

# ゲノム・オミックス医療における ビッグデータの利用

東京医科歯科大学 難治疾患研究所

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

## 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・Precision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

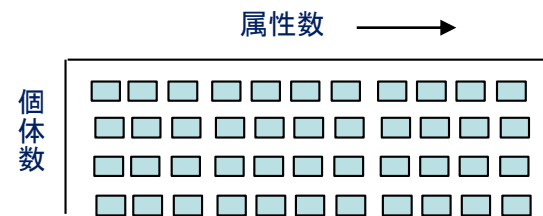
- Population医療情報・医療政策

# 新しいタイプの医療ビッグデータ

- ゲノム医療における網羅的分子情報のビッグデータ (L. Chen)

- **Small “Big Data”**

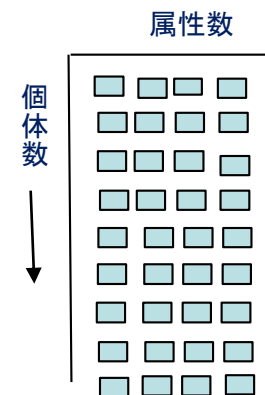
- 1 個体に関するデータ数が膨大
  - 全ゲノム配列x30で100Gbpなど
- 従来の統計学が無効  $p \gg n$



- 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
- 属性数十項目 サンプル数が多数



新しいデータ科学の必要性

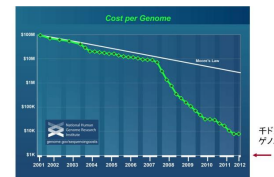
# 新しい種類のビッグデータ (網羅的分子情報)

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

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



DNA Sequencing Cost: the National Human Genome Research Institute

現状 米国ではすでに20以上の医療施設で、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。

NHGRIの  
リスト

Organization	Year
University of Washington	2009
Mayo Clinic	2010
Washington University	2010
Stanford University	2010
Massachusetts General Hospital	2010
University of California, San Diego	2010
University of Michigan	2010
University of Texas at Austin	2010
University of California, Los Angeles	2010
University of Colorado	2010
University of Pennsylvania	2010
University of Texas at Dallas	2010
University of California, Berkeley	2010
University of Washington	2010
University of Michigan	2010
University of California, San Diego	2010
University of Texas at Austin	2010
University of Colorado	2010
University of Pennsylvania	2010
University of Texas at Dallas	2010
University of California, Berkeley	2010

WGS/WESによる原因遺伝子変異・がんドライバー遺伝子  
同定



ウィスコンシン  
医科大学病院

Genome sequencing  
program患者選択委員会

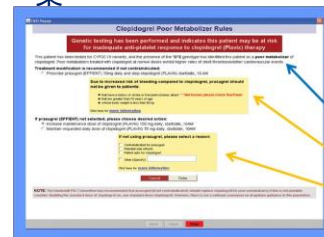


ベーラー  
医科大学

Whole genome laboratory  
を設置。In-houseでシーケ  
ンス

その他にMayoClinic, Washington, Partnerヘルスケア  
など

薬剤代謝酵素の多型性の判別による個別化投  
薬



クロビドゲル処方  
電子カルテの警告画面  
処方名プラスティック・投薬情報  
ステント留置手術の後に処方

CYP2D6の多型性で2\*2の場合は  
代謝機能が低いので血中が蓄積する  
薬剤投与の注意は十分である  
この患者の場合は(2\*2)プラスゲル  
(商品名エフィエント)に替えるか  
分量を2倍にすると警告している



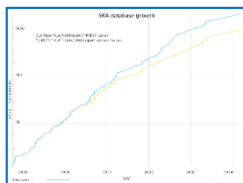
# 医療ビッグデータの流れ

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

2000兆塩基 (2 Pb)  
が登録 (SRA)



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

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

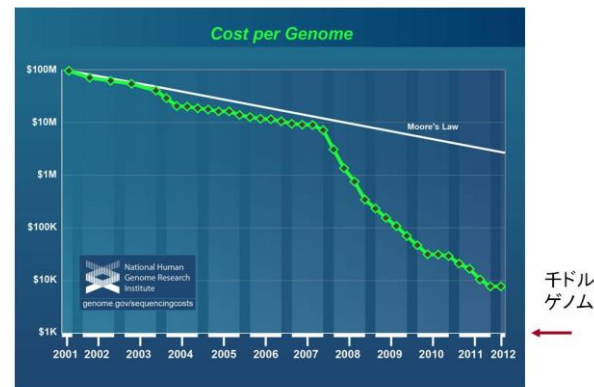
# 次世代シーケンサのインパクト

## 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 (経費1/5)



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!

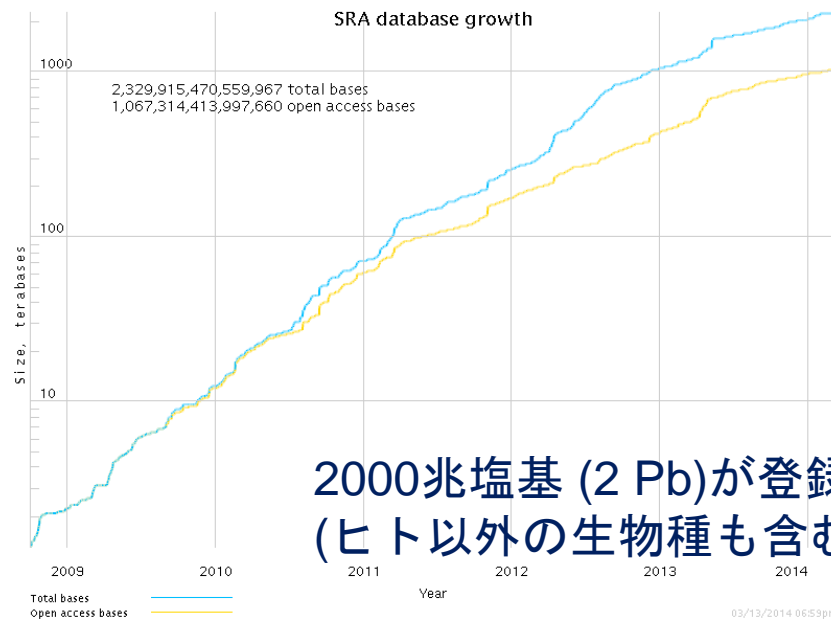
# クリニカル・シーケンシング

- ヒトゲノム配列を直接読み、個人に特徴的な変異情報から治療戦略を検討する
- ヒトゲノム情報の解析方法
  - **全ゲノム解析 (WGS)**
    - すべての染色体領域を解析対象とする
    - 30億塩基 (3 Gb) を最低30xのカバレッジで読むため、1000億塩基 (100 Gb) 程度の情報が生まれる
  - **全エクソーム解析 (WES)**
    - コーディング領域だけを、配列特異的なオリゴマーを用いて濃縮
    - 試薬によって異なるが、6000万塩基 (60 Mb) を平均100x程度のカバレッジで読み、1検体あたり6 Gbとなる
- 例：イルミナHiSeq2500
  - ラピッドラン・モード: 27時間、100 Gbで実現可能な解析
  - 30x ヒト全ゲノム解析 1検体
  - 100x エクソーム解析 15検体
    - 60 Mbライブラリ × 100x × 15検体 = 93 Gb



# 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)

**ムーアの法則(計算機の性能) を遥かに超える速度で蓄積**

# 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

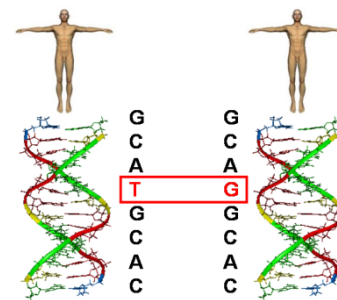
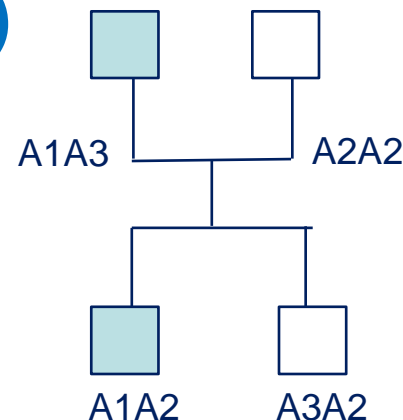
H.Tanaka: Genome and Omics Medicine

—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  
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

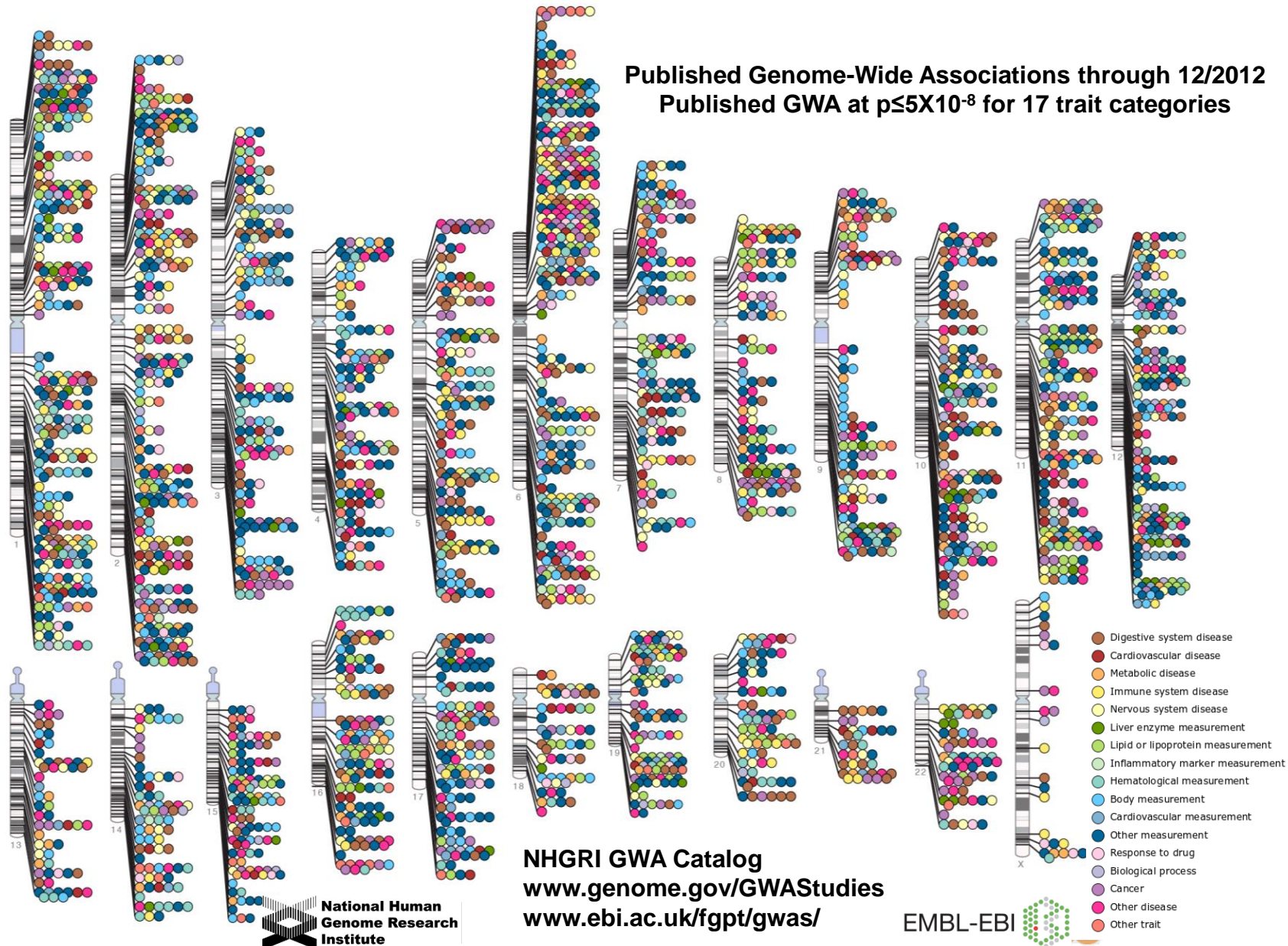
環境因子: 例, 喫煙肺がん相対リスク4.8

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

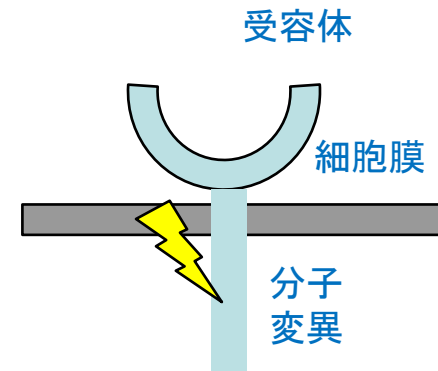
疾患	原因遺伝子・感受性遺伝子	座位
ハンチントン病	HTT(huntingtin) 遺伝子	4p16.3
筋ジストロフィー	DMD(dystrophin)遺伝子	Xp21.2
嚢胞性線維症	CFTR遺伝子	7q31.2
家族性大腸ポリポーシス	APC 遺伝子	5q21-q22
色素性乾皮症	XPA遺伝子(Aグループ)	9q22.3
炎症性大腸炎	CARD1 遺伝子など	16q21
統合失調症	COMT 遺伝子など	22q11.21-23
2型糖尿病	CAPN1 0遺伝子/ADIPOQ遺伝子	2q37.3/3q27
晩発性アルツハイマー	APOE遺伝子など	19q13.2
高血圧	AGT 遺伝子/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

# 疾患関連遺伝子のマップ



# 網羅的分子医学 第1パラダイム テーラメイド投薬

- **薬理ゲノム学**(Pharmacogenomics: PGx)  
薬剤個別化投与(Personalized medication)
  - PD (薬動力学) 薬剤応答多型性
    - 例 チロシンキナーゼ阻害のPD薬剤多型性  
イレッサ(gefitinib) exon19欠失, 21:L858R変異
  - PK (薬物動態学) 代謝酵素多型
    - ADMET,薬剤代謝酵素の個別性,P450(CYP)
    - 化学薬剤の登場から150年進化で淘汰されず



代謝酵素多型	個別化応答を示す主な薬剤
チトクロームP450(CYP)	
CYP2D6	$\beta$ 遮断剤
CYP2A6	5-FU(抗がん剤)
CYP2C19	オメプラゾール(抗潰瘍剤)
CYP2C9	ワーファリン(抗凝固剤)
CYP3A4	irinotecan、paclitaxel(抗がん剤)
NAT2(Nアセチルトランスフェラーゼ2)	isoniazid(結核治療剤)
TPMT(チオプリンメチル転移酵素)	6-MPメルカプトプリン(抗がん剤)
UGT1A1(UDPグルクロン酸転移酵素)	塩酸irinotecan(抗がん剤)

# 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

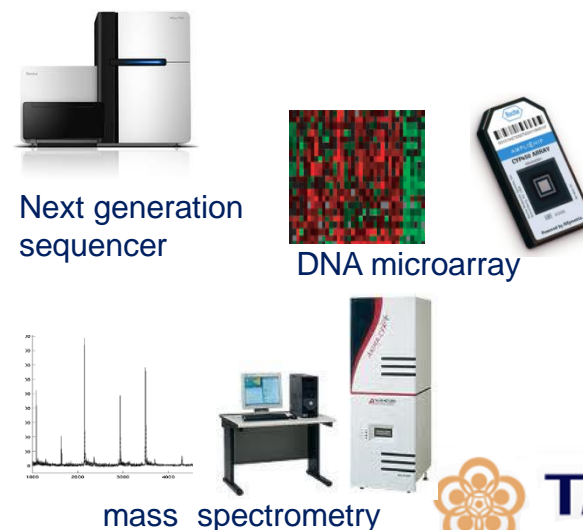
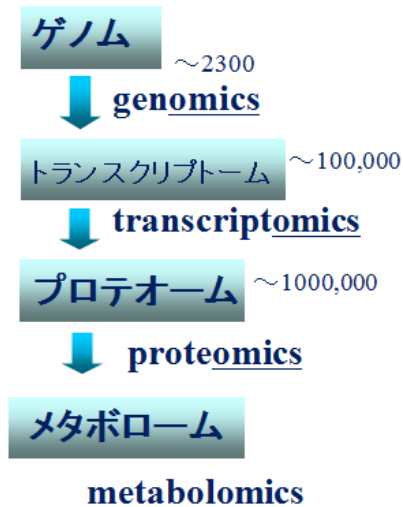
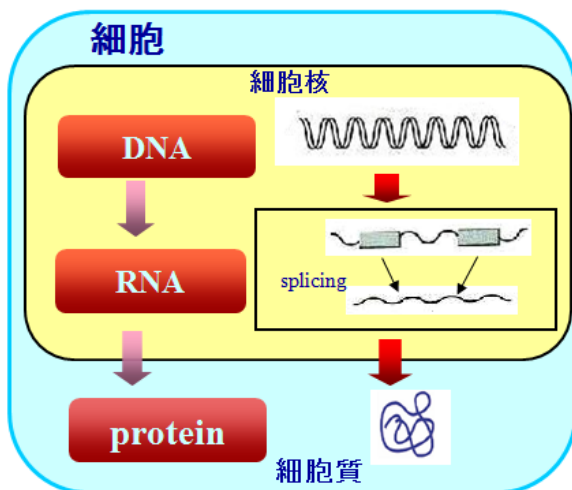
- 後天的ゲノム変異・遺伝子発現プロファイル・プロテオーム・メタボロームなど
- 網羅的分子表現型 "molecular phenome"

- 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

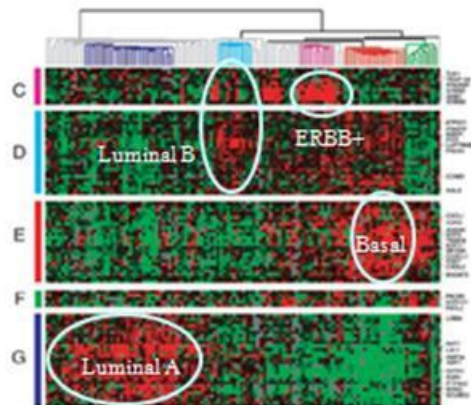
- 臨床症状や病理変化より早く変化→予測医療・先制医療





# 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(intrinsic)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"> <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>
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"> <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>
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"> <li>• Computer-based family history tool for perception of breast, ovarian</li> <li>• SLC01B1</li> </ul>

# 主なゲノム/オミックス医療の 臨床実装病院

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 <u>rx</u> 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 <u>gliomas</u> and <u>glioblastomas</u> for biomarker based RCTs</li> <li>• Targeted <u>rx</u> (such as RET inhibitor) of metastatic solid tumors based on tumor mutation status</li> </ul>
Morehouse	<ul style="list-style-type: none"> <li>• <u>Exome</u> sequencing of 1200 early onset severe African American hypertension cases and 1200 controls</li> </ul>
Duke	<ul style="list-style-type: none"> <li>• Computer-based family <u>hx</u> 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, <u>pharmacogenetic</u> analysis, identification of families for further research
Baylor	Whole <u>exome</u> and whole genome sequencing in <u>Mendelian</u> disorders to improve <u>diagnosis</u> .
<u>Geisinger</u>	<ul style="list-style-type: none"> <li>• Selection for gastric bypass surgery <u>vs</u> other <u>wt</u> 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 <u>pts</u> randomized to genetic counseling <u>vs</u> usual care</li> <li>• CYP2C19 testing in interventional cardiovascular procedures for <u>clopidogrel</u>.</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 <u>PGx</u> genotyping in children
Vanderbilt	Pre-emptive <u>PGx</u> genotyping for <u>clopidogrel</u> , 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>• <u>PGx</u> driven selection/dosing of antidepressants</li> <li>• CYP2C19 genotyping for antiplatelet <u>rx</u> post PCI</li> </ul>
Inter-Mountain	Tumor-based screening for Lynch syndrome

# 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

# 第1世代 ゲノム医療の臨床

# ゲノム配列解読の臨床応用 Clinical Sequencing

Nic Volker



- 3才の男子。2才のころから、原因不明の腸疾患で、腸のいたるところに食べると潰瘍が発生
- 130回の外科的切除手術を行うが、再発を繰り返す。これ以上行う治療がなくなった
- Nicの全エキソンの配列を次世代シーケンサ決定
- 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。

# いくつかの米国病院のゲノム医療

## Wisconsin大学病院

- Genome sequencing program
  - 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：
  - アメリカ病理学会 (CAP)およびClinical Laboratory Improvement Amendments (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

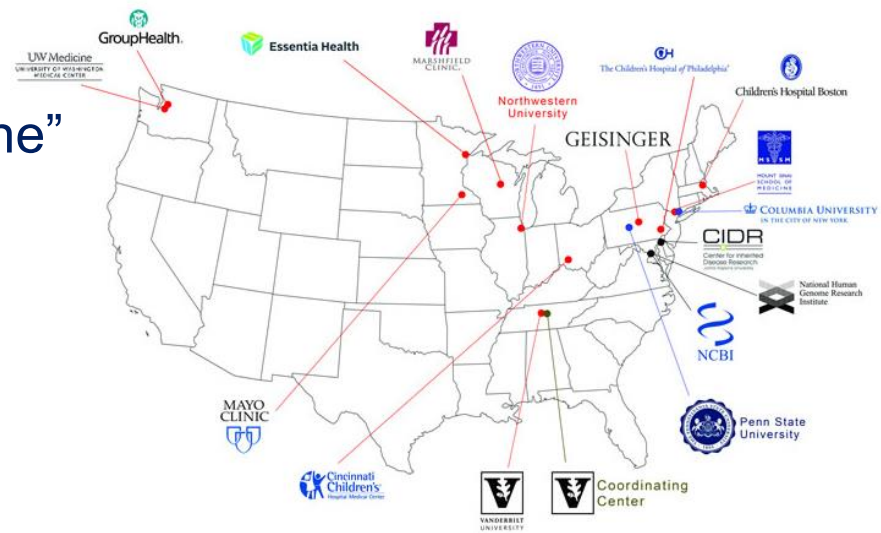
Clinical Sequencing

難治性がん

- 例 胆管がん

原因不明遺伝病

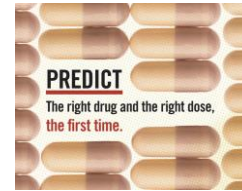
- “診断オデッセイ”



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



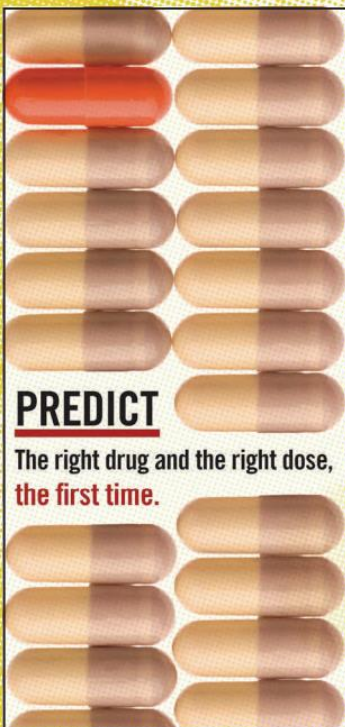
クロピドグレル処方  
電子カルテの警告画面  
商品名プラビックス：抗血栓剤  
ステント留置手術の後に処方

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

# PREDICT



## FDA Labels with Pharmacogenomic Biomarkers

Class	Sample Drugs	Enzyme
<b>Cytochrome P450</b>	Quinidine	CYP2D6
	Carvedilol	CYP2D6
	<b>Clopidogrel*</b>	<b>CYP2C19</b>
	Metoprolol	CYP2D6
	Prasugrel	CYP2C19
	Propafenone	CYP2D6
	Propranolol	CYP2D6
	Ticagrelor	CYP2C19
	<b>Warfarin*</b>	<b>CYP2C9</b>
	Atomoxetine	CYP2D6
	Fluoxetine	CYP2D6
	Fluvoxamine	CYP2C9
	Risperidone	CYP2D6
	Tiotropium	CYP2D6
	<b>Tamoxifen*</b>	<b>CYP2D6</b>
	<b>Tacrolimus*</b>	<b>CYP3A5</b>
<b>Other</b>	<b>Simvastatin*</b>	<b>SLCO1B1</b>
	<b>Azathioprine*</b>	<b>TPMT</b>
	<b>Warfarin*</b>	<b>VKORC1</b>

Drug	Estimated Adverse Events Prevented Over 5 Years
Abacavir	3
Azathioprine	17
Clopidogrel	79
Simvastatin	19
Tamoxifen	15
Warfarin	265
<b>Total</b>	<b>398</b>

\*Denotes drug-genome testing in the development pipeline for PREDICT.

# わが国での ゲノム・オミックス医療

- 研究費を用いた試行的ゲノム医療であるがわが国でもいくつかの医療施設でゲノム・オミックス医療が行われている。
- **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

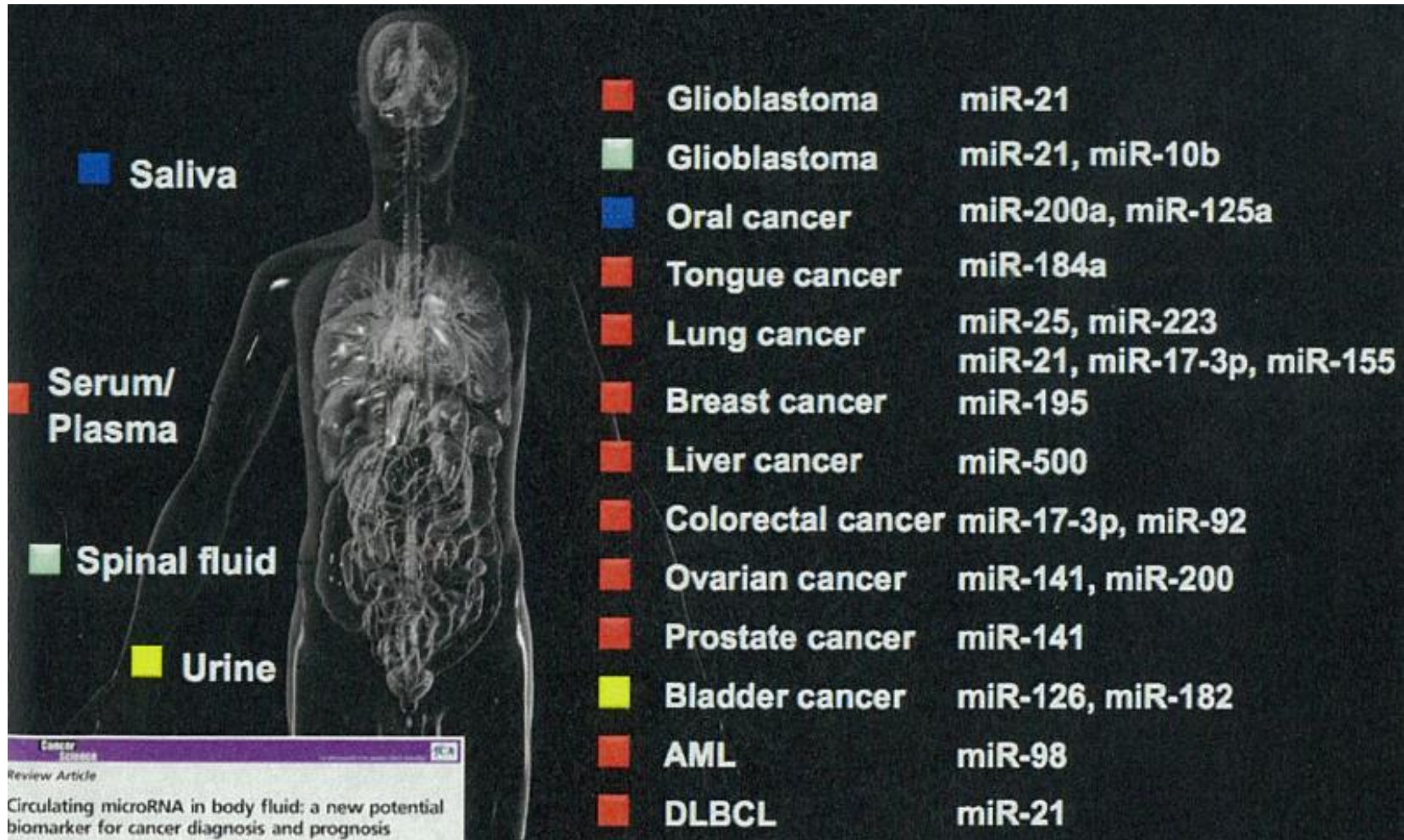
# 第2世代 オミックス医療の臨床

# 第2世代オミックス医療への期待

## 液体細胞診 (Liquid Biopsy)

- すでに乳がんでは手術サンプルの遺伝子発現プロファイルが米国では必須 OncotypeD, MammaPrintなどになっているが、手術標本やバイオプシーが必要
- 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



# ゲノムオミックス医療 ビッグデータベース構築の 動向



# 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)

# NIH「ビッグデータから知識へ」計画

“Big Data to Knowledge” (BD2K) initiative

- 以前の同様な国家プロジェクト 1999
  - “生命医療情報科学技術計画(BISTI)”
  - “Biomedical Information Science and Technology Initiative”
- **BD2K: “Big Data to Knowledge” Initiative 開始 2013**
  - 「データ・情報学に関するNIH長官諮問委員会」WG
    - WG on Data and Informatics for Advisory Committee to the Director (ACD) of NIH
  - 焦点化したワークショップ, 知識発見指標や人材養成のためのCOEに対する提案の募集
  - 研究費の配分 2014年から
  - データ科学のための副長官 (Associate Director of Data Sciences) を任命 **Boune, 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」
  - 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創成
  - Center of Excellence in Data Science
    - Uni. 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
- 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・PubMedのDB版
  - UCSD: BioCADDIEを中心にDDI研究開発

# Obama大統領と Precision Medicine Initiative

- 2015年，一般年頭教書で発表
- 個別化医療、**精密医療**（precision medicine：適確医学）の推進
- 100万人のコホート研究、GxE
- 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

# 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)

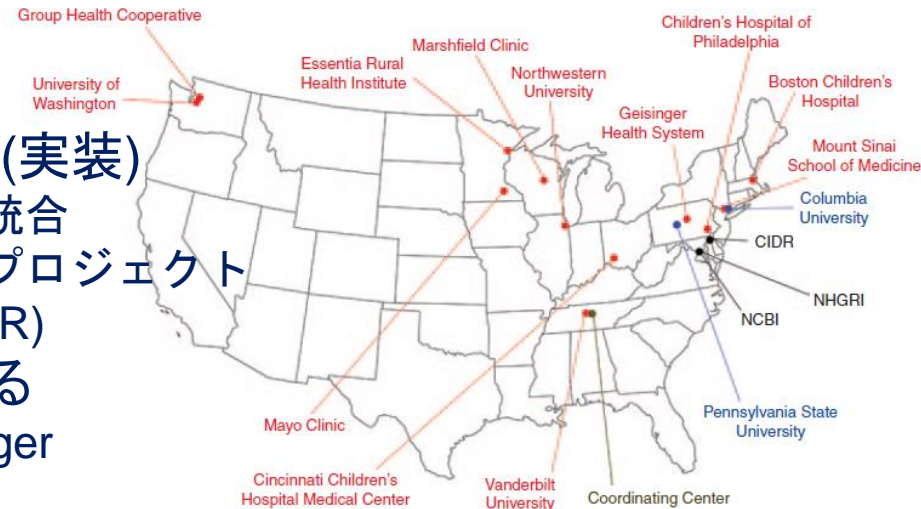
- 電子カルテを通して臨床phenotypingを行う
- EMR : 臨床phenotypingとBiorepositoryに基づくGWASが可能か (EMR-based GWAS) 利用に関するELSI側面も検討
- eMERGE-I: Mayo Clinic, Vanderbilt大学, Northwestern大学など5施設

- 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



# 臨床データの表現型形式化 (Phenotyping) の問題

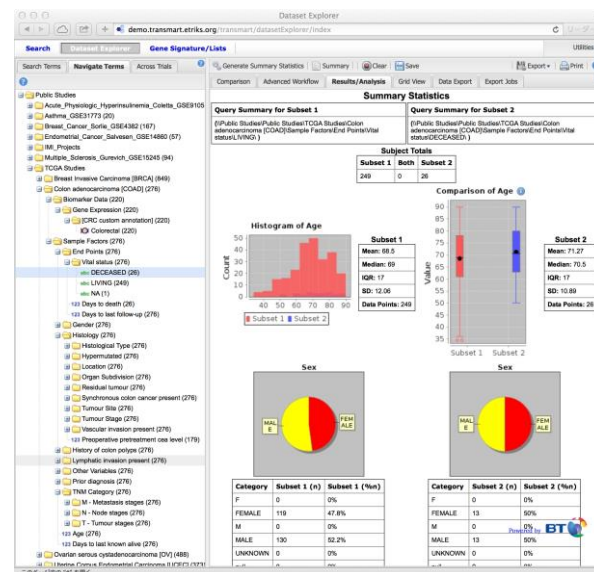
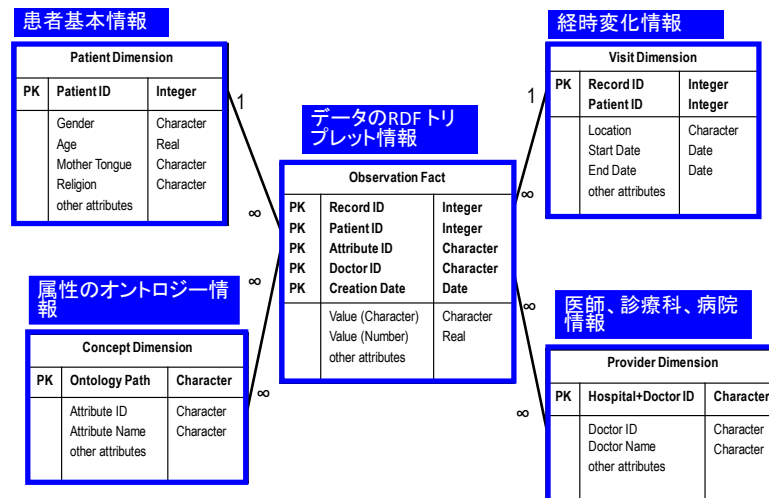
## i2b2 (Informatics for Integrating Biology and the Bedside)

- 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)のソフトウェア
- 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.



# Precision Oncology— 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)

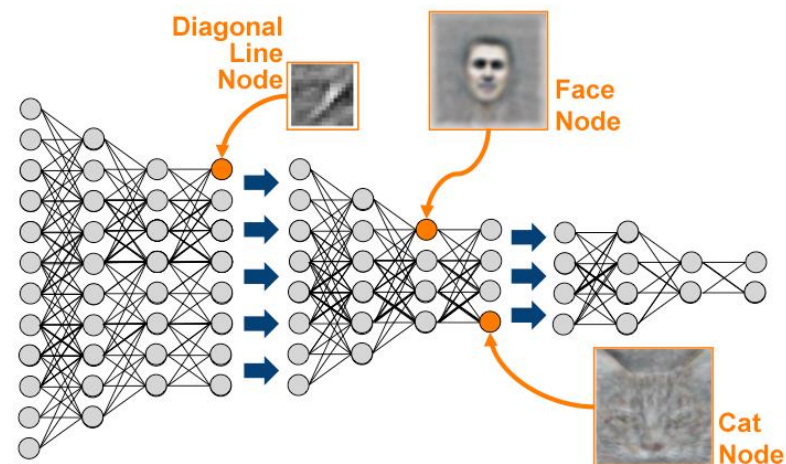
# Precision Oncology - CancerLiQ

- **Learning systemの不可欠性:** 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、ガイドライン
  - 病院の患者訪問によって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。
- 2013年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

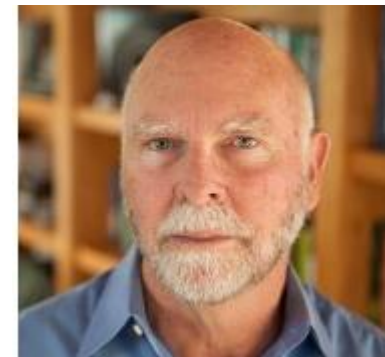


# Precision Oncology – Cancer Commons

- 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
- **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



# 米国民間ゲノムデータベース



- 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7000億円医療費削減、HiseqX 5sets
  - 一年**40000ゲノム**（幼児から老人まで、患者・健常者も）  
収集し最大のゲノムDBを作る、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に
  - Och F（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探し、今後拡大

# 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グデータ

ゲノム情報 + 臨床・環境情報(EMR等)

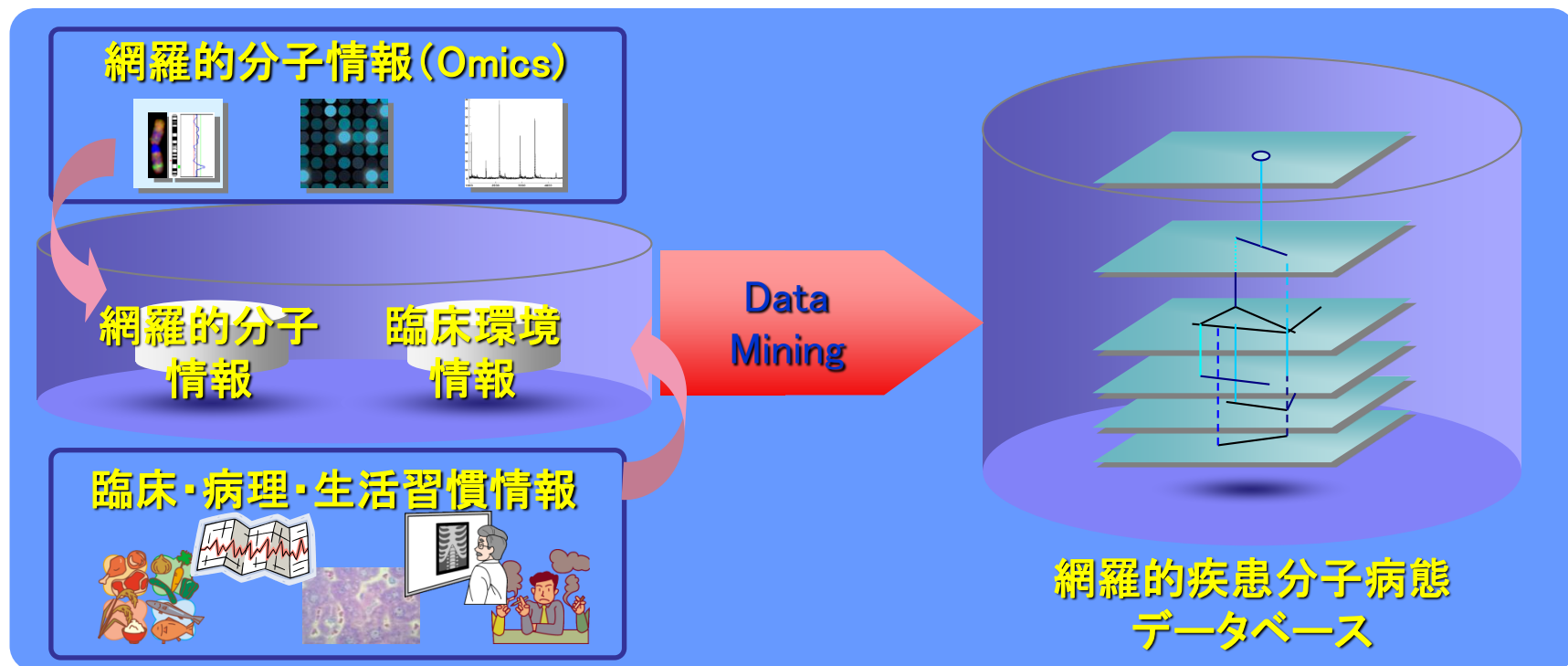
知識発見(learning) システム

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

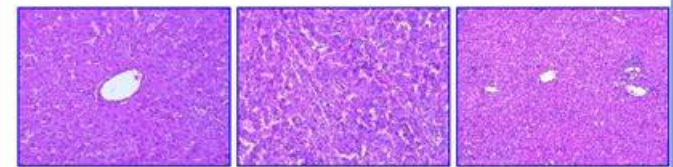
1: 0101032 Hepatocellular carcinoma (肝臓腫瘍)  
治療 肝S5重区域切除+S8部分切除  
病理 T=T1,N=N0,M=M0

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

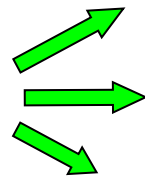
3: 01017849 Hepatocellular carcinoma (肝臓腫瘍)  
治療 肝右葉切除 (右門脈癌栓術後)

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 :



1ヶ月  
3ヶ月  
半年  
1年  
3年  
10年

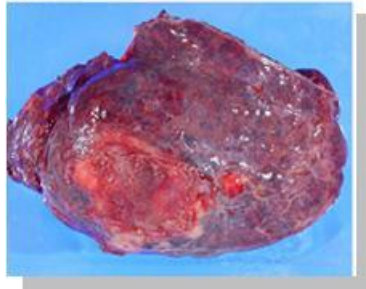
主病名  
基本情報  
症状  
背景  
疾患  
生活歴  
検査所見  
血液  
生検  
画像診断  
Angiography  
US (超音波)  
CT  
PET  
MRI  
内視鏡検査  
上部  
下部  
治療  
内科的治療  
外科的治療  
放射線治療  
予後  
病理  
肉眼所見  
組織所見  
分子情報

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



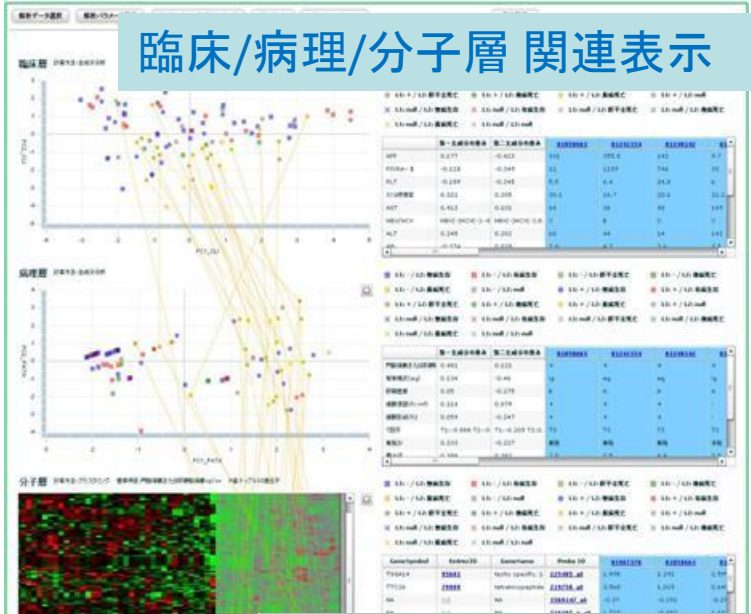
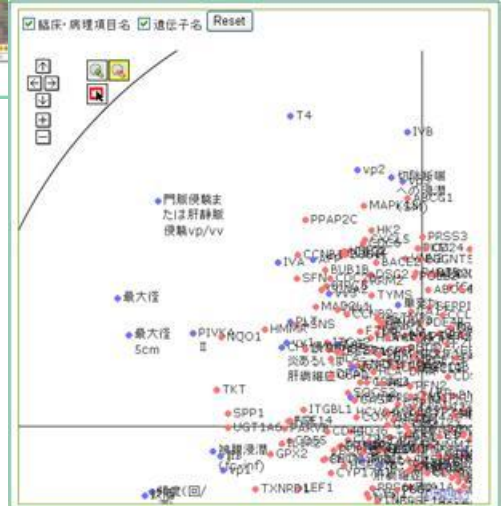
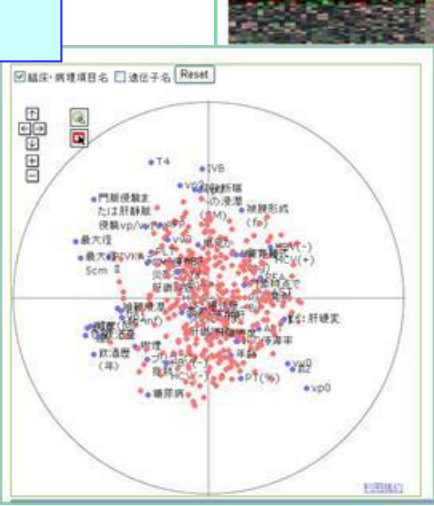
# 臨床オミックス解析

## 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解明

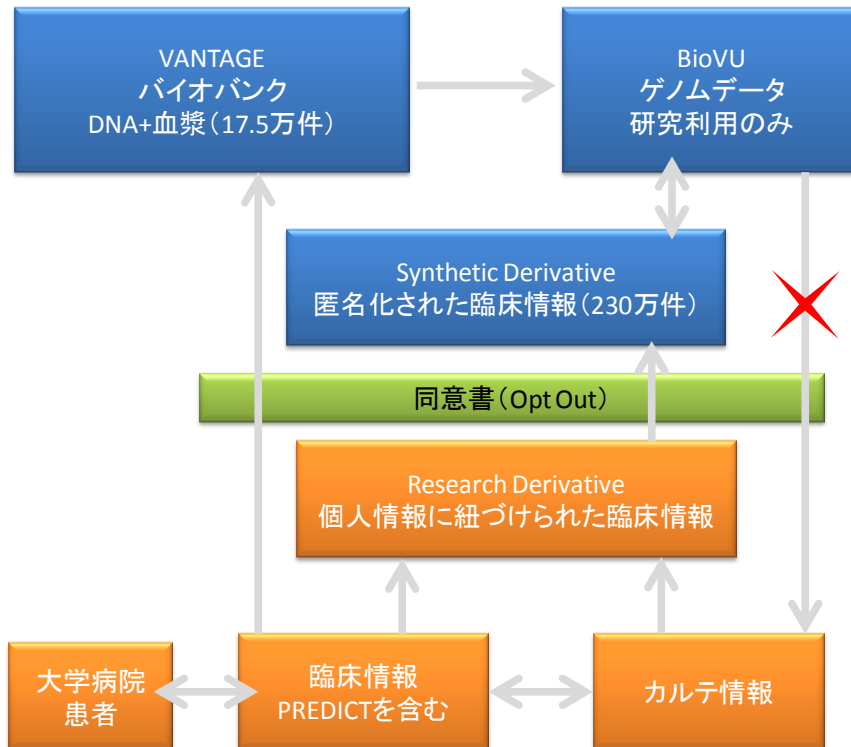
# ゲノム・オミックス医療を支える情報システム

- 統合臨床オミックスデータベース
  - 各病院に装備するか、センター方式で病院群が共有するか、2つの方式がある
  - 地域医療連携と連動する方式も考えられる
  - Population（健常人）型 Biobankとも連携が考えられるが、暫くは疾患コホート Biobankとの連携の方が容易である。
- 臨床表現型とゲノム・オミックス情報の相関関係の表現方式が重要である



# 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)

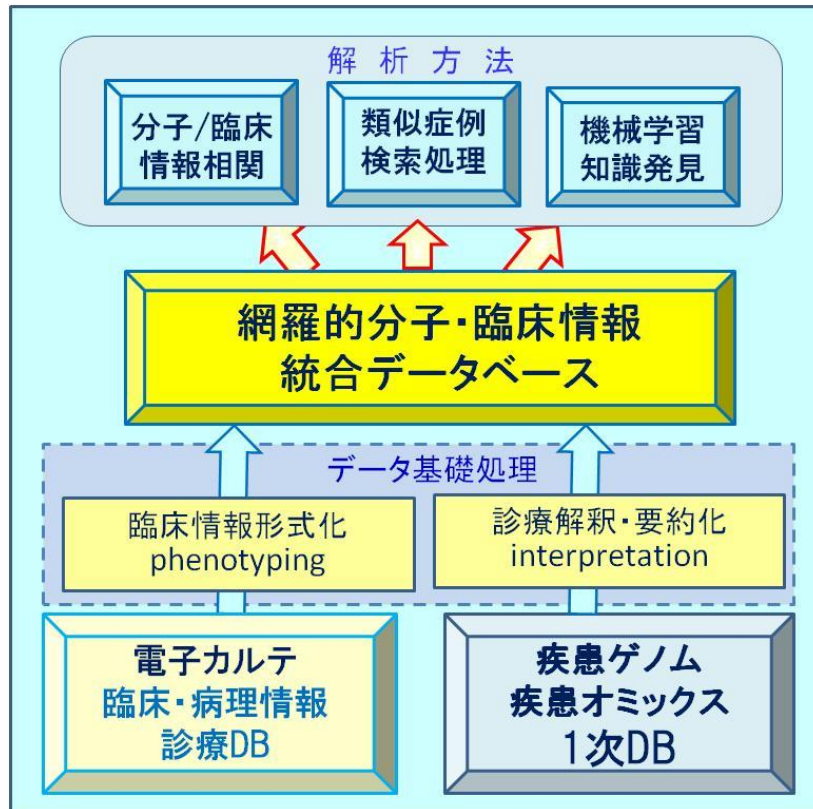
## バイオバンクと遺伝子解析

**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

# 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

# 生涯にわたる個別化予防

# 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
  - 全欧州 **BBMRI** (Biobank/Biomole. Res. Infra.) 250Biobank統合
- **東北メディカルメガバンク (Tohoku Medical Megabank)**
  - 地域住民コホート (80000) : 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 (70000) 。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, **total 150,000 subjects**
- そのほか、長浜 Nagahama コホートなど各地でBiobank

# 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

# 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



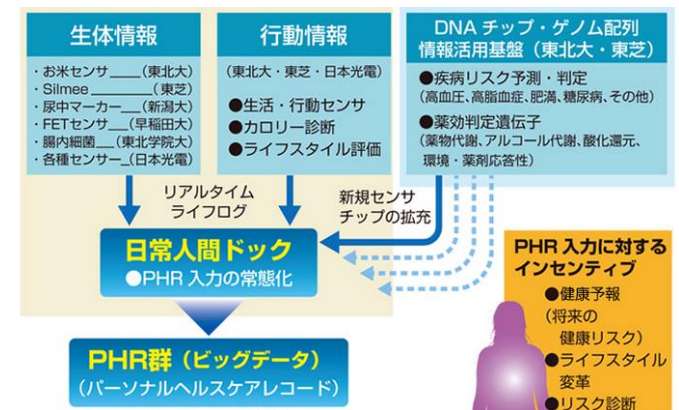
ECG; EEG; Skin Conductivity; EVG

- 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



東北大学COI 「さりげないセンシング」

# Dr. John Halamka

Professor at Harvard Medical School



Straight from the Shoulder

John Halamka, M.D.

PGP#2

N ENGL J MED 353:4 WWW.NEJM.ORG JULY 28, 2005

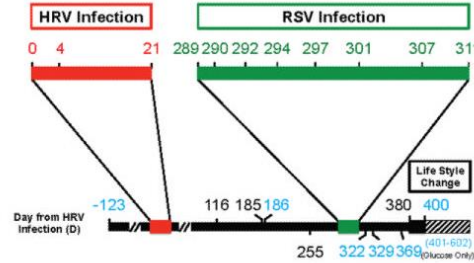
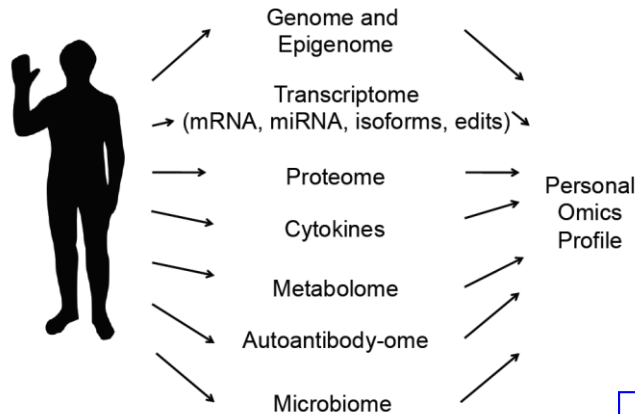
The New England Journal of Medicine

- PGP#2

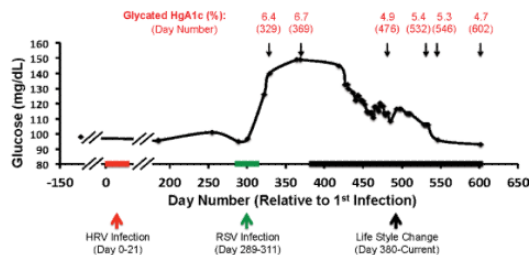
- 患者に薦める前に自分で「人体埋め込み用RFID」を体験し、あらゆる面から調査
- 米粒ほどの大きさで、電波を使って外部の装置と情報をやり取り可能
- データベースと照合することで、同RFIDを埋め込んだ人の医療情報などが取り出せる

# iPOP (integrated Personal Omics Profiling)

## Personal "Omics" Profiling (POP)



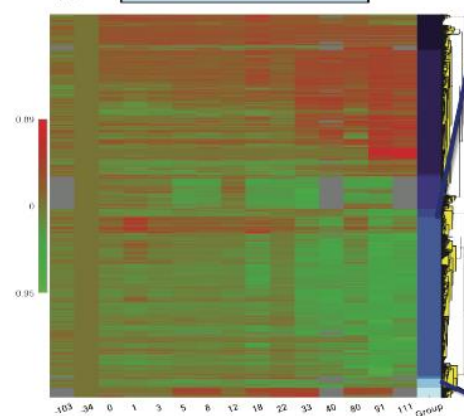
統合個別化オミックス  
 プロファイルの時系列  
 分析 (Fourier分析) に  
 より統合オミックスの  
 乱れを検知



## Integrated Omics clustering



### (I) Autocorrelated data clusters

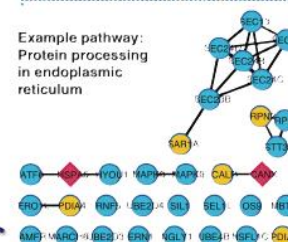


RSV

Full Reactome (FI) known pathway map for cluster:



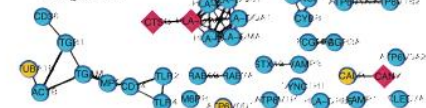
Example pathway: Protein processing in endoplasmic reticulum



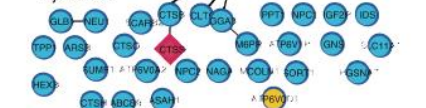
Dynamic expression pattern observed in:



Example pathway: Phagosome



Example pathway: Lysosome



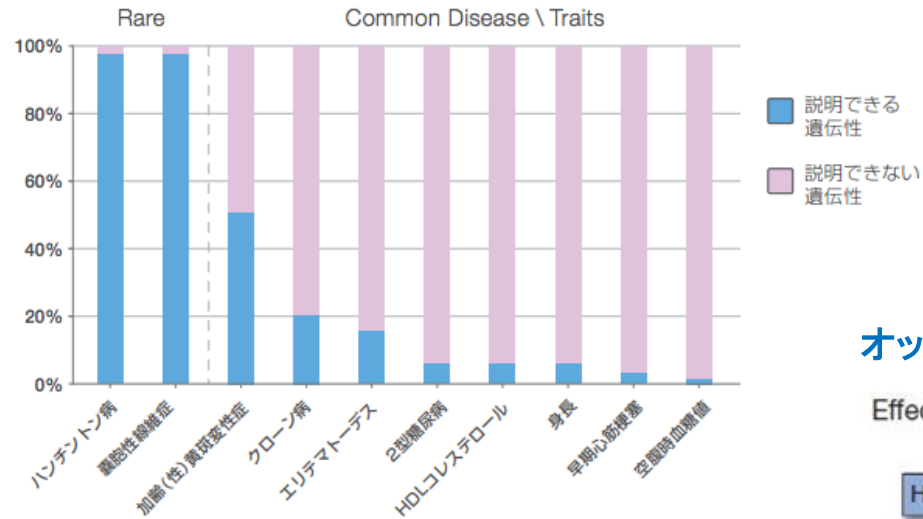
Example pathway: Insulin





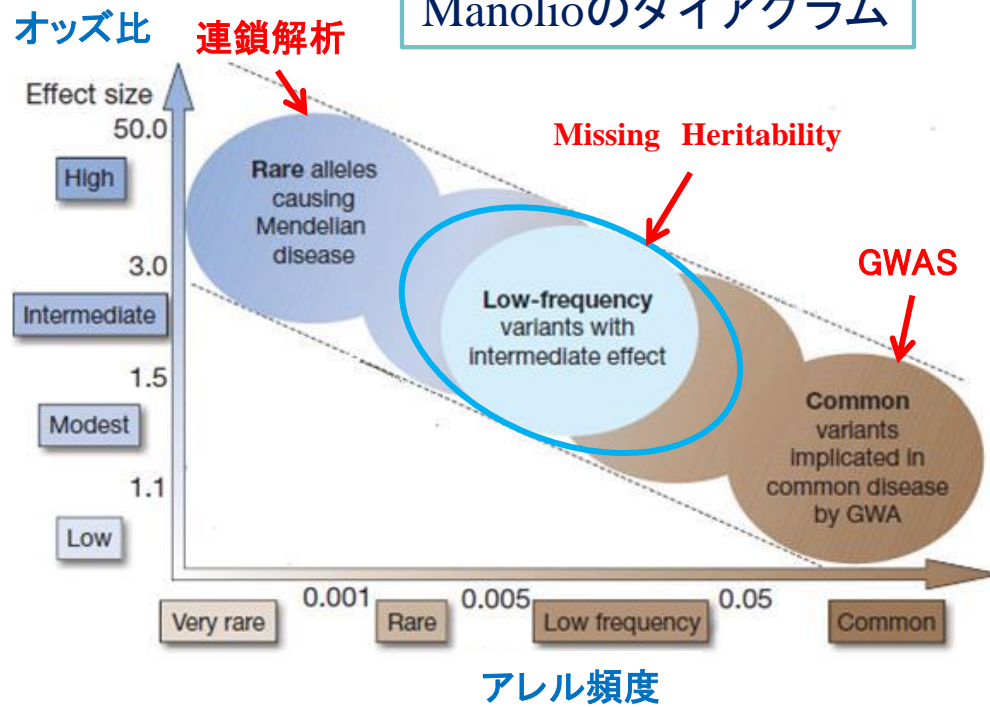
# 疾患理解の深化:失われた遺伝継承性

## Missing Heritabilityとシステム分子医学



連鎖分析やGWASでは遺伝素因のほとんどを説明できていない

Manolioのダイアグラム



■  $\pi_{\text{explained}} = h^2_{\text{explained}} / h^2_{\text{all}} < 1$   
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

E.Lander; The mystery of missing heritability: Genetic interactions create phantom heritability、PNAS109,2012

Teri A. Manolio, Francis S. Collins et al. Finding the missing heritability of complex diseases, vol 461, 2009



# 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**

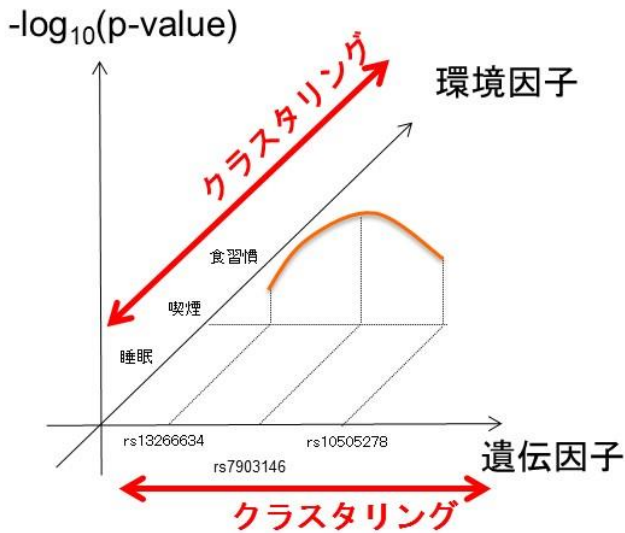
		CYP1A2 Phenotype ≤ Median		CYP1A2 Phenotype > Median	
		Likes rare/medium meat	Likes well-done meat	Likes rare/medium meat	Likes well done meat
Non-Smoker	NAT2 Slow	1	1.9	0.9	1.2
	NAT2 Rapid	0.9	0.8	0.8	1.3
Ever-Smoker	NAT2 Slow	1	0.9	1.3	0.6
	NAT2 Rapid	1.2	1.3	0.9	<b>8.8</b>

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

# 疾患の発症に相乗的に関連する 遺伝因子・環境因子相互作用の同定



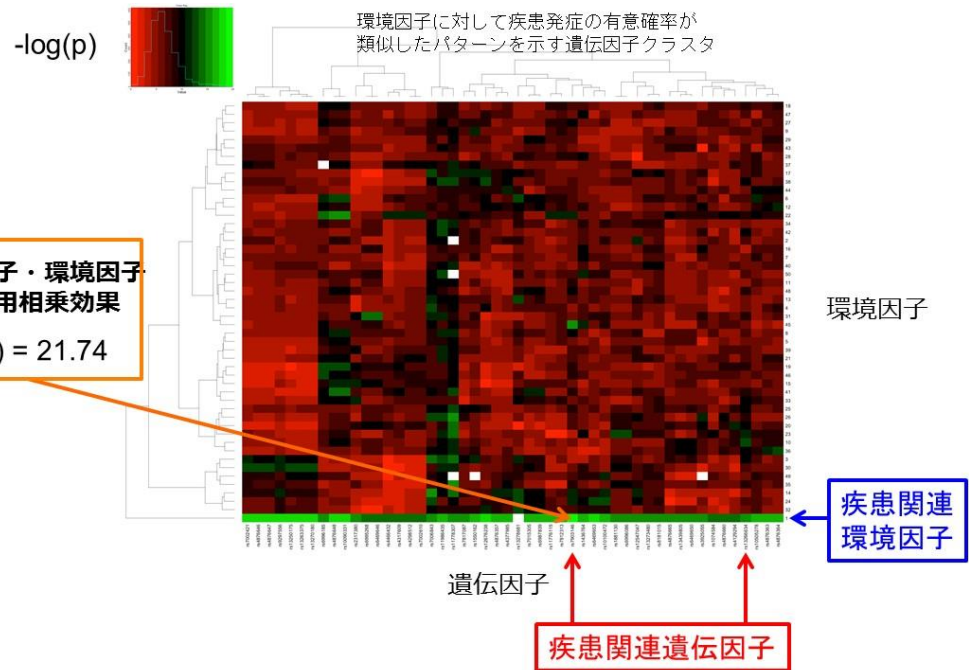
遺伝因子・環境因子相互作用の同定

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

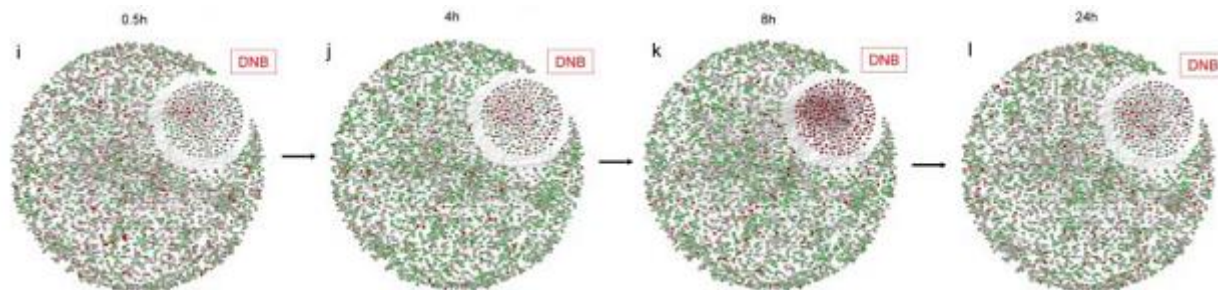
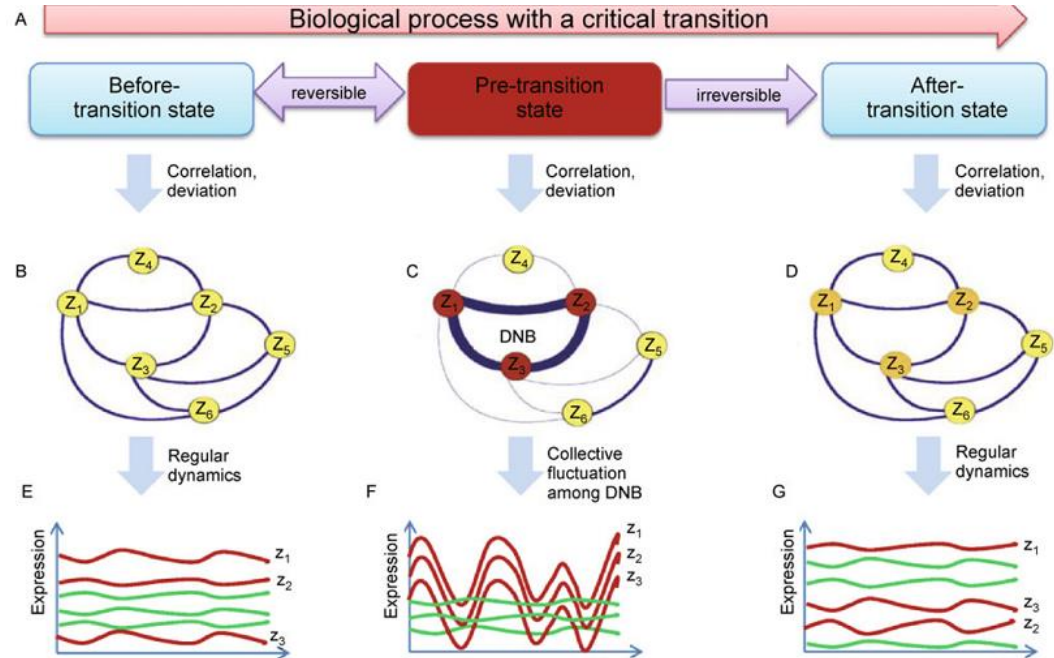
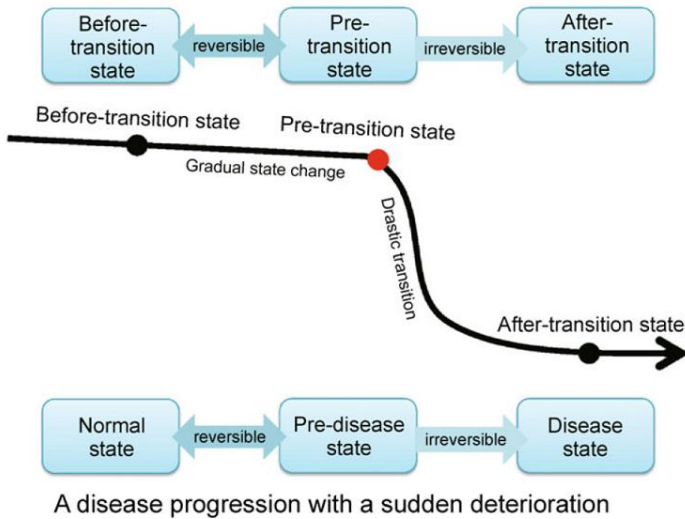
unaffected		environment	
		0	1
genotype	0 (aa)		
	1 (aA)		
	2 (AA)		

遺伝因子・環境因子  
相互作用相乗効果  
 $-\log(p) = 21.74$

GxE heat map



# Chenの疾患発症のネットワーク バイオマーカー



# Life-long (life-course) health care

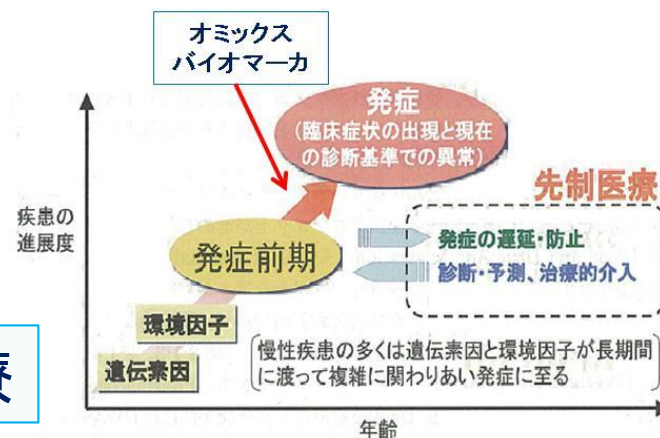
- 日常的生体モニタリングの目標
  - Life-long healthcareにおける疾患発症の予知へ
- 発症予測と分子的バイオマーカー
  - 先制医療 (preemptive medicine) へ
  - Liquid biopsy (液性バイオマーカー) 血中循環RNA, DNA計測, がん・アルツハイマー症などの予知
  - DIY genomicsの発展
  - 製薬会社も先制医療薬へ



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

ご清聴ありがとうございました

