

ビッグデータ医療における 人工知能（AI）の展望

東京医科歯科大学 名誉教授 特任教授（医療データ科学推進室）
東北大学 東北メディカル・メガバンク機構 特任教授
機構長特別補佐（情報・システム担当）

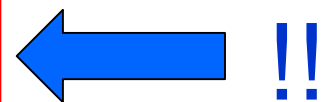
田中 博



医療ビッグデータ時代の到来

- (1) 次世代シーケンサ (Clinical Sequencing)による「ゲノム/オミックス医療」における網羅的分子情報収集/蓄積
- (2) Biobank/ゲノムコホート普及による分子・環境情報の蓄積
- (3) モバイルヘルス(mHealth) によるWearable センサの連続計測による生理データの蓄積 (unobstructed monitoring)

急激な大量データの出現
コストレス化かつ高精度化



ゲノム : 13年→1日(1/5000) 3500億→10万円(1/350万)

個別化医療・医療の国民レベルの向上
医療/ヘルスケアの適確性の飛躍的な増大



医療におけるビッグデータ

I 次世代シーケンサによるゲノム情報

— 網羅的分子情報の急速な蓄積

II 大規模バイオバンクによる情報蓄積

— ゲノム情報・環境生活関連情報

III モバイルヘルスによる生理変量

— 連続的生理モニターによる情報蓄積

新しい
タイプの
医療ビッグ
データ

IV 表現型医療情報の蓄積

— 電子化の普及による医療情報の蓄積

旧来のタイプの
医療データの
大容量化

医療の「ビッグデータ革命」

～ゲノム・オミックスデータの基軸的な特徴～

＜目的もデータ特性も従来型と違う＞

従来の医療情報(IV)の「ビッグデータ」

N -Big Data ($n \gg p$)

医療情報・疫学調査では属性数：数10項目程度

- 目的：Population MedicineのBig Data
⇒個別を集めて「集合的法則」を見る

網羅的分子情報(I~III)などビッグデータ

P -Big Data ($p \gg n$)

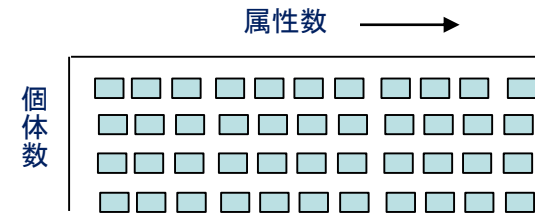
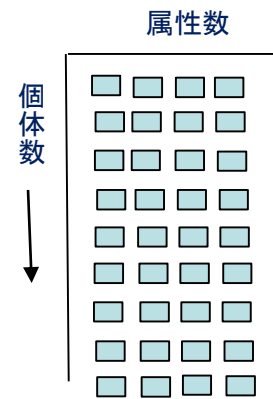
1個体に関するデータ属性種類数が膨大

属性に比べて個体数 少数:従来の統計学が無効

「新NP問題」：多変量解析:GWASで単変量解析の羅列

- 目的：例えば医療の場合Personalized Medicine

⇒大量データを集めて「個別化パターン」の多様性を抽出



新しいデータ科学の必要性

ビッグデータは 医療のパラダイムを変革する

- 医療は近年大きくパラダイム変換しつつある。
- 2000年頃から、「ビッグデータ医療」の概念出現の前に、パラダイム変換の概念として、次の2つの概念が提示されてきた

個別化医療

Personalized Medicine

先制医療

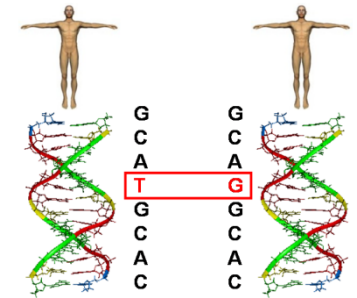
Preemptive Medicine

個別化医療

personalized medicine

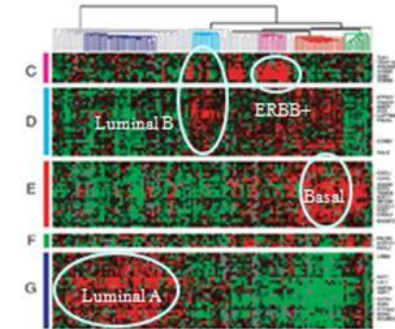
- 別表現
 - 層別化医療(stratified), 個別化医療(personalized), **P4 医療**(predictive, preemptive, preventive, participatory)
 - 主にゲノムなどの疾患機序の網羅的分子情報の進展による
- 2つの個別化医療の概念
 - 先天的遺伝的素因（発症の個人的遺伝素因）
 - 疾患原因遺伝子causative gene, 感受性遺伝子 disease susceptibility gene、HapMapプロジェクト(2003)から
 - 後天的要因（疾患の個性：がんの遺伝子変異）
 - 後天的遺伝子変異、特異的遺伝子発現、ICGC(2008)から

個別化医療の概念



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ります

- **疾患感受性遺伝子** (Disease susceptibility gene)
 - **多型性**：一塩基多型(SNP), 1億個位,
 - 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) 疾患相対リスク**
 - HAPMAPプロジェクト、1000ゲノムプロジェクトなど
 - **薬剤代謝酵素の多型性**→ 有効性・副作用の評価
- **遺伝子発現プロファイルによる疾患の内在的分類**
 - **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による分類
 - **疾患の進行度の評価**
 - 疾患組織や病態進行の段階に依存
 - 臨床症状や病理変化より早く変化→ **予測医療・先制医療**



個別化医療による 医療のパラダイム変換

- 従来のpopulation医学<One size fits for all>はもはや成り立たない。
 - 同一の病名で括られているが、内在的亜型 (intrinsic subtype) が多数存在する
 - 第1種：薬剤代謝酵素の多型性
 - 第2種：がんの亜型解析と分子標的薬の選択
- 医療の隅々に浸透するポピュレーション医学の桎梏を克服

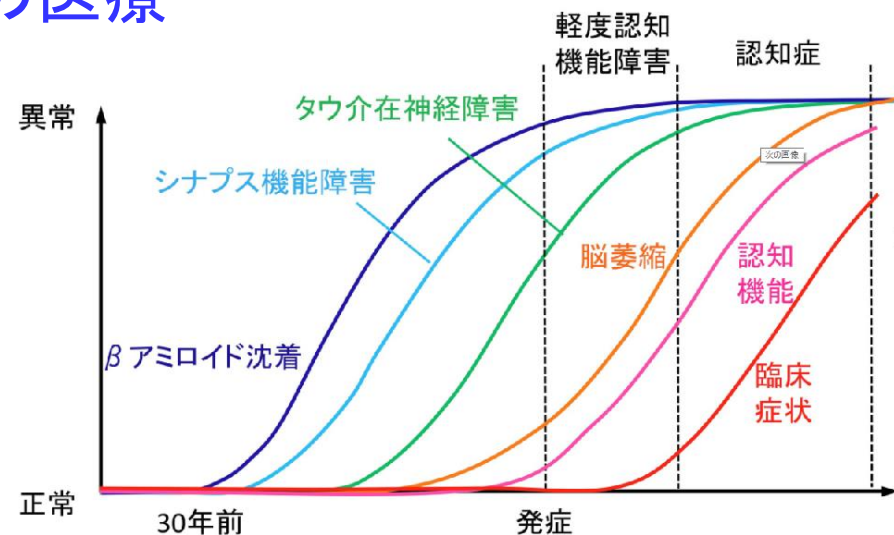
先制医療

preemptive medicine

- Preemptive medicine (2005)
 - NIH長官Zerhouni が2005年NIH長期計画で
 - “By making use of precise molecular knowledge (分子情報を使用) to detect disease before symptoms are manifest, and intervening before disease can strike.”
(発症前に検出・治療) .
- その後NIHの将来計画に頻出

先制医療による 医療のパラダイム変換

- 従来の医療：疾病罹患後の医療
 - 対応的 (reactive)
 - 機会主義的 (occasional)
- 疾患罹患後の治療
 - 治癒の困難性
 - 例：アルツハイマー治療
 - 医療経済的圧迫
- 早期診断・早期治療
- 予測医療・予防医療(proactive)
- 疾患に関する見方の変化
 - 生涯にわたる疾患把握
- **Life-course-oriented healthcare**



Jack CR Jr, et al. Lancet Neurol. 2010;9:119-128.

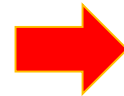
医療の長期的パラダイム変換

Population医療



個別化医療

Reactive 医療



Proactive 医療

Occasional 医療



Life-long 医療

ビッグデータ医療はパラダイム変換を実現可能にする

ビッグデータの効果(20~30年)

I ゲノム・オミックス情報

疾患成立機序の解明

II 大規模バイオバンク情報

長期生涯疾患過程の解明

III モバイルヘルス情報

短期生涯疾患過程の解明

Disease Big Data

個別化医療・先制医療

生涯医療・先制医療

Life-long Big Data

医療の「ビッグデータ」革命は どんな既存のパラダイムに挑戦しているか

- Population medicineのパラダイム転換
 - <One size fits for all>のPopulation医療はもはや成り立たない
 - 個別化医療 “Personalized (Precision) medicine”
 - 個別化医療を実現するために<個別化・層別化パターン>を網羅的に調べる：どこまでの粒度で個別化・層別化すればよいか
- Clinical research（臨床研究）のパラダイム転換
 - 臨床研究を科学にする従来の範型RCTは、個別化概念に破綻した
 - <statistical evidence based>呪縛からの解放
 - 「標本」統計・「推測」統計学に限定されない臨床研究
 - Real World Data: ビッグデータ知識生成（BD2K）

ゲノム・オミックス医療の展開

わが国での「ゲノム・オミックス医療元年」

2015年AMED（医療研究開発機構）発足

「ゲノム医学実現推進協議会」（中間報告）2015.7

「全国遺伝子医療部門連絡会議(10.18)」

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

NGS臨床応用セッションに

研究費を用いた会員の中で「臨床応用を実施している部門は12施設」

アンケート結果が発表。東大病院ゲノム医学センターなど

25～40%程度の原因遺伝子同定

AMED：IRUD（Initiative on Rare and Undiagnosed Disease）

未診断疾患の原因遺伝子をIRUD拠点病院が審査して解析セン

ターがシーケンシング。その後、DB化する。米国UDP、

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

がんの網羅的分子診断と個別化治療

— 国立がん研究センター東病院

- ドライバー遺伝子の診断。分子標的薬の治験グループに割当て

— 静岡県立がんセンター 上記と同様の内容のプロジェクト

— 京大腫瘍内科（OncoPrime）、岡大、北大、千葉大 診療施設併設型BB

ゲノム医療では、米国と水を空けられている。しかし、Biobank/Genomic Cohortでは大きく遅れてはいない

医療ビッグデータの時代の到来と 米国の最新の状況

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

ゲノム医療の2つの流れ

- 米国の流れ

- 次世代シーケンサの急激な発展による「シーケンス革命」からの怒濤の展開（2010から）
- 「治療医学」レベル質的向上のためにゲノム情報を取り入れた臨床実装の推進
 - 稀少疾患の原因遺伝子変異の同定
 - がんのドライバー遺伝子変異の同定と分子標的薬の選択
 - 薬剤代謝酵素の多型性の同定と個別化投与

- 欧州の流れ

- 社会福祉国家の理念より国民医療（医療の国民レベル）の向上
- 「予防医学」レベル質的向上のためにゲノム情報を取り入れたバイオバンク推進
- 大規模前向きpopulation型バイオバンク/ゲノム・コホートの確立
 - 遺伝的素因だけでなく環境要因（生活習慣）との相相互作用を解明し、「ありふれた疾患」発症を予測し、これに基づいて個別化予防する。
 - 疾患を発症前に対応して発症を防ぐ「先制医療(preemptive medicine)」や「予測医療(predictive medicine)の実現を目的

次世代シーケンサのインパクト

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

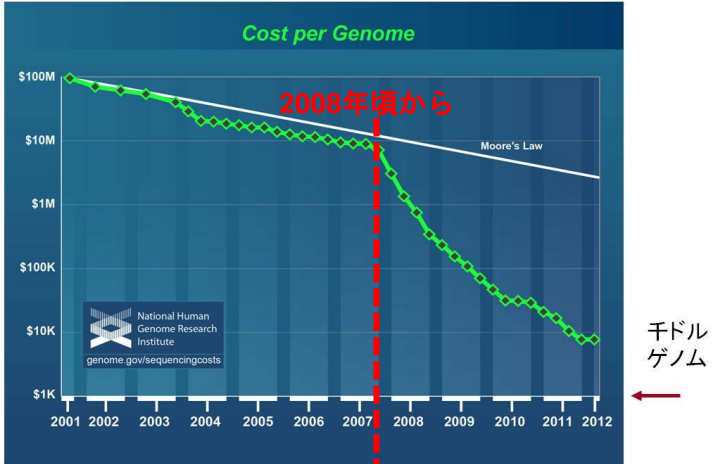
急速な高速化と廉価化 ヒトゲノム解読計画13年,3500億円⇒1日,10万円

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



	HiSeq2500		Ion Proton
本体価格	約1億円		約3500万円
モード / チップ	ハイアウトプット	ラピッドラン	Ion Proton I
解析時間	11日	27時間	2時間
リード長 (bp)	2 x 100	2 x 150	200
データ産出量 (Gb)	約600	約120	10
試薬コスト (ヒト1人全ゲノム)	数十万円		不可 エクソームのみ

HiSeq X システム 10台構成 (経費1/5)



DNA Sequencing Cost: the National Human Genome Research Institute

シーケンス革命 2007/8

ゲノム(配列決定)機器の進歩は、計算機のムーアの法則を越えている!

ゲノム医療やゲノム創薬の臨床実現を本当に推進したのは、「ヒトゲノム解読計画」ではなくて、2007年に起こった「シーケンス革命」であった。



米国におけるゲノム医療の 怒濤の展開

米国におけるゲノム医療の開始

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

(1) 原因不明先天的疾患(undiagnosed disease)

原因遺伝子の臨床の現場で(POC)の診断

次世代シーケンサの爆発的發展を受けて

Wisconsin 医科大学での全エキソーム解析

(2) 薬剤の代謝酵素の多型性の検査

臨床の現場で電子カルテの警告(診療支援)

Vanderbilt大学病院の先制ゲノム薬理

ゲノム医療：少数の予想される遺伝子の変異を調べる候補遺伝子アプローチはすでに「遺伝子医学」で行われていた。あらかじめ候補遺伝子を決めず、網羅的でデータ駆動的なゲノム解読（ゲノム網羅的アプローチ）によって変異を見出す医学である

ゲノム・オミックス医療の進展とビッグ・データ

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



2010

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

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

Wisconsin医科大学
臨床シーケンス初例
大きなインパクト

第1世代

Early adopter
時期

Baylor医科大学
Mayo Clinicなど
後続病院多数

2013
前後

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

ビッグ
データの
概念

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

第2世代

国規模の計画/全国Consortium
時期

2015

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

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

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

ゲノム医療の最初の臨床実装

ゲノム医療の第1の流れ 未診断病のClinical Sequencing



Nic Volker

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



XIAP (X連鎖アポトーシス阻害タンパク質遺伝変異 TGT(cysteine)→TAT(tyrosine) (203番目)
アポトーシスの阻害因子 免疫系が腸を攻撃する自己免疫を阻害 これまでのヒトゲノム配列で見出されていない ショウジョウバエからチンパンジー見いだせず

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



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



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

- Wisconsin医科大学 Genome sequencing program

- Nic君に続いて（翌年3月まで6例）

- 候補選択（nomination）

- 従来の検査・診察で診断困難な症例

- Multidisciplinary 患者選択委員会でレビュー

- 6-8時間のアセスメントとカウンセリング

- 32 全ゲノム, 550 全エクソーム（2015年4月まで）

- アメリカ病理学会（CAP）およびClinical Laboratory

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

- データ解析：in-houseのBIで

- Baylor医科大学病院 2番手（すでに準備?）

- Wisconsinに続いて臨床ゲノム配列解析

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

- In-houseでシーケンシング/変異分析

- CAP/CLIA認証の検査室を病院内に立ち上げる。

- 臨床分子遺伝学者によって解析・結果報告

- そのほかにWashington大学、Partnerなど多数つづく



Wisconsin
小児病院

Wisconsin 医科大学 (MCW)



Froedtert 病院

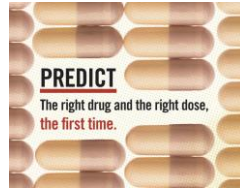


Baylor医科大学



ゲノム医療の第2の流れ

薬剤代謝酵素多型性のゲノム医療 バンダービルト大学病院



■ PREDICTプロジェクト

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

Pharmacogenomic Resource for
Enhanced
Decisions in Care and Treatment



Clopidogrel Poor Metabolizer Rules

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

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

Treatment modification is recommended if not contraindicated:

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

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

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

Click here for [more information](#)

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

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

If not using prasugrel, please select a reason:

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

Click here for [more information](#)

Cancel Order

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

Back Home Close

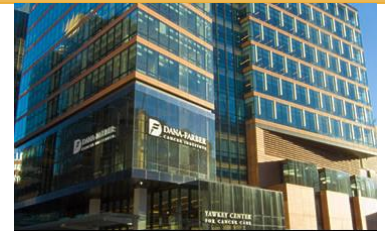
クロピドグレル処方
電子カルテの警告画面
商品名プラビックス：抗血栓剤
ステント留置手術の後に処方

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

この患者の場合(*2/*2)プラスゲレル
(商品名エフィエント)に替えるか

分量を2倍にしると警告している

ゲノム医療の第3の流れ



著名ながんセンター Dana Faber /MD Andersonなど

臨床シーケンスにより難治性がんのドライバー変異の同定する

組織限局的な後天的ゲノム変異のクリニカル配列解析
がんゲノムアトラス (TCGA : 2006年~) および
国際がんゲノムコンソーシアム (ICGC : 2008年~)

50種のがんを500症例の全ゲノム配列解析

2012頃から成果発表と始まった(我が国も肝臓がん)

患者個人70余の変異、全集合で3000を超える変異

がんを推進させるDriver変異と偶発的なPassenger変異

ゲノム・オミックス医療の 3つの流れ

2008年

2009年

2010年

2011年

2012年

2013年

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

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

ゲノム多型性の認識
.Hapmap2002開始
GWAS研究の興隆

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

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

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

ゲノム・オミックス医療
臨床実装(clinical implementation)

ゲノム/オミックス医療－米国の状況

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

NHGRI Working Groupのリスト

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

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

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

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

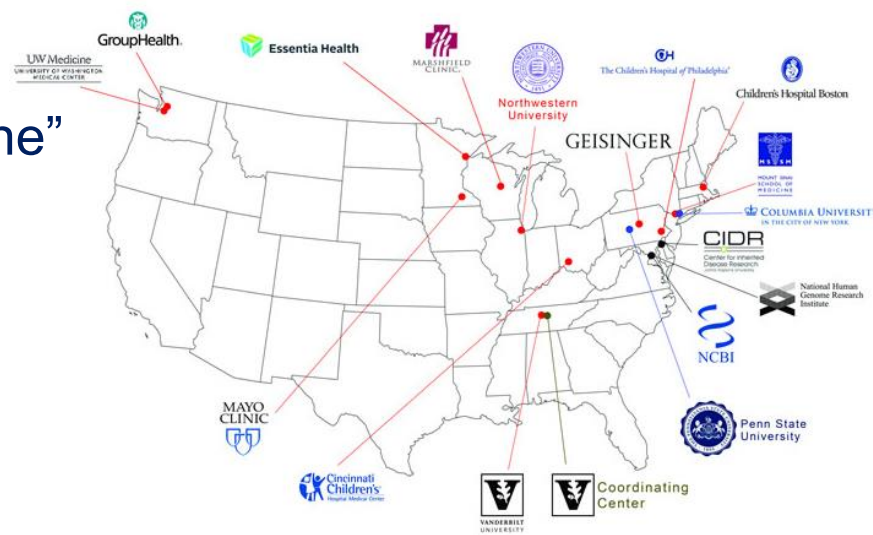
Clinical Sequencing

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



医療ビッグデータ時代の到来（米国）

ゲノム医療の実践

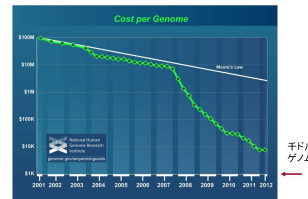
第1段階 ゲノム医療の発展

次世代シーケンシングの臨床普及 (2010~)

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

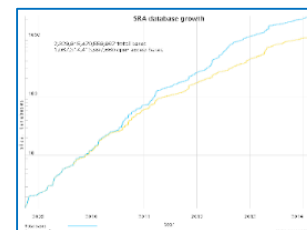
米国では数十の著名病院で実施

ゲノム・オミックス情報の蓄積



DNA Sequencing Cost: the National Human Genome Research Institute

2000兆塩基 (2 Pb)
が登録(NCBI:SRA)



医療ビッグデータ

第2段階 医療ビッグデータ時代

医療情報との統合

電子カルテからの
臨床フェノタイプ

医療ビッグデータ

学習アルゴリズム

ゲノム医療知識

人工知能AI



MayoClinicでは
10万人患者WGS

ゲノム・オミックス医療の進展とビッグ・データ

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

2010

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

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

Wisconsin医科大学
臨床シーケンス初例
大きなインパクト

第1世代

Early adopter
時期

Baylor医科大学
Mayo Clinicなど
後続病院多数

2013
前後

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

ビッグ
データの
概念

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

第2世代

国家政策/全国Consortium
時期

2015

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

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

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

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>

国家戦略としての「医療ビッグデータ」

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

疾患の3要素—— 遺伝素因、環境要因、臨床表現型

① Genome/Omics情報（遺伝的素因、分子表現型）

1. 先天的な遺伝的素因：遺伝子変異、多型性(遺伝子型)
生得的で「生まれ持って有する」変異や多型性の情報
2. 後天的に生成された特定組織の遺伝子変異（がん）
3. 疾病とともに後天的に生成され、臨床的にいろいろな疾患形成過程の時点で計測された各種疾患オミックス・プロファイル情報(注1)
疾患の今後の病態進行を予測するバイオマーカの役割
疾患の内在的なサブタイプを現すバイオマーカの役割
(注1) 「分子表現型」として遺伝子要因・遺伝型情報としては扱わない

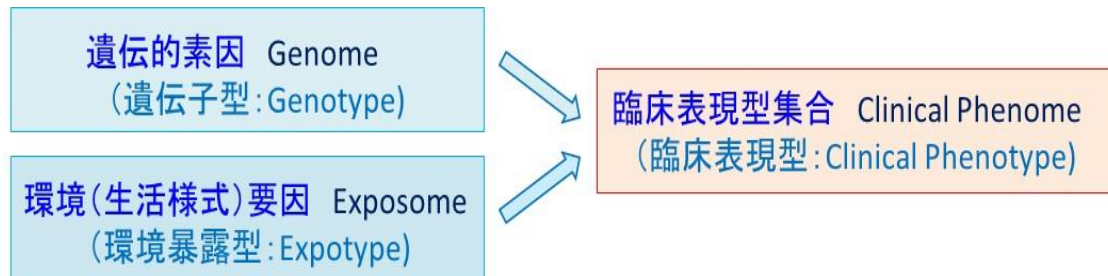
② Exposome情報(環境要因・暴露因子・生活様式・介入要因)

1. 環境情報
 - (1) 物理化学的環境情報
環境物質など疾病発症に関係のある環境因子
 - (2) 病原体をはじめとする疾患発症に関連する生物学的環境情報
病原的な細菌やウイルスなどの環境情報
2. 生活習慣・様式情報（注2）
 - (1) 食事内容・食習慣等の情報
 - (2) 運動量などに関する情報

③ Clinical Phenome（臨床表現型集合）情報

1. 主観的臨床表現型
主訴などの患者の訴え
2. 客観的臨床表現型
 - (1) 臨床検査などの各種の検体検査情報
 - (2) 生理検査や内視鏡検査などの生体検査情報
 - (3) X線・MRI・CTなどの医用画像情報
3. 薬剤治療などの治療的介入に関する情報（注2）
 - (1) 薬剤処方の種類と投与歴
 - (2) 手術や処置などの介入情報

(注2) 本来なら環境介入情報であるが、臨床経過情報の1種として臨床表現型情報の項目に含める場合も多い
3. 病態経過・転帰など
予後や転帰など疾病経過に関する情報



臨床表現型との統合(phenotyping)

臨床表現型データ 検証的抽出、「非構造化データ障壁」

electronic **ME**dicinal **R**ecords + **GE**nomics (NHRI-funded)

phase I (2007-2011) EMR-basedゲノム研究の探求

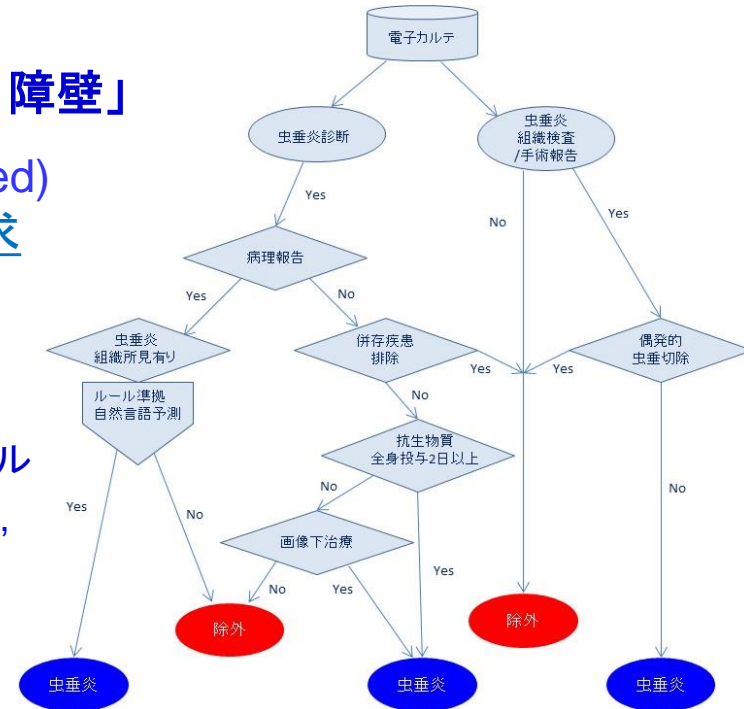
- EMR(臨床phenotyping)とbiorepositoryに基づくGWAS等 (EMR-based GWAS) が可能か (LHS)。
 - 開始時はGWAS全盛時代。ゲノム医療の臨床実装未着手
- 電子カルテより臨床表現型情報抽出 phenotypingルール
- 計画開始時参加施設：Mayo, Vanderbilt Univ, Marshfield, Univ. Washington, Northwestern Univ.など5施設,

phase II (2011-2015) 臨床実装へ舵を切る

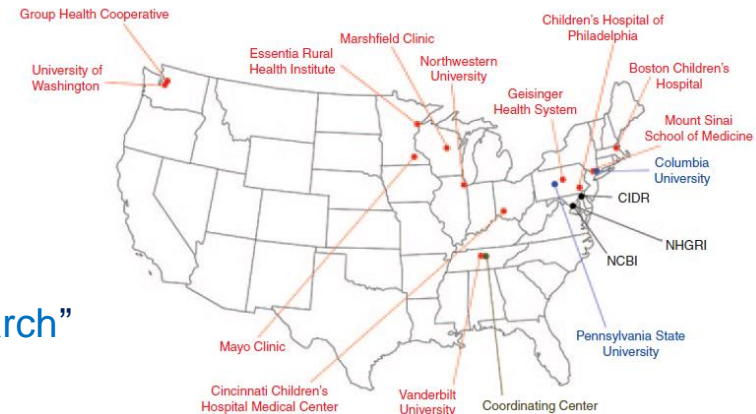
- MCWの臨床実装のインパクト, Vanderbiltの先制PG x
- 電子カルテと遺伝情報の統合
 - 電子カルテへのゲノム情報の統合
 - **PheKB** (Phenotype Knowledge Base)
 - ゲノム医療の実装、PGxの臨床応用
 - 結果回付 **Return of Result**, ELSI等
- 4つのサイトが新しく加わる
 - 小児病院グループとMount Sinai, Geisinger

phase III (2015) より始まる

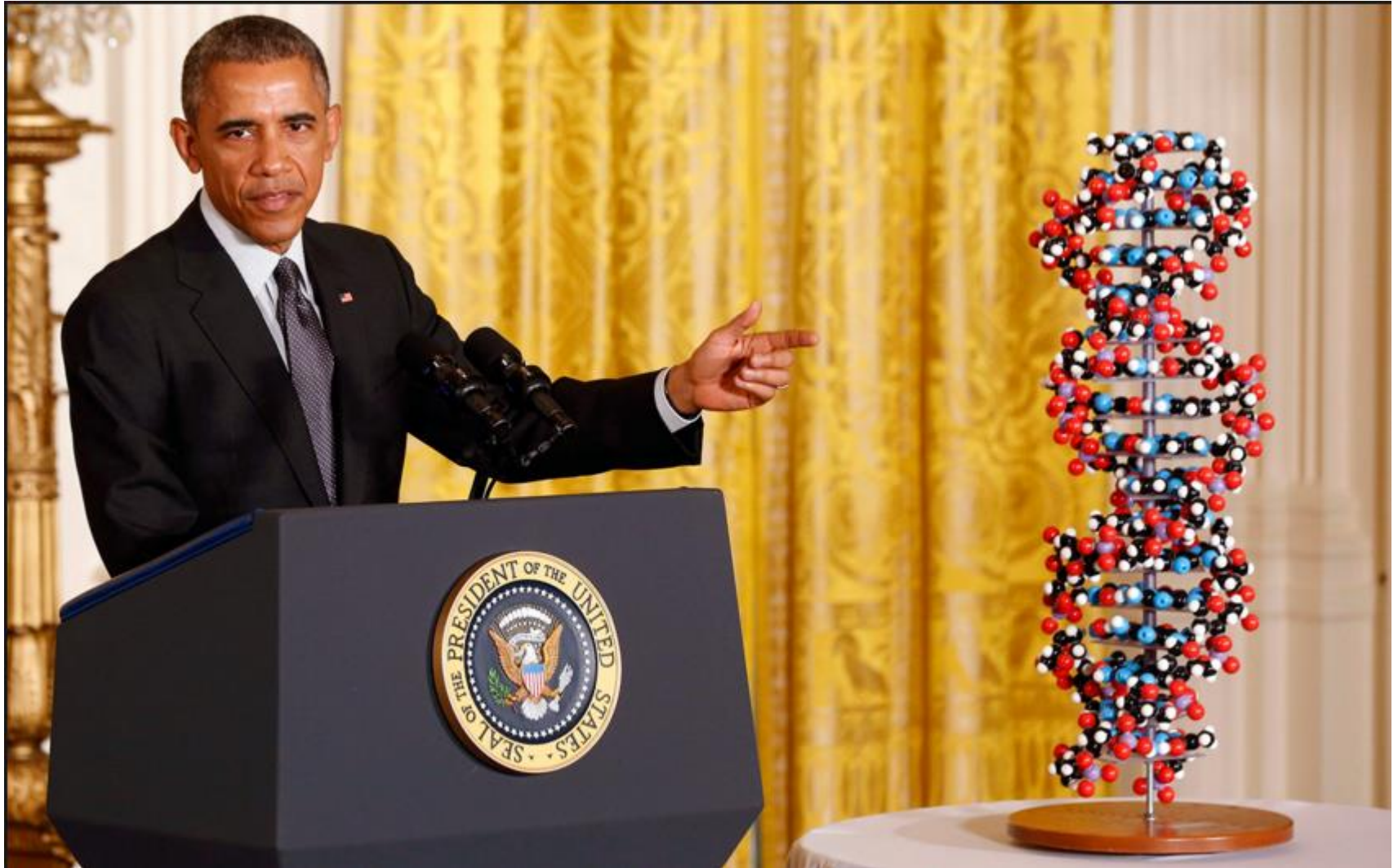
- NHGRIのコンソーシアムと連携
- とくに**CSER** “Clinical Sequencing Exploratory Research”



PheKB: phenotyping ルール



オバマ大統領 Precision Medicine Initiativeを開始



2015年1月 大統領一般年頭教書演説

Precision Medicine とは何か

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

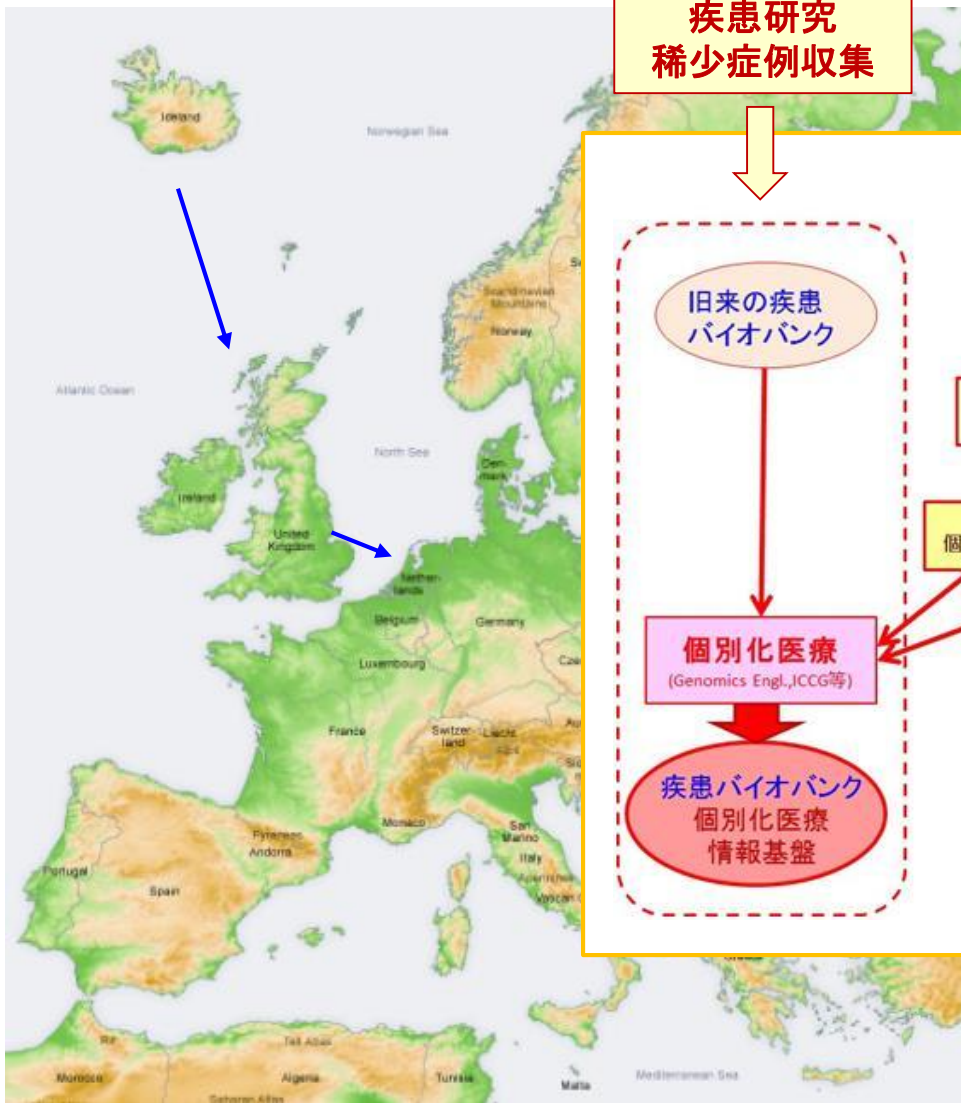
趣旨：基本は、個別化医療 Personalized Medicine の概念と変わらないが、目指していたのは診断/治療の個人化ではなく層別化であることを明確化

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

医療ビッグデータ時代の到来による個別化医療の拡張

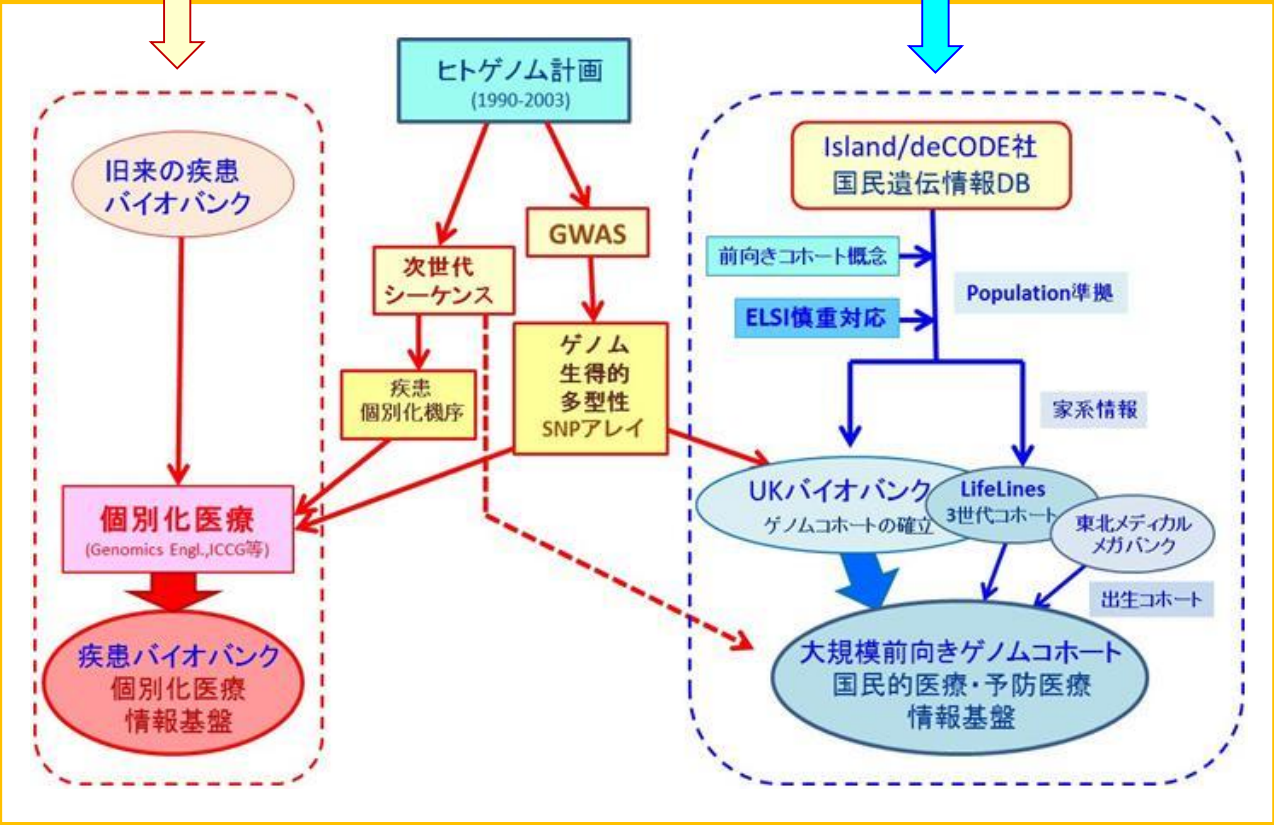
- (1) 遺伝素因 X 環境(生活習慣)要因のスキーマ重視
SNPや変異 (Genome)だけでなく環境・生活習慣要因(Exposome) の重視、疾患発症は2つの要因の相互作用を明快に強調。電子カルテの臨床表現型 (Clinical Phenome)も疾患発症後には不可欠。3つの成因の重視
- (2) 日常生理モニタリング情報の包摂
モバイルヘルス(mHealth)・wearable sensorによる大量継続情報収集の重視
- (3) ゲノムコホート・Biobankの重視
Precision Medicineを実現する基礎として、ゲノムコホート/Biobankが必要であることを認識。Real world dataの重視

第2の流れ 欧州のバイオバンクの普及



疾患研究
稀少症例収集

「集合的遺伝情報」による
国民レベルでの医療向上



UKバイオバンクとそのインパクト

- **＜集合的遺伝情報＞**の価値の認識
 - アイスランドdeCODE社
 - 「全国民の遺伝情報データベース」の概念
 - 「**集合的な遺伝子情報**」を用いて国民の医療の未来を拓く
 - deCODE社の提起した**＜「集合的遺伝子情報」を基盤として新たな医療を切り拓く＞**欧州型ゲノム医療は、UKバイオバンクによって引き継がれた。
- **前向きpopulation準拠ゲノム・コホート」と**
いう＜先見の明＞←1998年、英国政府は医学研究審議会（MRC）に「**国規模で使えるDNAコレクション**」
- 「**大規模large scale**」「**住民準拠population-based**」「**前向きコホート研究prospective cohort**」という、**基軸的な概念**がすでに全部取り入れられていた
- **標的は「ありふれた病気」（多因子疾患）の病因**に関する**遺伝的素因と環境要因の複雑な相互作用を解明**する
- 参加者の**その後の疾病発生を追跡するゲノム・コホート**
- **環境要因との相互作用を観測する**

Biobank/ゲノムコホートへの期待

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

Biobankとゲノムコホート

• バイオバンクの目的・機能の変化

- 従来は再生医療のための生体標本や臨床研究の資料保存、近年はゲノム医療の基盤としての役割
- **ゲノム/オミックス個別化医療、創薬の情報基盤**
 - **疾患型BioBank**：全国的・全世界規模で疾患罹患患者の網羅的分子情報（ゲノムなど）とそれに対応する臨床表現型（臨床検査、医用画像、処方歴、手術歴、病態経過、転帰など）の収集。**疾患ゲノムコホート**
- **個別化予防の情報基盤**
 - **Population型BioBank**：「健常者」前向きコホート。調査開始時の網羅的分子情報（ゲノム）と臨床環境情報（exposome）を集めて、生涯を追跡するゲノム・コホート

• 欧米のBiobank

- **英国 UK biobank**
 - 50万人の健常者。40～69歳（2006-2010, 62Mポンド), 2011-16, 25Mポンド
 - 健診データ（血液・尿・唾液サンプル、生活情報）を集め、健康医療状況を追跡する。
- **英国 Genomics England,**
 - 2013開始、2017年までに 10万人のゲノム 配列収集。
 - 最初の対象は稀少疾患（患者・家族）、がん患者、最初はEnglandのみ
- **欧州 BBMRI (Biobank/Biomole. Res. Infra.)**
 - 250以上の欧州各国のBioBankを統合
- **オランダ Lifeline**
 - 165000人北部オランダ 2006年開始 30年間の追跡、3世代コホート（世界初）
- **Precision Medicine Initiative Genome Cohort**
 - 100万人のゲノムを集める

最近の米国の状況

米国で国が推進する 各種プロジェクト

プロジェクト 略称	プロジェクト名	プロジェクトの内容	予算 (百万ドル)	期間
UDN	Undiagnosed Diseases Network	NIHの未診断疾患プログラム (UDP)を拡張して稀少・新規疾患を診断する	67.9	2013-17
NSIGHT	Newborn Sequencing In Genomic medicine and public Health	新生児のゲノム・シーケンスの利用可能性を探求する	10	2013-16
CSER	Clinical Sequencing Exploratory Research	ゲノム・シーケンスを臨床診療に応用するときのインフラ、方法、問題点を探求する	65	2012-16
eMERGE II	Electronic Medical Records and Genomics	ゲノム情報と電子カルテの融合。臨床フェノタイピング。	56	2011-15
RoR	Return of Results	ゲノム研究における個別遺伝情報を患者・家族に回付してよいか、時期・仕方に関する研究	5.7	2011-13
IGNITE	Implementing Genomics in Practice	患者のゲノム所見を臨床診療に組み込む方法の開発と普及	32.3	2013-16
ClinGen	Clinical Genome Resource	臨床診療に適切なバリエーションの共通情報の開発と普及	25	2013-16

米国ゲノム医学会（ACMG）の ワーキンググループ

- 「偶発的所見」（IF: incidental finding）
 - ROR（治療行為が対処可能（actionable）な56個の遺伝子変異のリスト）
 - 016年にはに二次的所見（the secondary findings）という名称で（注）、59 遺伝子（前リストから1種類を除き4種の遺伝子を追加）の新リスト

© American College of Medical Genetics and Genomics

Genetics
in Medicine

ACMG POLICY STATEMENT

ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing

Table 1 Conditions, genes, and variants recommended for return of incidental findings in clinical sequencing

Phenotype	MIM-disorder	PMID-Gene Reviews entry	Typical age of onset	Gene	MIM-gene	Inheritance*	Variants to report*
Hereditary breast and ovarian cancer	604370 612555	20301425	Adult	<i>BRCA1</i> <i>BRCA2</i>	113705 600185	AD	KP and EP
Li-Fraumeni syndrome	151623	20301488	Child/adult	<i>TP53</i>	191170	AD	KP and EP
Peutz-Jeghers syndrome	175200	20301443	Child/adult	<i>STK11</i>	602216	AD	KP and EP
Lynch syndrome	120435	20301390	Adult	<i>MLH1</i> <i>MSH2</i> <i>MSH6</i> <i>PMS2</i>	120436 609309 600678 600259	AD	KP and EP
Familial adenomatous polyposis	175100	20301519	Child/adult	<i>APC</i>	611731	AD	KP and EP
<i>MYH</i> -associated polyposis; adenomas, multiple colorectal, <i>FAP</i> type 2; colorectal adenomatous polyposis, autosomal recessive, with pilomatricomas	608456 132600	23035301	Adult	<i>MUTYH</i>	604933	AR [†]	KP and EP
Von Hippel-Lindau syndrome	193300	20301636	Child/adult	<i>VHL</i>	603200	AD	KP and EP
Multiple endocrine neoplasia type 1	131100	20301710	Adult	<i>menin</i>	603200	AD	KP and EP
Multiple endocrine neoplasia type 2	171400	20301710	Adult	<i>RET</i>	603200	AD	KP and EP

eMERGE III (2015年から2019年)

①専用のシーケンスセンターの設立

- eMERGE IIの時に、研究チームが9施設と1つの調整センターの体制に増加した状態は変わらないが、シーケンシングやジェノタイピングの拠点として、2つの施設、すなわちブリンガム婦人病院/ブロード研究所 (Brigham and Women's Hospital/Broad Institute) とベイラー医科大学 (Baylor College of Medicine) に集約された。

②25000人の次世代シーケンスによる遺伝情報回付110遺伝子の検証

- 第3期での大きなプロジェクトは、偶発的所見で、ACMGが遺伝子情報回付すべきとした56遺伝子リストに加え、eMERGEのメンバー9つの医療施設が各施設6遺伝子ずつ計54遺伝子を追加して出来た臨床有用性の高い110個の遺伝子の遺伝子リストについて、25000人のシーケンスを行い (eMERGEseq) の、その臨床的有用性を確かめるプロジェクトである。

③臨床表現型フェノタイピング・ルール作成の継続

- 臨床表現型の電子カルテからの抽出に関しては、II期の成果である、病名などを電子カルテ情報で確定するフェノタイピングのルールを引き続き策定し、60から80のルールに拡張する。

④遺伝子情報回付におけるELSI研究

- 遺伝情報の回付を患者に行う場合、患者と臨床医との関係性を含めたELSI研究を重点化する。

⑤CSERとの共同研究を通じた連携の強化

- 遺伝子情報回付に関してはCSERでもワシントン大学のNEXTmedicineプロジェクトでは、先に触れたように114個の遺伝子リストを提案している。またCSERコンソーシアムでは、遺伝子検査の回付に関して多くの研究グループが存在する。従って、CSERの各グループと連携することは、大きな成果をもたらす。

100万人コホート“ALL of Us”

- プレジジョン医療の最大の転換点は、米国が多因子疾患のゲノム医療を重視し始めたことである
- 2017年からの参加者のリクルート開始
- All of Us組織体制
 - ①主導施設： 2016年からパイロット研究が始まっており、PMIコホート研究の中心施設はバンダービルト大学医療センターであり、Data and Support Center（プロジェクト支援データセンター）として、昨年71.6百万ドル（約80億円）の予算がNIHから措置された。
 - ②共同施設： パイロット研究は、バンダービルト大学の主導のもとに、ブロード研究所、コロンビア大学医療センター、ミシガン大学公衆衛生学部、ノースウェスタン大学、テキサス大学生命情報学部である。Verilyライフサイエンス（かつてのGoogle Life Science）も参加する。
 - ③バイオバンク：メイヨクリニックに置く。そのほかに医療提供組織としてカリフォルニア・プレジジョン医療コンソーシアムや地域のプレジジョン医療コンソーシアムが参加している。

“National Cancer MoonShot”計画

National Cancer Moonshot計画が2016年から5年間の予算でNCI主導ので始まった。予算総額は、3年間で10億ドル（約1000億円）の超大型プロジェクト

①患者の直接の参加ネットワークの組織化

- 臨床治験にいつでも参加できるがん患者のネットワークの組織化。

②がんの免疫療法の重点的推進

- がんの免疫療法は一方では目覚ましい成果を上げているにもかかわらず、一部の患者だけである。免疫療法の効果の機序の研究を推進する。

③がんの耐性の超克

分子標的薬剤も当初は著しい効果を上げて耐性によって効かなくなる。体制のメカニズムを解明し、それを防ぐ

④がんデータ共有の国規模エコシステムの形成

がん患者の分子情報や臨床病態情報を広くデータ共有するエコシステムを構築する。

⑤小児がんのドライバー研究の強化

小児のがんの主要なドライバー分子（融合がんタンパク質など）に対する研究の強化

⑥がん治療の衰弱化副作用の最小化

化学療法などで患者を衰弱させる副作用をおこす。これを最小限にする。

⑦実証された予防医学の利用の拡張と早期のがん検出戦略

有効だと証明された予防法の拡張と早期がん検出プログラムの重視

⑧過去のデータ蓄積から知識発見して将来の患者病態を予測する

⑨3次元がんアトラスの構築

がん細胞と微小環境との相互作用を捉える

⑩新しいがん技術の開発

この暫定的な方針を見ても、超大型がんプロジェクトNational Cancer Moonshotの今後の展開を期待したい。

ビッグデータ医療における 人工知能の期待



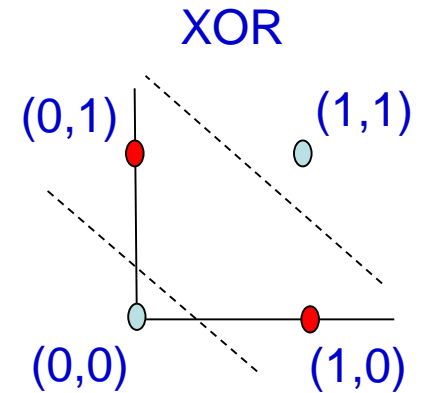
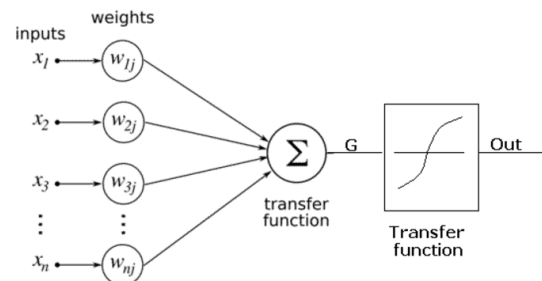
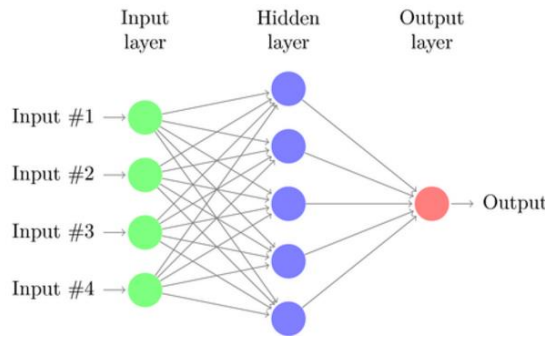
人工知能の最近の話題

- 「アルファ碁」 (Google DeepMindによるコンピュータ囲碁プログラム) が2016年3月に数多くの世界戦優勝経験のあるプロ棋士李世石 (Lee Sedol : 九段) に挑戦し、**4勝1敗と勝ち越した**
 - チェス : IBM 「Deep Blue」 が1997年に当時の世界champion, カスパロフ氏 (ロシア) に勝利
 - 将棋 : ボンクラーズ, 2012年米長永世棋聖に勝利
 - 「アルファ碁」にはニューラルネットワーク (Deep Learning) が使われた。評価経験則が人間によってコードされていない
 - 最初、棋譜に記録された熟練した棋士の手と合致する手をさすように訓練され、次に、ある程度の能力に達すると、強化学習を用いて**自身と多数の対戦 (3000万回) を行う**ことで上達した。
- 人工知能が1000万枚の画像を与えて「猫」を認識するニューロンをできたと2012年に発表



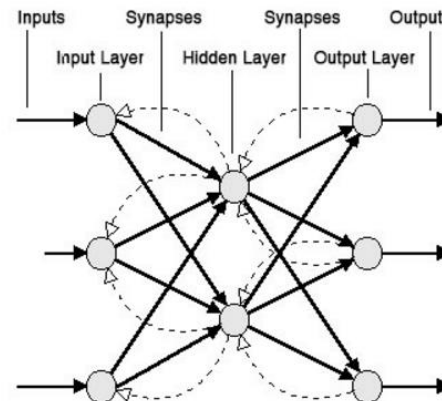
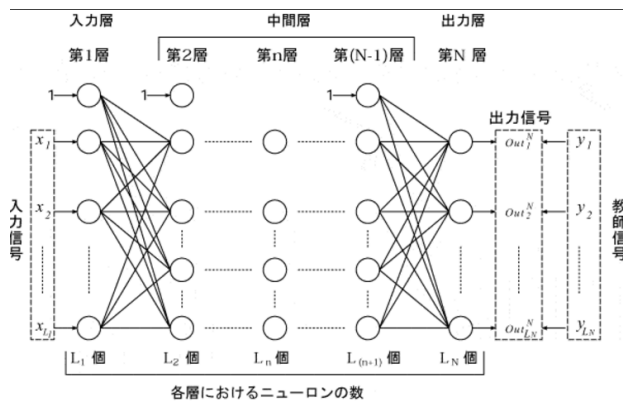
従来のニューロネットワーク

古典的Neural Network・パーセプトロン(1970年代)



多層Neural NetworkとBack projection (1980年代)

線形分離できない

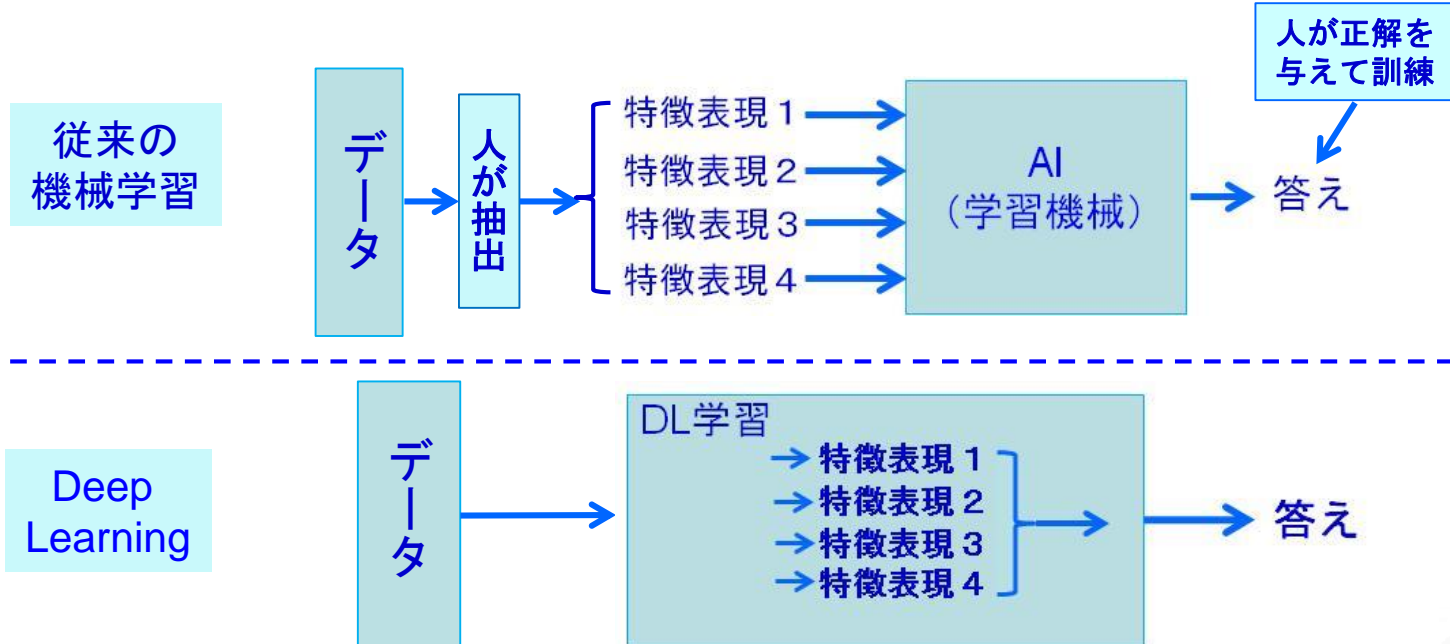


Back Propagation (1986 Rumelhart)
 望ましい出力との誤差を教師信号として与える事により、次第に結合係数を変化させ、最終的に正しい出力が得られるようにする。結合係数を変える事を学習と呼ぶ。この学習方法には、最急降下法(勾配法)が使われる。出力層へ寄与の高いノードの重みの変更。

多層にわたる逆伝搬で修正感度減衰

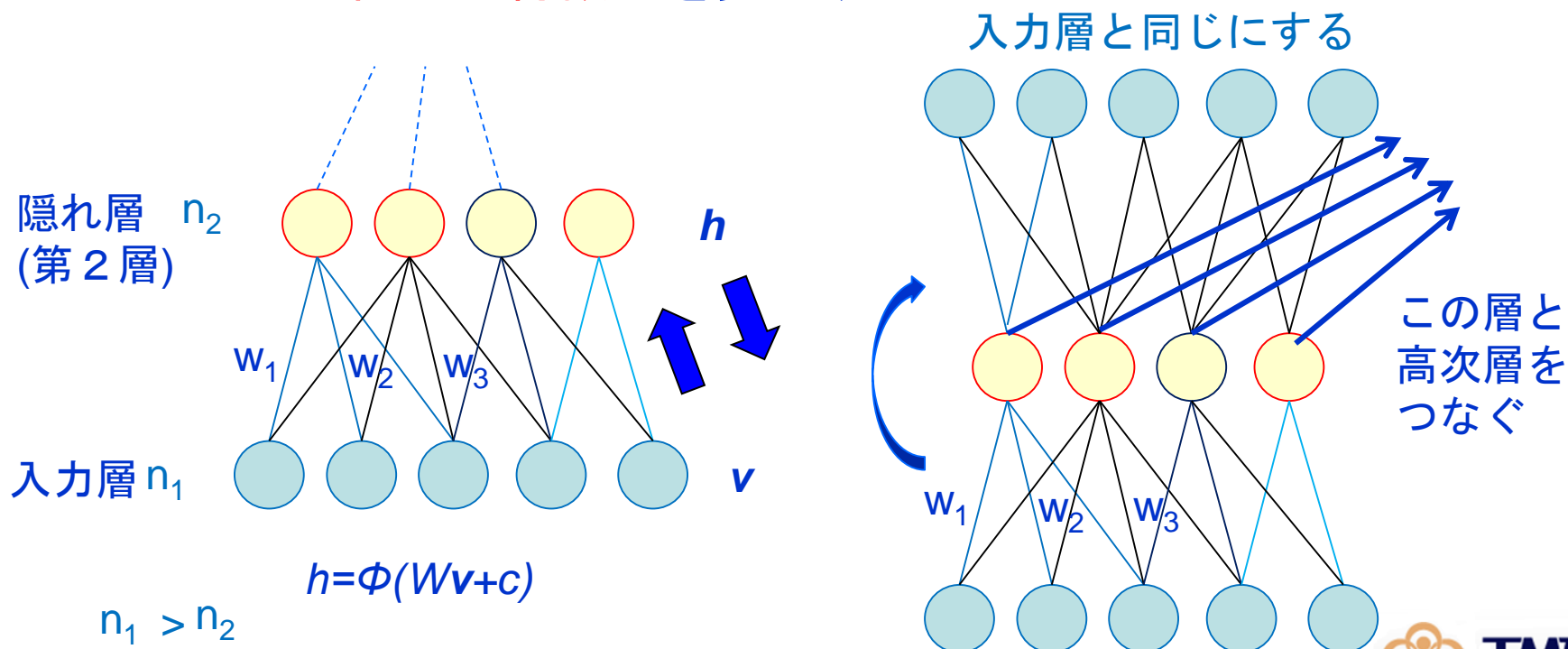
Deep Learning による 人工知能革命

- 機械学習のこれまでの限界
 - 「教師あり学習」
 - 分類対象の特徴と正解を与え学習機械（AI）を構築
- Deep Learningの革命性
 - 「教師なし学習」
 - 対象の特徴表現や対象の高次特徴量を自ら学ぶ



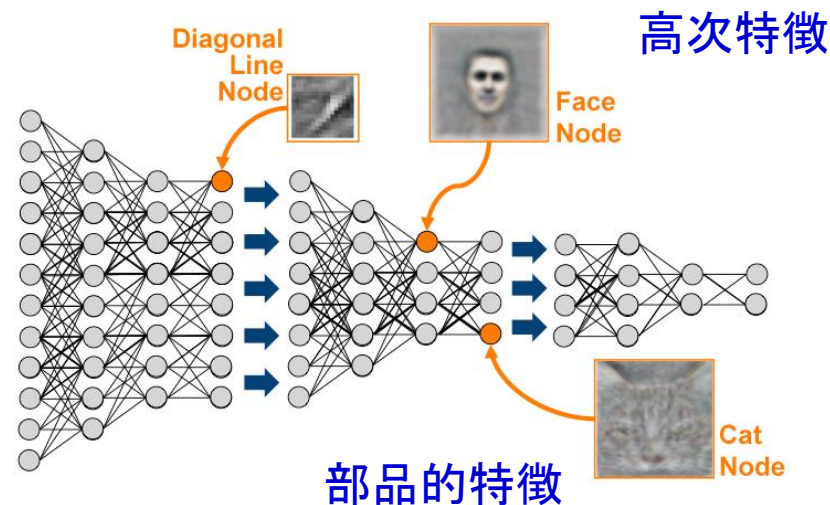
DLの革命点 Autoencode 1

- 対象に固有な**内在的特徴**を学ぶ自己符号化の原理
- 格段ごとに入力を少ない中間層を介して復元できるかを行なう
- 次元を圧縮されて可及的に復元する
 - できるだけ復元に**効果的な**特徴量を探索する
 - 内在的な特徴量**を見出す



DLの革命点 Autoencode 2

- 各層ごとに自己符号化を行うので**何層でも組める**
 - 各層間で「自己符号化」の積上げ (autoencoder stack)
- 第一層で学習した特徴量を使って次の階層を作るので**高次の特徴量**が作られる
- 特徴的表現と概念を結びつけるため「**教師あり学習**」が最後に必要。
- 自動特徴抽出によってこれまでの学習手法の限界を克服した
 - 内在的な特徴量による構造的な理解
- 人間の「思考の枠組み」を超えた正解の低次
 - 「**アルファGo**」が定石にない手で碁の名人に勝つ



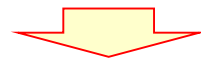
「ビッグデータ」のData 縮約原理

問題点 属性項目数(p) \gg サンプル数(n)

p : 数億になる場合あり n : 多くても数万、通常数千



これら膨大な属性変数がすべて独立ならばビッグデータの構造解析は不可能。単変量解析の羅列 (GWASのManhattan Plot) しか可能でない



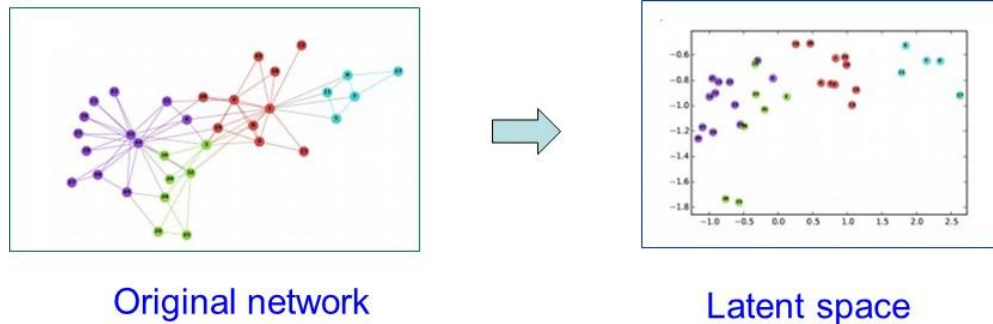
ビッグデータ・スパース仮説

ビッグデータは、多数であるが属性値数より少ない独立成分が基底となって、相互にModificationして構成されている。
(独立成分の推定は、サンプル数とともに増加する)

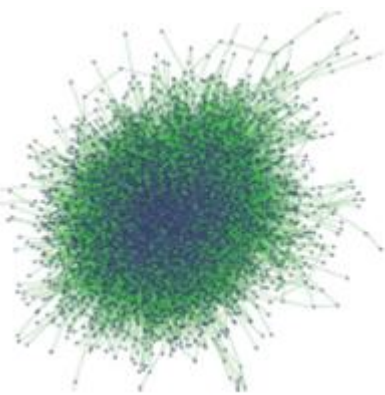
データ次元縮約の原理 (**principle of compositionality**)

Deep Learningによる Network Embedding

超多次元ネットワークをそれより遥かに低次元のLatent Spaceに写像



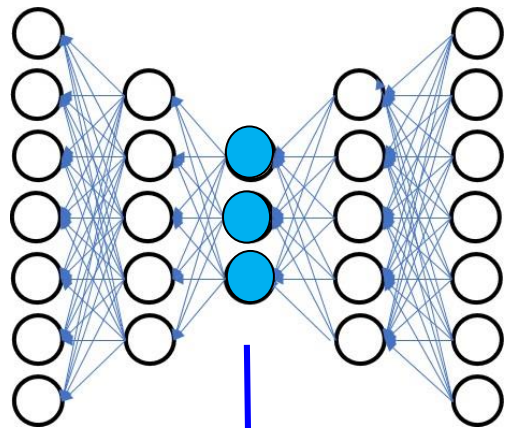
Structural Deep Network Embedding



大規模
ネットワーク

全節点の近接ベクトル

入
力
層

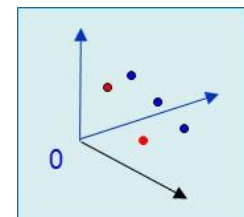


encoder

decoder

対
称
出
力
層

潜在空間
Latent space



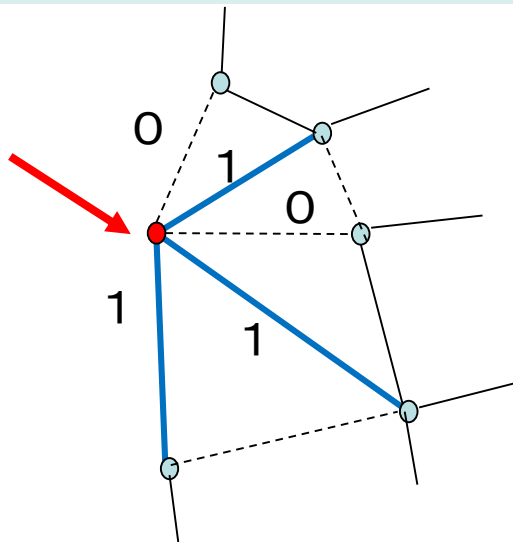
Deep Learning と SVD (singular value decomposition)の精度の違い

あるタンパク質相互作用ネットワークのノードに注目する

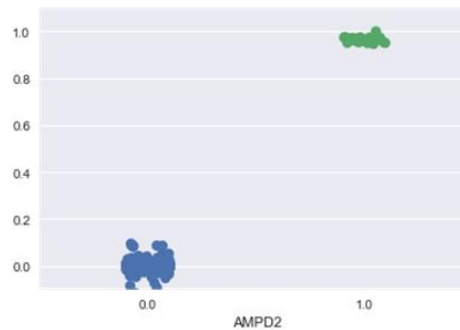
周りのノードで

結合しているノードは 1
 結合していないノードは 0
 とすると0, 1の近接ベクトルで結合を表現できる。

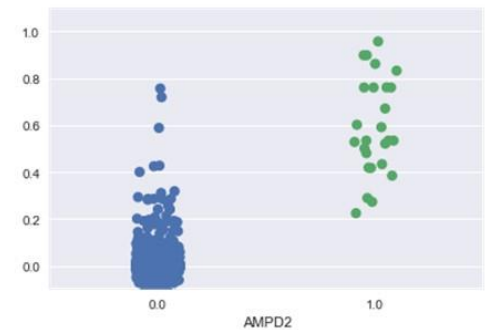
$$v_i = (0, 0, 0, 1, 0, 1, 0, \dots)$$



AMPD2 (adenosine monophosphate deaminase 2)
degree=26

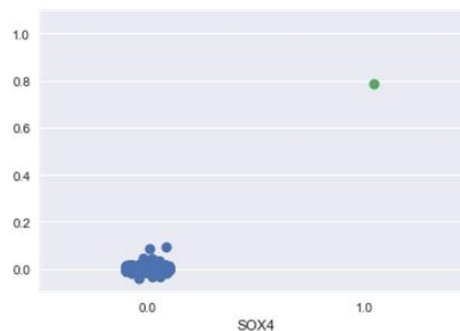


Autoencoder

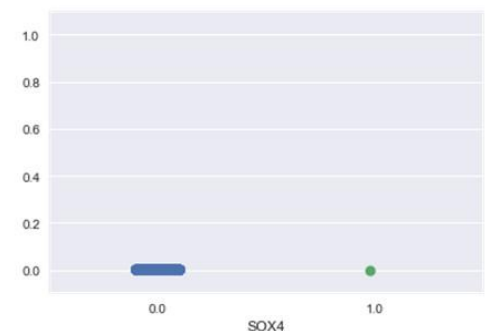


SVD

SOX4 (SRY-box 4)
degree=1



Autoencoder



SVD

N=8,502

Deep Learningによる 多次元ネットワーク縮約法

(Hase, Tanaka 2017)

- 医療・創薬ビッグデータへの応用性高い
- 超多次元ネットワーク情報構造の急増
 - ゲノム医療<網羅的分子情報–臨床表現型情報>
 - ゲノムコホートにおける<遺伝子情報–環境（生活様式）情報>
- Deep Learning-based Network Contraction
「DLネットワーク縮約法」

超多次元ネットワーク情報構造⇒
少数の特徴的ネットワーク基底に分解
- 線形分解ではない。非線形分解で基底への射影

第2世代の ゲノム・オミックス医療



ゲノム医療の第2世代

成功した臨床実装

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

しかし

多因子疾患の機序/発症予測は無着手である

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

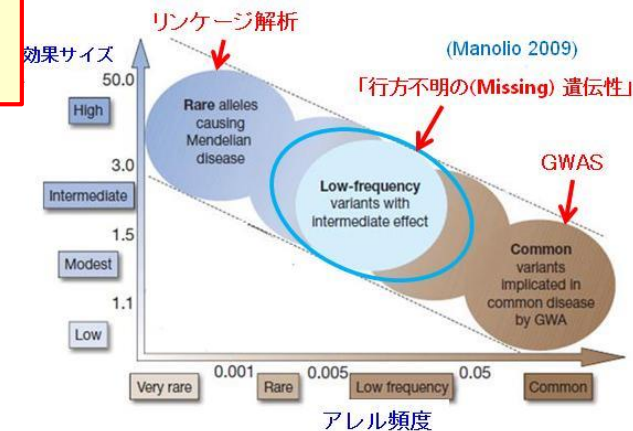
SNPの相対リスク
低い(1.1~1.3)理由
 $G \times E$ 組合せ特異的効果
を環境要因の平均



多因子疾患は個人の<遺伝的体質と環境要因>の
<相互作用の結果。シーケンスだけでは解明不能

疾患発症の遺伝要因と環境要因の相互作用は
加算的 ($G \oplus E$) でもなく乗算的 ($G \otimes E$) でもない
< (G, E) 組合せ特異的な効果 > である

例 大腸がんの遺伝要因と環境 (生活習慣) 要因



大半の疾患の基礎としての 「遺伝素因X環境要因」の相互作用

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

疾患発症の相対リスク=

遺伝要因(G:genome) X 環境要因(E:exposome)

相互作用は加算的でもなく乗算的でもない

<(G,E) 組合せ特異的な効果>である

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



発達プログラム説 DOHaD

(Developmental Origin of Health and Disease)

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



オランダ
飢饉 (1944)

環境因子

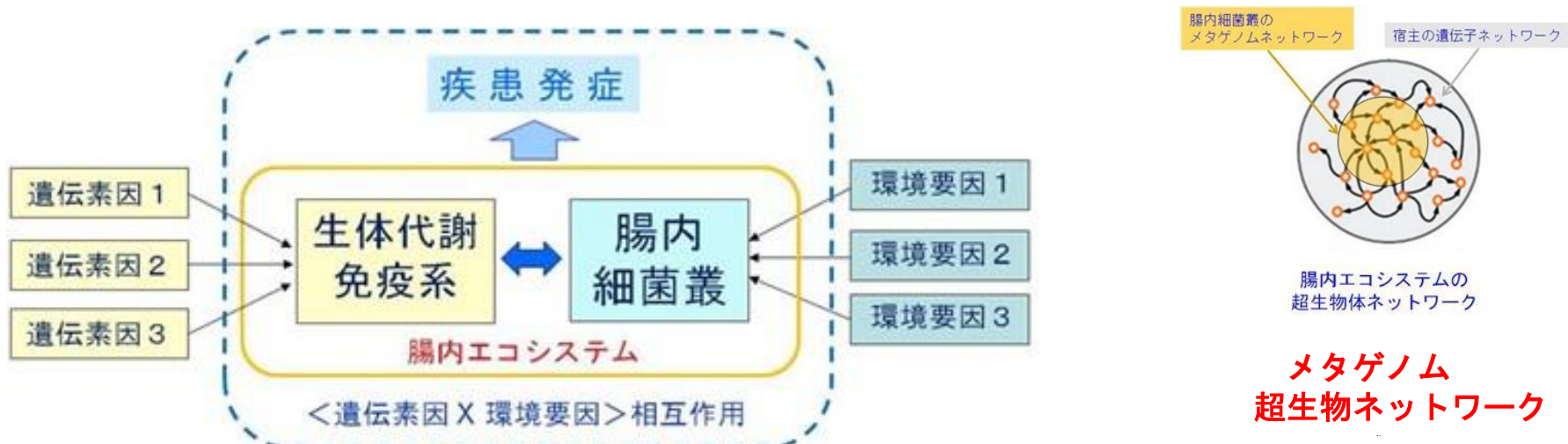
Epigenome変化

遺伝子発現調節

疾病発症

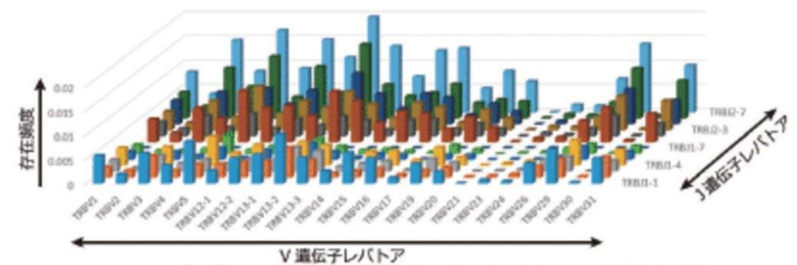
腸内細菌叢microbiome：メタゲノム

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



免疫ゲノム

- 可変領域や相補性決定領域（特にCDR3）のDNAやRNAを次世代シーケンサ(HTS)で解析
- レパトア解析
 - 抗原受容体全体のプロファイルを俯瞰的に把握できる
 - V(D)Jなどの成分を基軸として3次元表示可能。
 - 疾患罹患とともに瞬時に全体像が変化する。
 - 網羅的病態全体像を提示する
 - VDJの使用頻度
 - 多様性(diversity)の変化
 - 疾病/加齢レパトア分布変化
- 臨床シーケンスに含まれる
- 3次元分布の特徴分析



(レパトア・ジェネシス社)

第2世代のゲノム・オミックス医療

- 生涯的全体性においてその個人の疾患可能性の全体性を把握し、個別化予防、個別化治療に取り組む
- ゲノム・オミックス情報と医療・健康
 - **Clinical Sequencing**のインパクト
- **第1世代ゲノム医療**
 - ゲノムの変異・多型性の個別性に基づく
- **第2世代のゲノム医療**
 - 多因子疾患が対象、環境情報との相互作用
 - エピゲノム、メタゲノム・免疫ゲノムなど
 - 遺伝子X環境の相互作用を反映する**メタ・オミックスのバイオマーカ**が必要

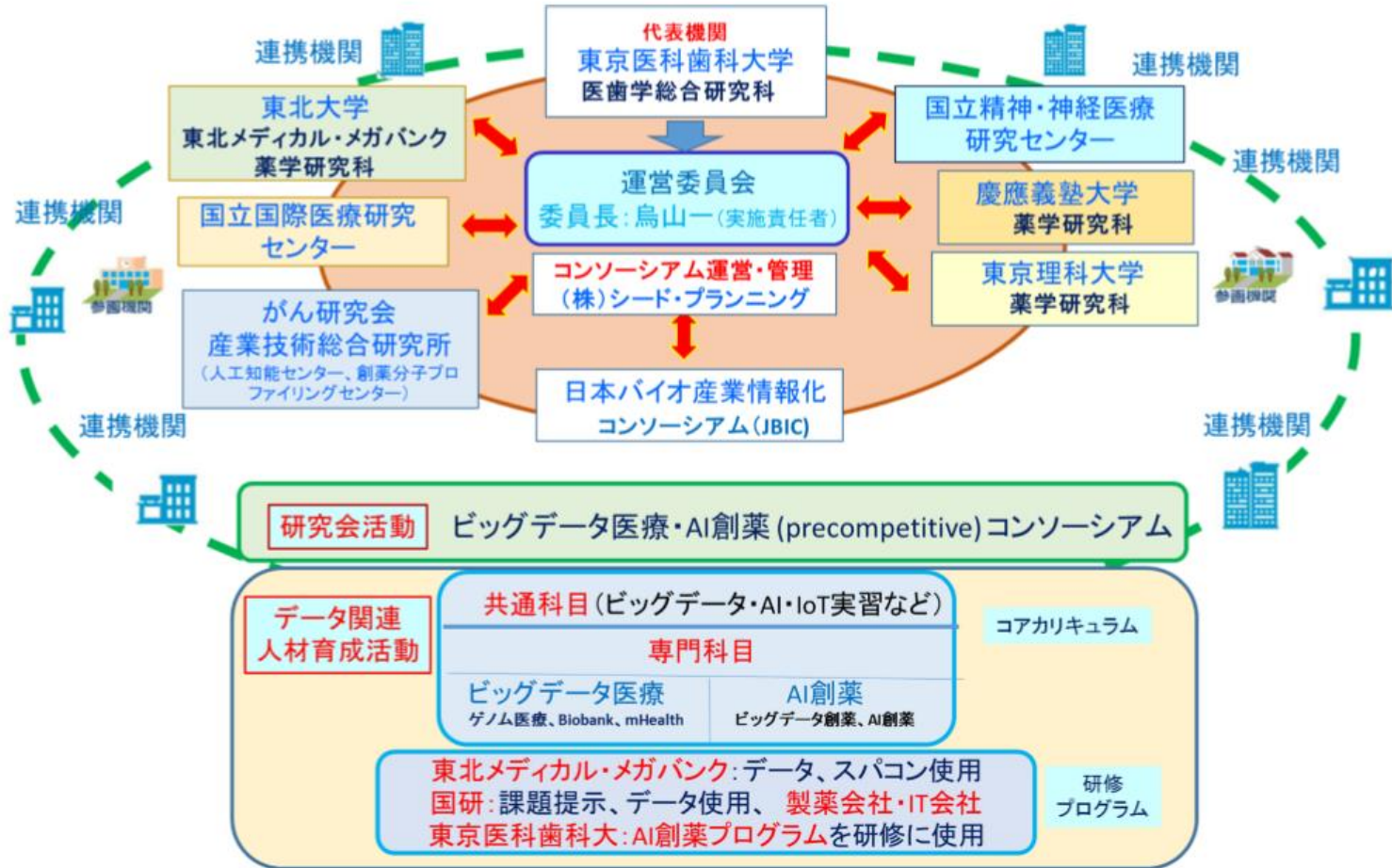
疾患メタ・オミックス修飾

今後の戦略・方向

- ビッグデータ医療時代：次元縮約
- Deep Learningによる〈多次元ネットワーク情報構造〉の縮約
 - ビッグデータ医療への適応可能
 - ゲノム医療の〈網羅的分子情報－臨床表現型〉の相関ネットワーク構造
 - バイオバンクの〈遺伝素因－環境要因〉と発症
- AI医療の「枠組み」実行方向は「見えてきた」

ヒトの仮説駆動的な〈知〉とAIのデータ駆動的な〈知〉との
「共創的cocreatingな〈知〉」が
これからの人類の未来の進むべき途の探索を可能にする

ビッグデータ医療・AI創薬コンソーシアム



2018.2.23-25, Harvard/MIT/TMDU - Datathon

ご清聴有難うございます



その他の米国での動き

「学習する医療システム」 Learning Health System

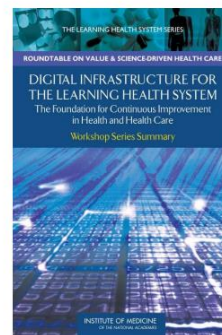
新しい生物医学知識が臨床実践に給されるまで17年
臨床データを用いて医療を実施しながら医療を改善

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

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

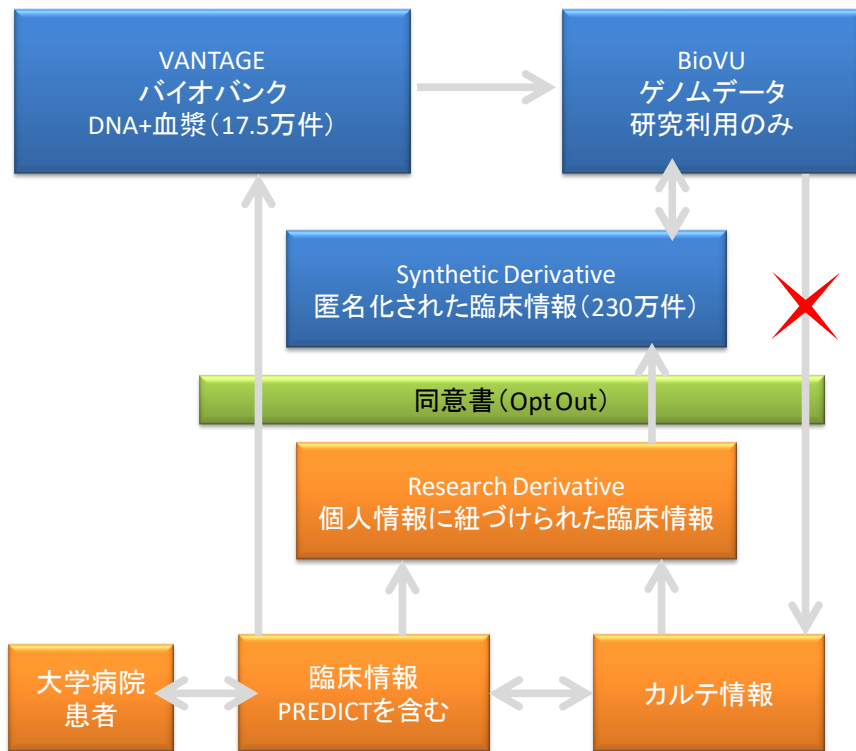
Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care

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



LHSの代表例 BioVU

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



電子カルテ

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

バイオバンクと遺伝子解析

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

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

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

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

個別化（層別化）医療の概念の普及とRCTの限界

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



ビッグデータと機械学習

- **The ASCO (米国臨床癌学) CancerLinQ initiative**

- 診療の現場(EHR)から大量の診療データを集め分析
- 新しい臨床治験へのガイドライン作成
- 17万人のがん症例データベースを構築。各がんについて1～2万人の症例を集める
- 学習システムを構築し治療知識を統計学習、ニューロネットを駆使して学習。

BigDataにおけるLearning systemの不可欠性

- 2013年に、CancerLinQのプロトタイプを完成、10万人以上の乳がんを蓄積、完全規模へ継続構築中

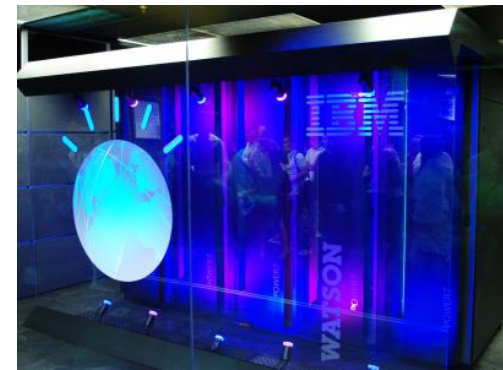
- **IBM Watsonのがんセンターへの普及**

- Memorial Sloan-Kettering Cancer Center (MSKCC) などと共同
- Watsonを母体に**The Oncology Expert Adviser software (OEA)**開発
- 他にNew York Genome Centerとglioblastoma (グリア芽細胞腫) 知識生成

- **Cancer Commons initiative**

- Rapid learningのインフラ整備
- 目的：患者の個別症例と最新の知識を更新
- 個々の患者の”Donate Your Data”(DYD)登録

- Google X project, “Human Longevity Inc.”



IBM Watson
Learning Big
Data

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