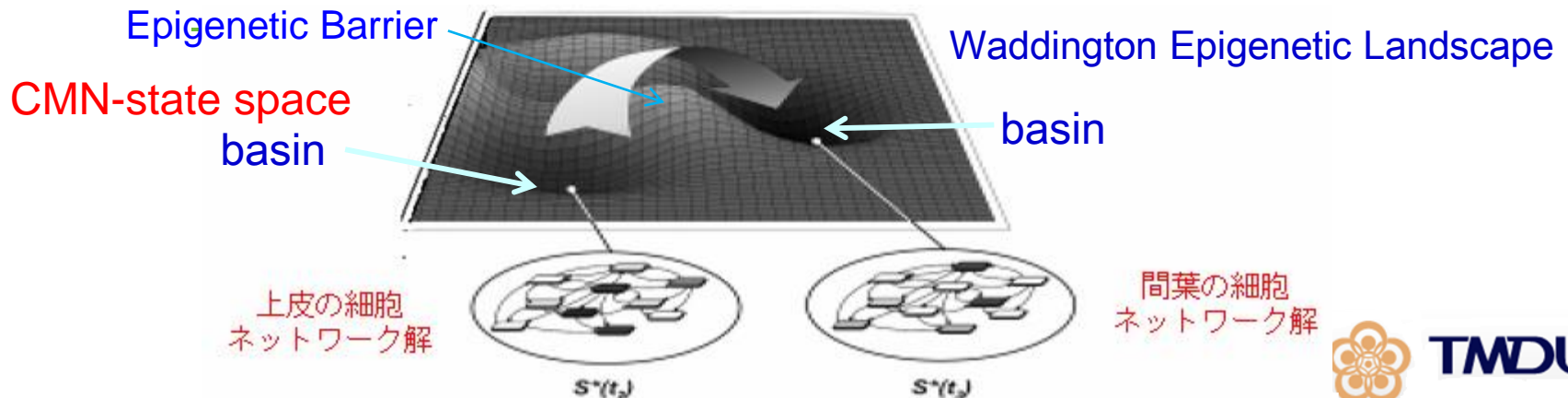
- Epithelial-mesenchymal transition (EMT)
 - non-motile, polarized epithelial cells, embedded via cell-cell junctions, dissolve their cell-cell junctions and convert into individual, non-polarized, motile and invasive mesenchymal cells
- EMT は生物に取っても重要な発生現象
 - Gastrulation (原腸陥入)、神経堤形成など
- EMT は、細胞分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



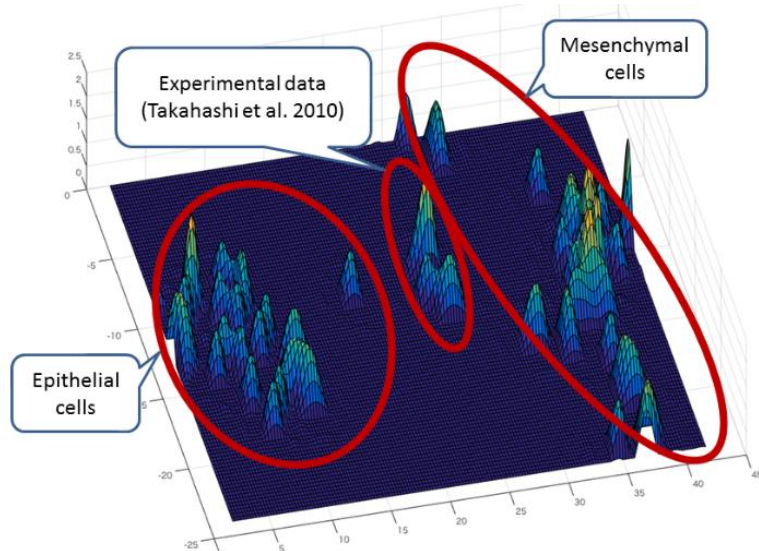
がんの浸潤転移

細胞分子ネットワーク空間の大域変異

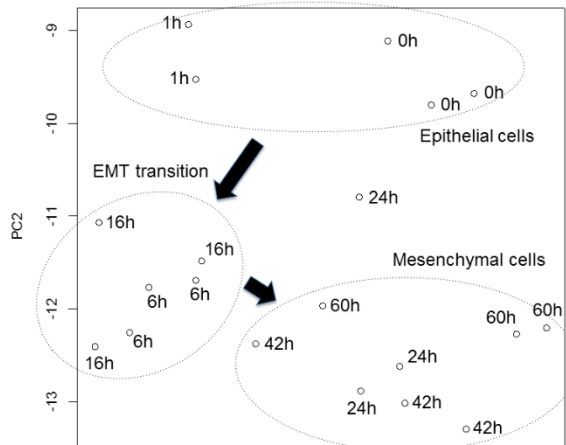
- 細胞分子ネットワーク状態 (CMN状態空間)
 - 細胞分子ネットワークはいくつかの安定解を持つ
 - 各安定解は細胞種類に相当する(約220種類の安定解)、細胞ネットワークが作るCMN空間のエピゲノム地形(準ポテンシャル)「アトラクター状態」これらは多能性幹細胞から分化解である
- EMT (上皮・間葉転移) Epithelial-mesenchymal Transition
 - 細胞分子ネットワークの上皮安定解が間葉安定解へと移行
 - 全ネットワーク状態の転移であるから「相転移」といえる
- がんの転移とは相転移
 - 安定解を不安定化し(ゲノム不安定性)、障壁を越えるためにはゲノ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

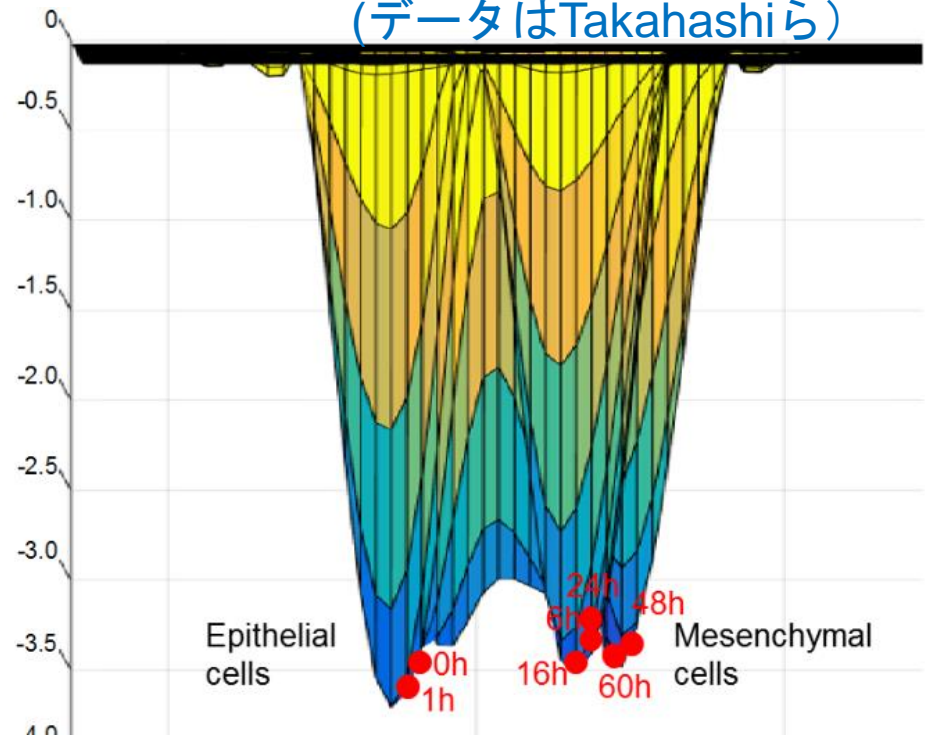


細胞ネットワーク空間の確率分布



がんの
EMT過程

網膜色素上皮細胞のEMTのqWEL
(データはTakahashiら)

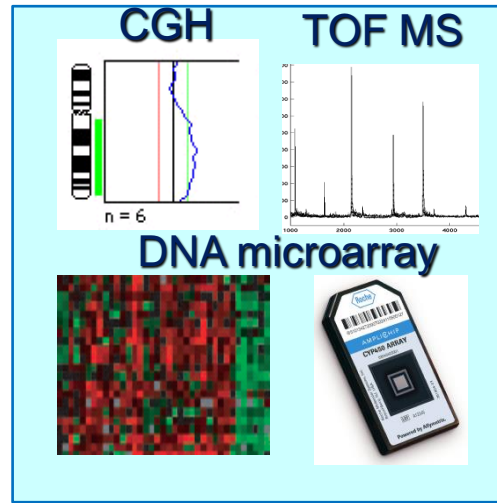


Precision Medicine

断層像
再構成技術

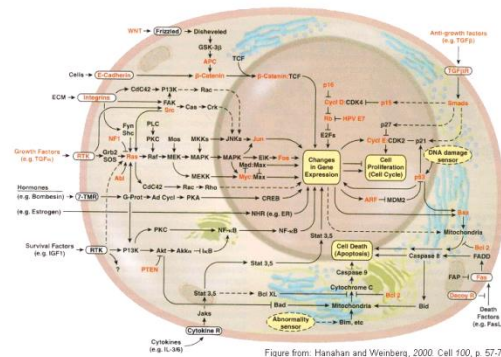
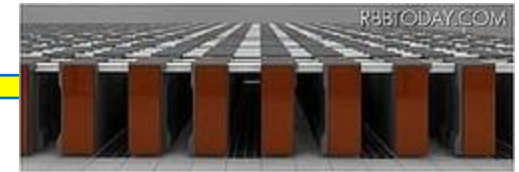


逆問題



疾患マルチ
オミックス
プロフィール

スーパー
コンピュータ



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



Life-long healthcareとPHR

生涯的健康医療に関する動向（1）

BioBankの国内外の普及

- 健常者の生涯にわたる疾患リスクを追跡調査
- 疾患発症の＜環境（習慣）要因X遺伝的素因＞を解明
- UKバイオバンク 50万人
 - 45歳から69歳の50万人の健常者の被験者について血液標本を2006年から2010年の間に集め、その健康医療状況を追跡する複合型のバイオバンク・プロジェクト
- 東北メディカルメガバンク（Tommo）15万人
 - 地域住民コホート（8万人）：宮城・岩手県の住民の健康医療記録を収集。コホートを10年間追跡。疾患の発症リスクなどを評価。東日本大震災後のPTSDやうつ病など
 - 3世代コホート（7万人）。2万人の新生児について両親、父親・母親の祖父母の計7名に全ゲノム配列情報を含んだ、遺伝・環境情報収集
- そのほか、長浜コホートなど各地でBiobank

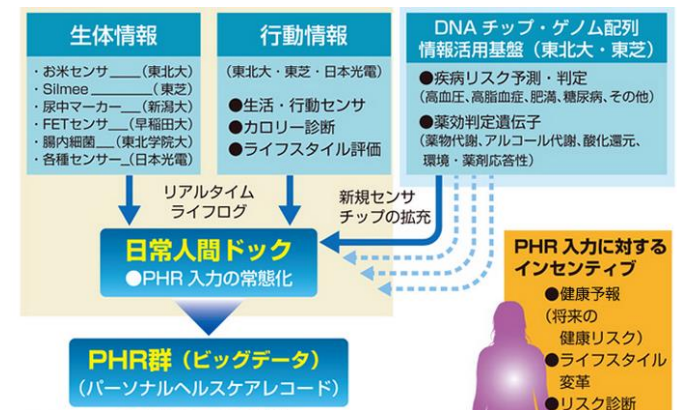
生涯的健康医療に関する動向（２）

生体センサリングの普及

- 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
 - コンティニューアなど幾つかの日常モニタリング, mHealthのプロジェクト



ECG; EEG; Skin Conductivity; EVG



東北大学COI 「さりげないセンシング」

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	8.8

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

次世代医療情報システムの基礎概念（1）

Life-long (life-course) health care

- 健康寿命延伸の情報基盤としてのPHR/EHR
 - 「生涯継続ヘルスケア」 (life-long, life course hc)
 - 従来のPHR：アレルギー副作用などのリスク情報
 - 近年のPHR：生涯的時間継続性における健康医療データ
- 全国に広がるBiobank計画における生涯疾患リスクと遺伝・環境要因の情報の蓄積
- Biobankを情報基盤とした地域医療情報連携
 - 次世代地域医療連携のあり方

次世代医療情報システムの基礎概念 (2)

Life-long (life-course) health care

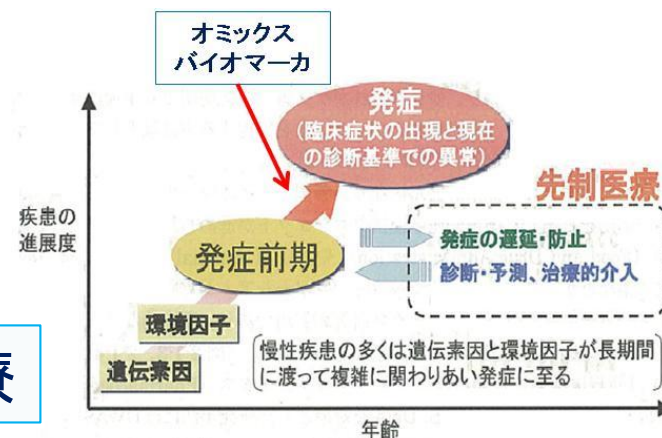
- 日常的生体モニタリングの目標
 - Life-long healthcareにおける疾患発症の予知へ
- 発症予知と分子的バイオマーカー
 - 先制医療 (preemptive medicine) へ
 - Liquid biopsy (液性バイオマーカー) 血中循環RNA, DNA計測, がん・アルツハイマー症などの予知
 - DIY genomicsの発展



Nanopour型シーケンサ

先制医療

Preemptive medicine (Zerhouni 2005)
By making use of precise molecular knowledge to detect disease before symptoms are manifest, and intervening before disease can strike.



reactive治療医学 → proactive予測医療

ゲノム・オミックス医療と ビッグデータ

ゲノム・オミックス医療臨床実装の普及
Clinical Implementation of Genome/Omics Medicine

詳細で大量のデータの蓄積

データマイニング・人工知能
データサイエンスの進展

データ駆動型医学の進展
BigData-driven Clinical Science

臨床科学は基本的にEmpirical

Biomed Pharma Data Scientist の必要性

- NIH BD2Kでも今年度の柱
- 新しい創薬戦略への適応
 - Chemicalに詳しい人材だけではなく、オミックス創薬、疾患ネットワーク準拠の創薬・DRなど生命・医療に準拠する創薬戦略にも
 - 国家的プロジェクト、precompetitive consortiumへの対応
- 新しい医療・健康戦略への対応
 - 精密医療 precision medicine
 - 参加型医療 participatory medicine (healthcare)
- 健康医療におけるデータ・情報の重要性
急激に拡大している

ご清聴ありがとうございました

