ビックデータとAI（人工知能）による創薬

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

田中　博

疾患ゲノム・オミックスの「ビッグデータ」と創薬

TMDU

医療ビッグデータ時代の到来

(1) 次世代シーケンサなどによる「ゲノム/オミックス医療」による網羅的分子情報蓄積

(2)モバイルヘルス(mHealth）によるWearable センサ情報の継続的蓄積（unobstructed monitoring)

(3) Biobankによるゲノム・コホート情報

大量データの急激な
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・予測医療
健康・医療の適確性の飛躍的な増大

医療の「ビッグデータ革命」
～何が新しいのか～

1）臨床診療情報
- 従来型の医療情報
  - 臨床検査、医用画像、処方、レセプトなど

2）社会医学情報
- 従来型の社会医学情報
  - 疫学情報・集団単位での疾患罹患情報

旧来のタイプの
医療データの
大容量化

3）新しい種類の医療ビッグデータ
- **網羅的分子情報・個別化医療**
  - **ゲノム・オミックス医療**
  - システム分子医学・Precision Medicine
- **生涯型モバイル健康管理**（mHealth）
  - ウェアラブル・生体センシング

新しいタイプの
医療ビッグデータ

TMDU

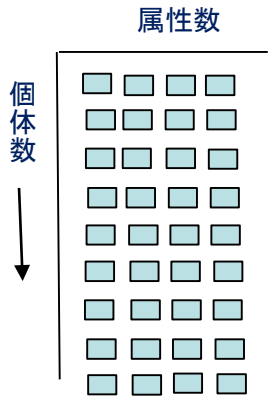
医療の「ビッグデータ革命」

～ゲノム・オミックスデータの基軸的な特徴～

＜目的もデータ特性も従来型と違う＞

従来の医療情報の「ビッグデータ」

Big “Small Data”　(n≫p)



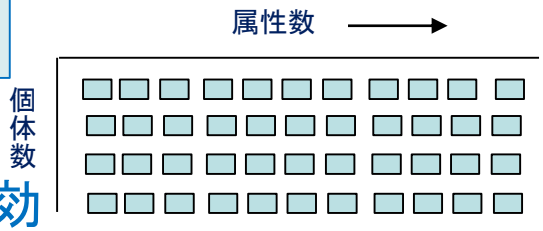
医療情報・疫学調査では 属性数：10項目程度

– 目的：Population MedicineのBig Data

⇒個別を集めて「集合的法則」を見る

網羅的分子情報などのビッグデータ

Small “Big Data”(p≫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」の多様性を摘出

新しいデータ科学の必要性

TMDU

医療の「ビッグデータ」革命は どんな既存のパラダイムに挑戦しているか

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

TMDU

医療ビッグデータの時代の到来と米国の最新の状況

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

次世代シーケンサのインパクト

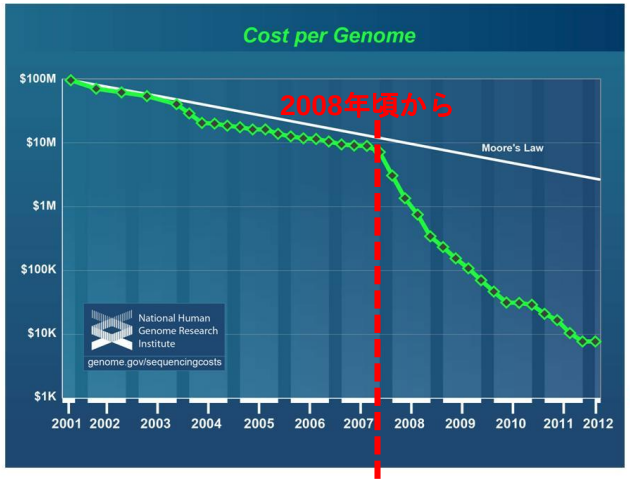
次世代シーケンサを始めとするhigh-throughput分子情報収集の急激な発展

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

2005～ NGS 454 (LS,Roche)
2007/8～454, Solexa (Ilumina), 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！

TMDU

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発展を受けて
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

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

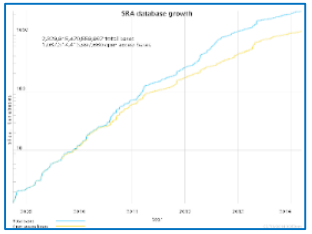
次世代シーケンシングの臨床普及(2010～)

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

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

TMDU

ゲノム・オミックス医療の３つの流れ

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

2008年
2009年
2010年
2011年
2012年
2013年

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

TMDU

ゲノム/オミックス医療ー米国の状況

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

- 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

NHGRI Working Groupのリスト

Institution	Major Projects
MC Wisconsin	Using whole genome sequencing to establish diagnosis in patients with currently undiagnosed genetic disorders
Mount Sinai	• 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	• 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	• 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 • SLCO1B1*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	• 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	• 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	• PGx driven selection/dosing of antidepressants • CYP2C19 genotyping for antiplatelet rx post PCI
Inter-Mountain	Tumor-based screening for Lynch syndrome

TMDU

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

第１世代

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

Early adopter 時期

Baylor**医科大学**
Mayo Clinicなど
後続病院多数

2013 前後

第２世代

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

国家政策/全国Consortium 時期

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

2015

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

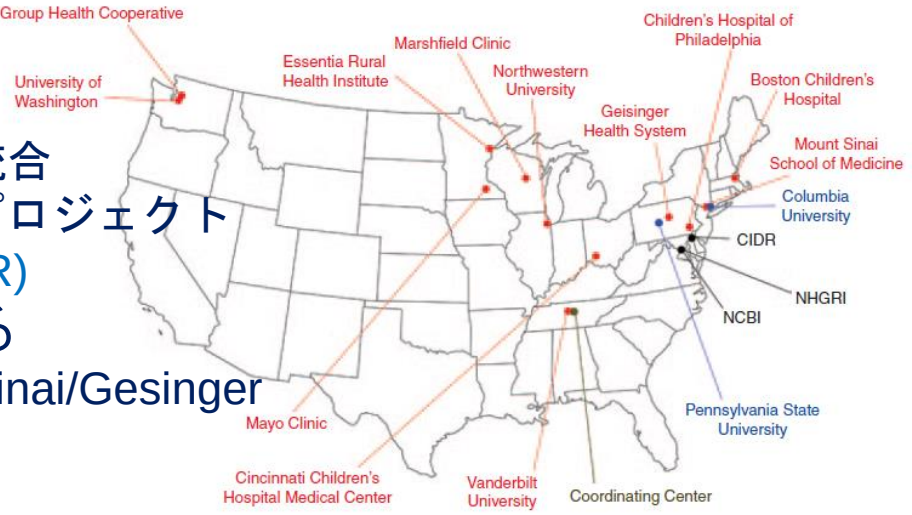
NIH “BD2K ”COE in Data Science, DDI (2014)
ASCO “CancerLinQ”, Cancer Common
“Precision Medicine (Obama) ” 1 M genomic cohort

臨床表現型 *eMERGE*プロジェクト

electronic MEdical Record + GEnome (NIH grand)

電子カルテからphenotyping

- phase I (2007-2011) 臨床表現型情報のタイピング
  - 電子カルテを通して臨床phenotypingするときの形式
  - EMR：臨床phenotypingとbiorepositoryに基づくGWASが可能か（EMR-based GWAS)。ELSI側面も検討
  - eMERGE-I: Mayo Clinic, Vanderbilt大学, Northwestern大学など５施設
- phase II (2011-2015) 臨床実装
  - 電子カルテと遺伝情報の統合
    - 電子カルテへのゲノム情報の統合
    - PGxの臨床応用に関する試行プロジェクト
    - 結果回付 Return of Result (RoR)
  - 4施設がeMERGE-IIより加わる
    - いくつかの小児病院とMount Sinai/Gesinger
- phase III：2015より始まる
- **CSER consortium**と連携
  - “Clinical Sequencing Exploratory Research” コンソーシアム NHGRIにより予算化



国家戦略としての「医療ビッグデータ」

NIH「ビッグデータから知識へ」計画

“Big Data to Knowledge” (BD2K)initiative

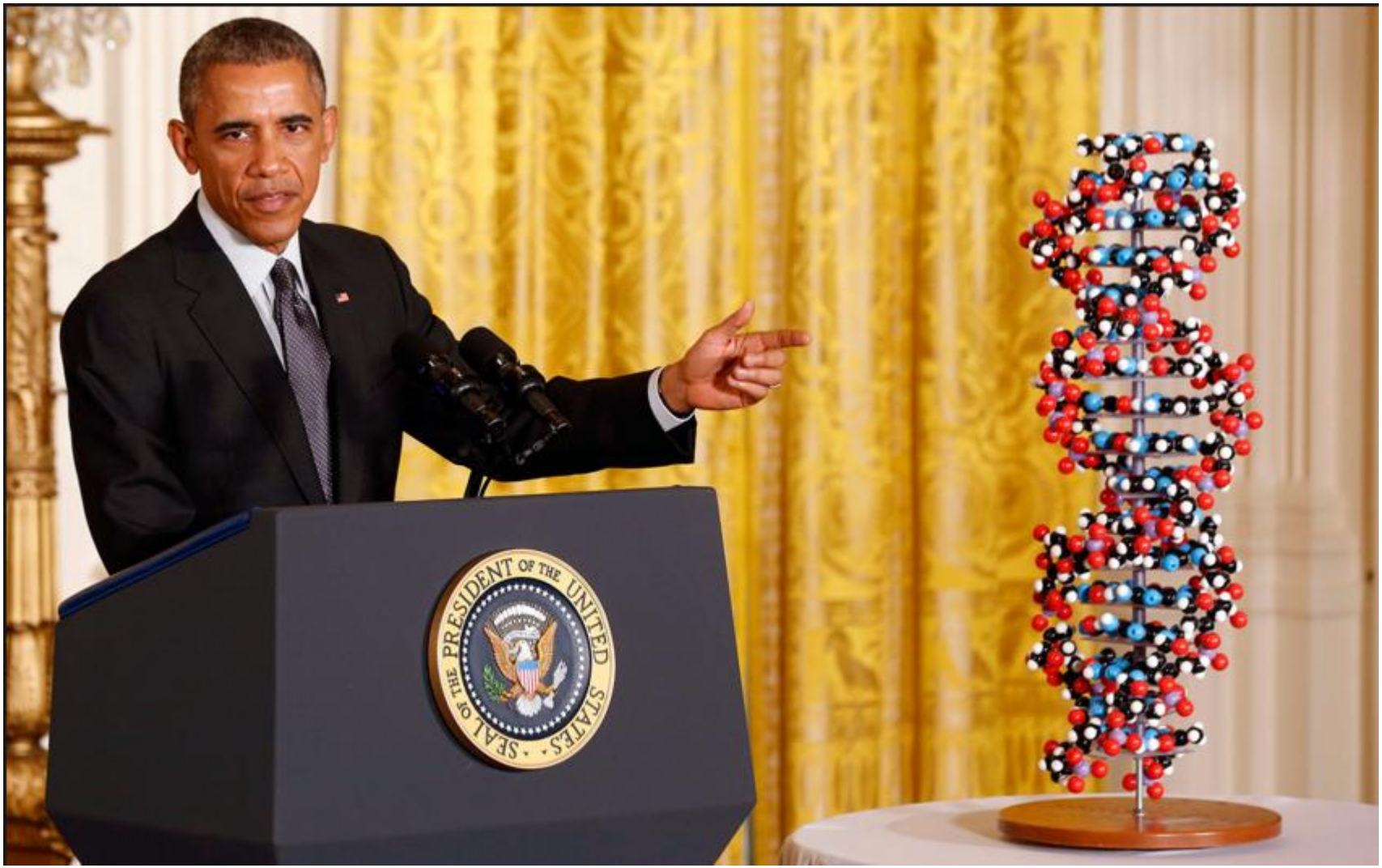
- BD2K: ”Big Data to Knowledge” Initiative 開始
  - 次世代シーケンサによるゲノム・オミックス医療の普及
  - 臨床シーケンス情報蓄積の大量化蓄積に対応して政策立案
  - 研究費の配分2013年に提案。計画実施は2014年から
  - **データ科学のための副長官**（Associate Director of Data Sciences）を医療情報の世界から任命　**Bourne, PhD.**
- Francis Collins長官談「NIH全規模での優先計画」
  - 生命医療研究に喫緊の重要性を持つ、指数的に増大する生命医療データを活用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

国家戦略としての「医療ビッグデータ」

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

TMDU

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

TMDU

Precision Medicineとは何か

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

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

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

医療ビッグデータ時代の到来による個別化医療の拡張

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

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

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

F.Collins

TMDU

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

TMDU

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 x 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

わが国でのゲノムオミックス医療の臨床実装

研究費を用いた試行的ゲノム医療であるが、いくつかの医療施設で
ゲノム・オミックス医療が試行されている

「ゲノム医学実現推進協議会」(中間報告）2015.7

「全国遺伝子医療部門連絡会議(10.18）」NGS臨床応用セッションに
会員の中で「臨床応用を実施している部門は12施設」
アンケート結果が発表。東大病院ゲノム医学センターなど
25～40%程度の原因遺伝子同定

AMED：IRUD（Initiative on Rare and Undiagnosed Disease)
未診断疾患の原因遺伝子をIRUD拠点病院が審査して解析センターがシーケンシング。その後、DB化する。米国UDP,
英国DDD(Deciphering Developmental Disorders),カナダForge (Finding of Rare Disease Genes)

がんの網羅的分子診断と個別化治療

- 国立がん研究センター東病院
  - ドライバー遺伝子の診断。分子標的薬の治験グループに割当て
- 静岡県立がんセンター 上記と同様の内容のプロジェクト
- 京大腫瘍内科（OncoPrime)、岡大、北大、千葉大 診療施設併設型BB

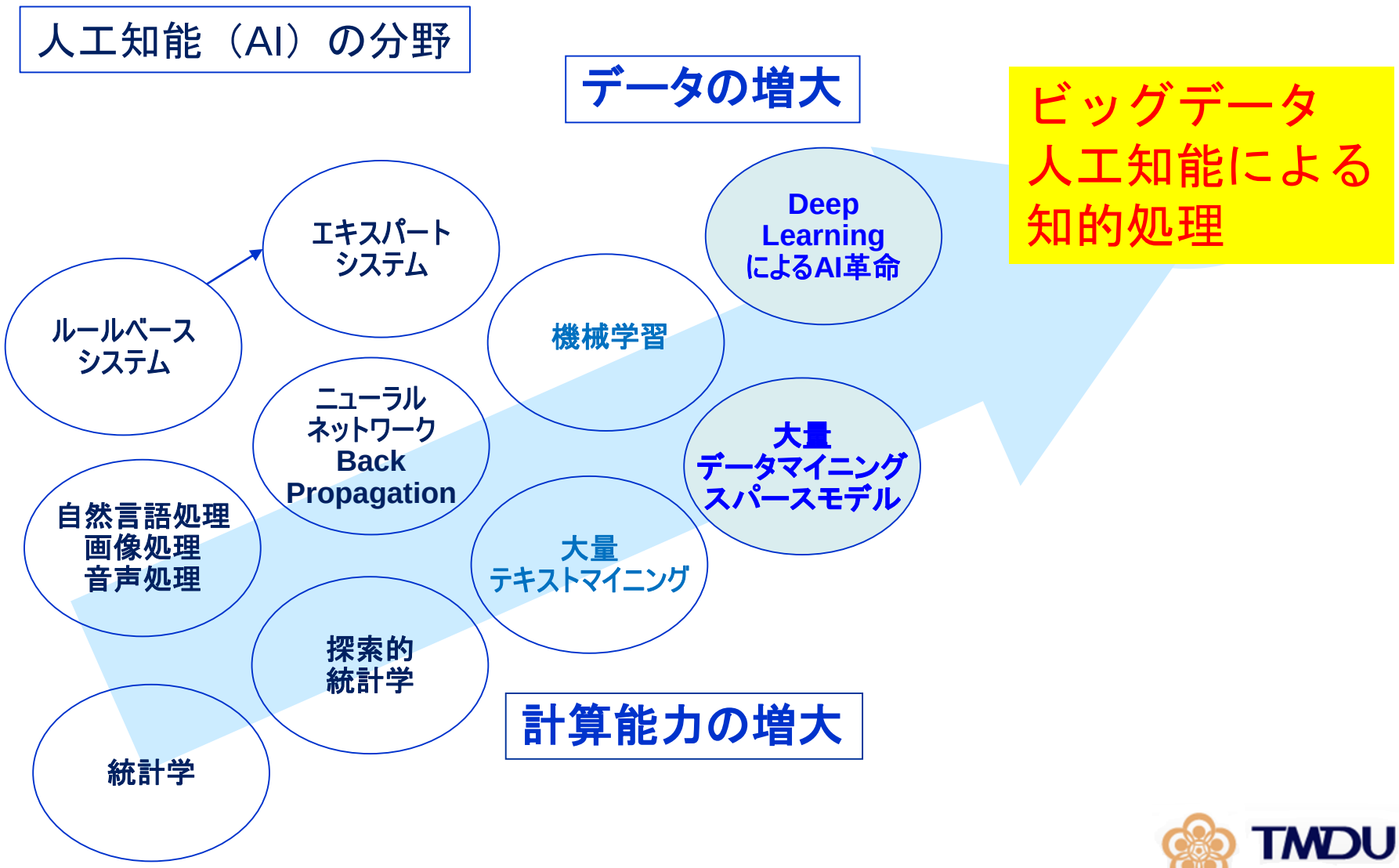
ゲノム医療では、米国と水を空けられている。しかし、Biobank Genomic Cohortでは我が国の状況はそれほど遅れていはいない。Biobank準拠のゲノム医療/創薬推進を行うべきである。また日本版eMerge計画を推進して臨床表現型情報の蓄積に邁進すべきである

AI医療・AI創薬

―人工知能への期待―

TMDU

人工知能への期待


人工知能（AI）の分野
データの増大
ビッグデータ
人工知能による
知的処理
ルールベース
システム
エキスパート
システム
ニューラル
ネットワーク
Back
Propagation
自然言語処理
画像処理
音声処理
探索的
統計学
統計学
機械学習
大量
テキストマイニング
Deep
Learning
によるAI革命
大量
データマイニング
スパースモデル
計算能力の増大

医療分野の人工知能の歴史

記号（シンボル）的知識処理

ニューロネットワーク処理

1970

問題解決の一般探索手法 GPS
解決木の高速探索（ゲーム）

ニューロネットワーク
3層の学習機械 Perceptron
入力層、隠れ層、出力層

1980

推論システム（if-thenルールシステム）
知識の表現と利用（専門家システム）
医療診断システム（Mycin, Internist-I)
大ブーム　医療から産業応用の期待波及

多層型ニューロネット
後方伝播 Back Propagation
結合係数修正アルゴリズム

期待消滅！

1990

しばらく停滞！

知識発見 機械学習
Machine Learning, KDD
診断知識のDBからの学習

ニューロネットワーク型
多層型ニューロネット
深層学習 Deep Learning
結合係数修正アルゴリズム
画像処理から創薬まで

2000

知識準拠診療支援（DSS)
医療ターミノロジー
医療オントロジー

ビッグデータと機械学習

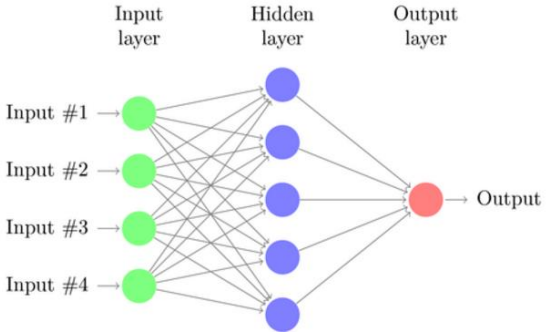
- **The ASCO (米国臨床癌学) CancerLinQ initiative**
  - 診療の現場(EHR)から大量の診療データを集め分析
  - 新しい臨床治験へのガイドライン作成
  - 17万人のがん症例データベースを構築。各がんについて１～２万人の症例を集める
  - 学習システムを構築し治療知識を統計学習、ニューロネットを駆使して学習。
    BigDataにおけるLearning systemの不可欠性
  - 2013年に、CancerLinQのプロトタイプを完成、10万人以上の乳がんを蓄積、完全規模へ継続構築中
- **IBM Watsonのがんセンターへの普及**
  - 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.”

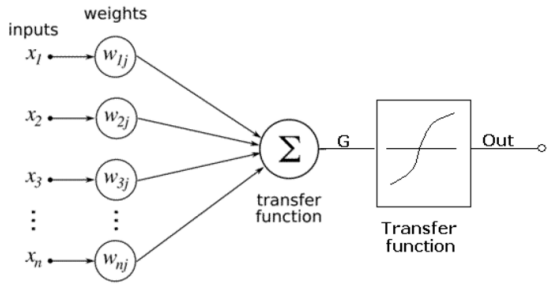
IBM Watson Learning Big Data

TMDU

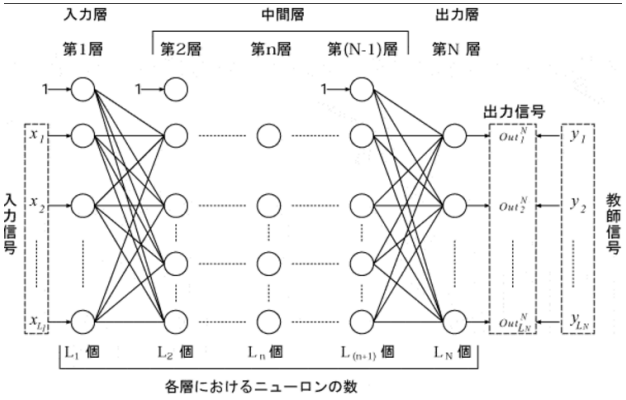
Deep Learningによる基軸成分の抽出

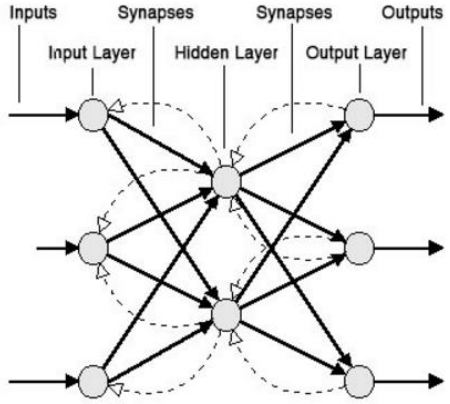
古典的Neural Network(1970年代）





多層Neural NetworkとBack projection (1980年代）



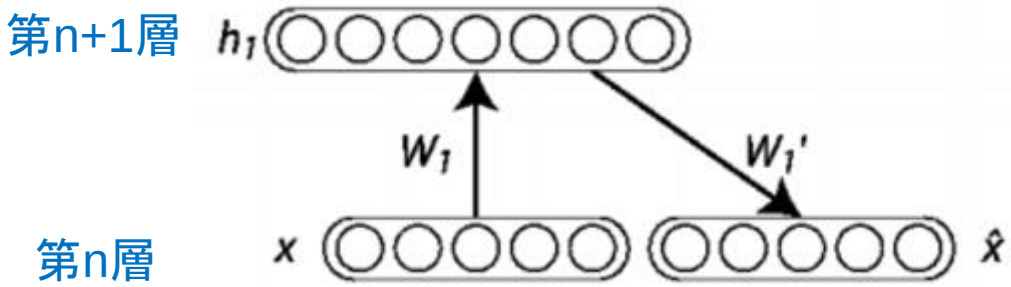


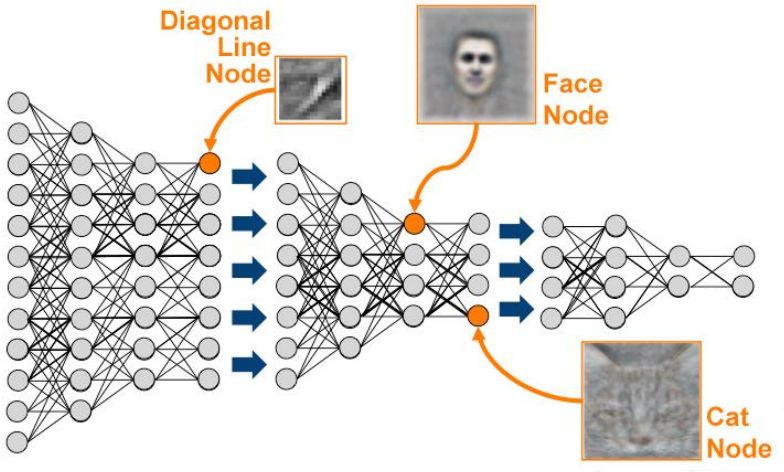
Back Propagation (1986　Rumelhart)
望ましい出力との誤差を教師信号として与える事により、次第に結合係数を変化させ、最終的に正しい出力が得られるようにする。
結合係数を変える事を学習と呼ぶ。
この学習方法には、最急降下法（勾配法）が使われる。出力層へ寄与の高いノードの重みが変更。

Deep learning どこが新しいか

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

（１）最初に「教師無しデータ」を利用して、各レイヤーのパラメータを一層ずつ調整。
（２）最初の層を学習する場合は入力を変換し逆変換をかけ元の入力と比較し一致するようにパラメータを更新。
少ない表現力で入力の情報を表現するようにパラメータが調整される。入力情報を最も良く表現できるような関数が抽出。
基本的な特徴情報が取得される。
（３）autoencoder：変換をかけて元の信号に戻せるように学習する方法
（４）第一層の結合係数は固定して次の階層の学習に入る
（５）最後の層が学習できれば、最後は逆伝播で微調整する

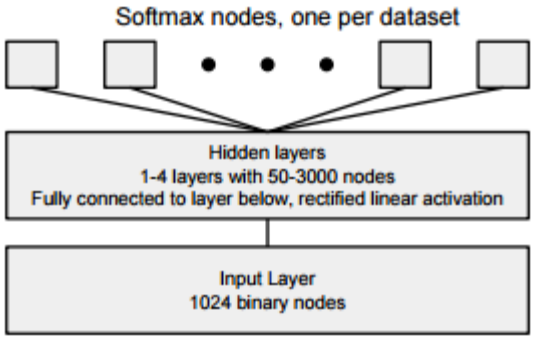




Deep learning ：創薬からの注目

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

バーチャルドラッグスクリーニングに対する deep learningによるツール開発
"Massively Multitask Networks for Drug Discovery"

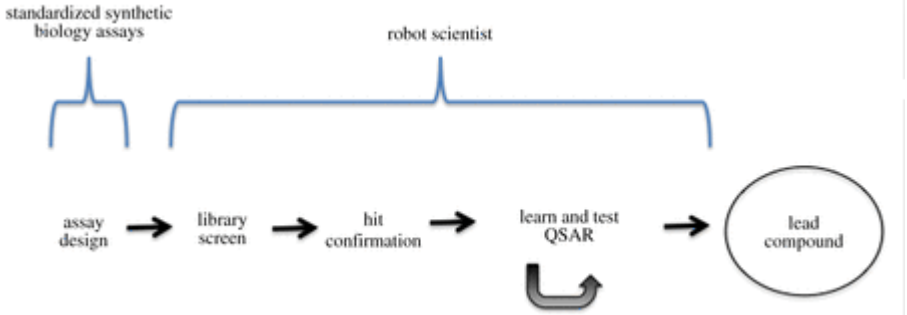


Massively Multitask Networks

そのほかのAI創薬の話題

- Berg社のAI創薬
  - AIを方法として膵臓がんの抗がん剤を開発中
  - 膵臓がんと非患者の14兆のゲノム・オミックス情報を比較。
  - 調節不全パスウェイのシステム推定
  - システム薬理学的AIIによる創薬（詳細不明）
- マンチェスター大学（ Cambridgeとも共同）
  Artificially-intelligent Robot Scientist for new drugs
  - ライブラリースクリーニング, ヒット化合物の確証, リード化合物などの自動化
  - 構造活性相関（Quantitative Structure Activity Relationship）(QSAR) を反復学習する
  - 熱帯病、寄生体のDHFR（ジヒドロ葉酸還元酵素：薬剤耐性）を標的にして学習、細胞を合成生物学操作
    血管新生阻害因子（抗がん剤）をDR候補を探索
  - 最上位にコンセプト木（“root: assay triple screen”など)

Robot scientist Eve at work



TMDU

Real-World-Dataを用いた創薬育薬の戦略の将来性

―RCT, EBMからの呪縛の解放―

TMDU

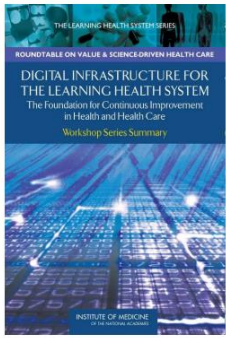
「学習する医療システム」
Learning Health System

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

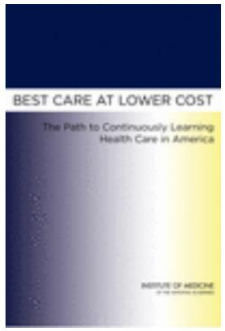
- IOM “Clinical Data as a Basic Staple of Health Learning”
- 医療システムのデジタル化（IT化）は必然の傾向である
- 「ルーチンの医療活動から集められたデータ（形式的臨床研究と違って）がLHSを支える鍵である」
- データを共有することによって学習して医療システムを改善
- RCTは「黄金基準」であるが、通常の医療システムの外で実施されている。医療が実際対象とする患者集団を代表しているのか。
- RCTは時間が掛かり費用もかかる
- 有効な知識の蓄積の速度が加速する

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

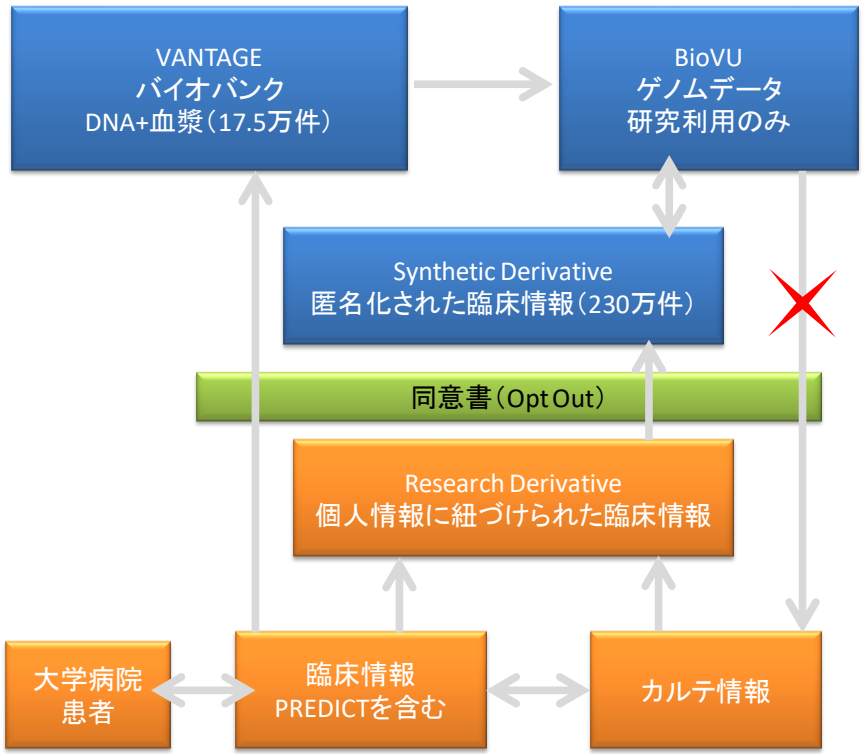
Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care



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。BuiVU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。

TMDU

個別化（層別化）医療の概念の普及とRCTの限界

- 個別化・層別化の概念の浸透
- RCTの治験集団とReal World Dataの乖離
  - すべての個別化パターンを包摂した治験集団は現実には不可能
  - 現在の治験集団
    - 大半のRCTは医療現実の外の「人工的な環境」
    - 高齢者・妊婦はいない、欧米では黒人とくに青年は含まれない
- 将来へ向けたプラットフォームの確立
  - 母集団に近いReal World 医療データが収集可能
    ⇒データの大規模化の「相転移」
  - Real World Data時代の臨床研究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
    RCTとRWDの融合としての
    registry-based clinical randomized trial
  - 我が国の戦略　段階的移行

Biobank準拠の創薬・治験

- 疾患レジストリ―/疾患型バイオバンク準拠型ランダマイズ治験
  - スウェーデンのSWEDEHEART
  - Registry-based randomized clinical trial
  - 疾患レジストリ―の登録患者から治験に適した治験対象者を選び
  - 選んだ集団で治験薬・対照薬をランダマイズして割付ける
  - 治験のエンドポイントは疾患レジストリーの追跡で観測される
  - 観測研究であるPopulation 型コホートでは困難か

ビッグデータ創薬の
有効性と将来展開

TMDU

創薬を巡る状況

- 医薬品の開発費の増大
  - １医薬品を上市するのに約700億円
- 開発成功率の減少
  - 2万~３万分の１の成功率
  - とくに非臨床試験から臨床試験への間隙
  - phase II attrition (第２相損耗)
- **臨床的予測性**
  - できるだけ早い段階でのヒトでの有効性・毒性の予測
- 臨床予測性の向上
  - 罹患者のiPS細胞を使う
  - ヒトのビッグデータを使う

15~17年

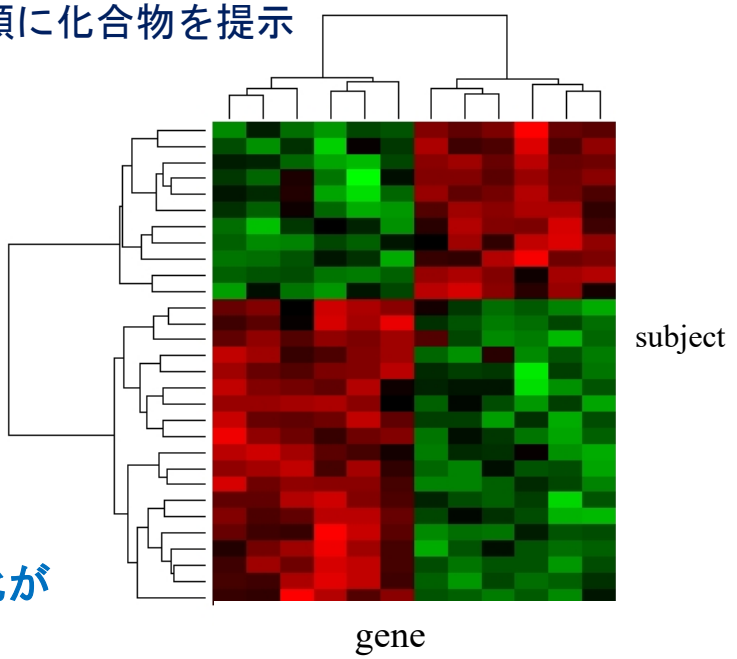
		化合物数（自社）	前の段階から移行できる確率	累積成功率
2~3年	合成(抽出)化合物数	652,336		
3~5年	前臨床試験開始決定数	203		1: 3,213
3~7年	臨床試験開始数	75	1: 2.71	1: 8,698
1~2年	承認申請	26	1: 2.88	1: 25,090
	承認取得	21	1: 1.24	1: 31,064

製薬協データ2011

製薬協ガイド2014

オミックス創薬の原理

- 薬剤特異的遺伝子発現（Drug-induced SDE)
  - CMAP：Connectivity Map
    - 薬剤投与による遺伝子発現プロファイルの変化
    - 米国 Broad Institute,1309化合物, MCF7,PC5など5 がんセルライン, 7000 遺伝子発現プロファイル
    - Signature (遺伝子発現刻印：差異的発現遺伝子の代表的集合）
      Signature of Differential gene Expression
    - DB利用：SDEをquery, 順位尺度で類似性の高い順に化合物を提示
    - 最近LINCS　100万サンプルへ拡張
- 疾病特異的遺伝子発現
  (Disease-associated SDE)
  - GEO（gene expression omnibus),
    - 疾病罹患時の遺伝子発現プロファイルの変化
    - 米国NCBI作成・運用　2万5千実験, 70万プロファイル
    - ArrayExpressもEBIが作成、サンプル数同程度



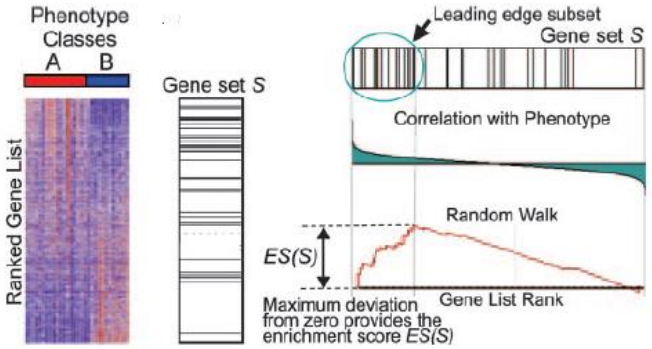
本来は、分子ネットワークの疾病/薬剤特異的変化が基本（第３世代網羅的医学）。

遺伝子発現プロファイル変化
≈　分子ネットワーク変化

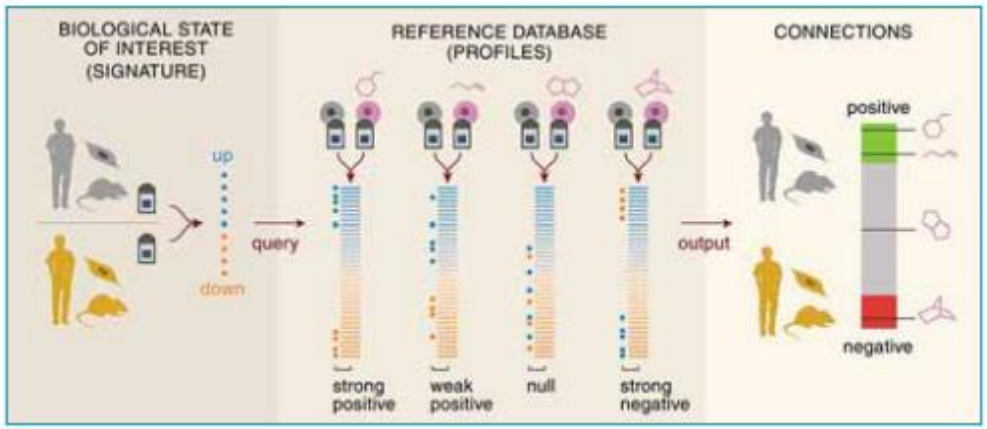
TMDU

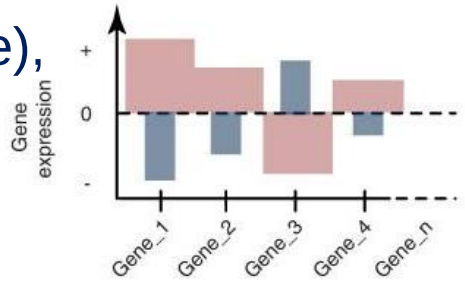
遺伝子発現プロファイルによる有効性予測

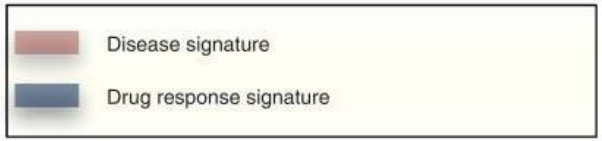
- 遺伝子発現シグネチャア逆位法 (signature reversion)
  - 薬剤特異的遺伝子発現シグネチャア
  - 疾患特異的遺伝子発現シグネチャア
  - 有効性予測：両者が負に相関する
  - Non-parametric な相関尺度で評価
    - Gene Set Enrichment Analysis（GSEA)：ES score
    - 対照と比較して順位づけられた遺伝子リストの上位に密集しているかの尺度
  - 例：炎症性腸疾患IBDに 抗痙攣剤(topiramate), 骨格筋委縮にウルソール酸



GSEA



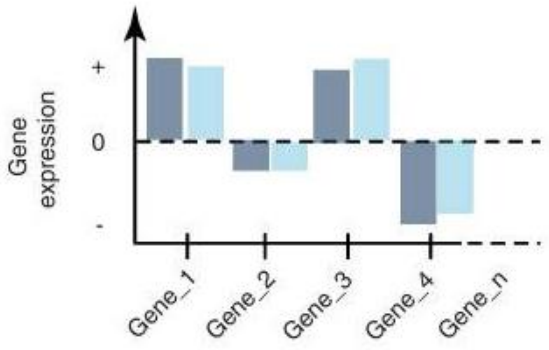




TMDU

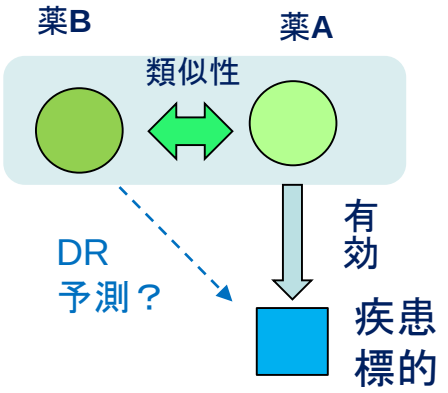
遺伝子発現プロファイルによる毒性予測

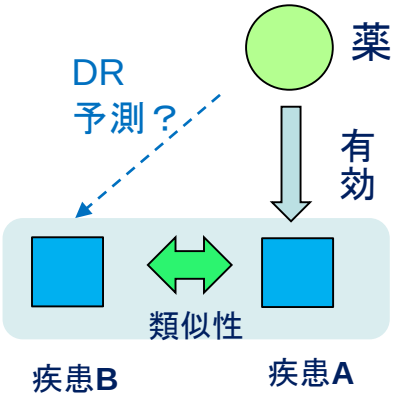
- 連座法 guilt-by-association ：
- 薬剤ー疾患間　副作用予測
  - 薬剤特異的シグネチュアと
  - 疾患特異的シグネチャアが
  - ノンパラメトリック相関　正
  - 毒性・副作用の予測
- 薬剤ー薬剤間
  - 薬剤ネットワークからのDR
  - Connectivity map から薬剤特異的遺伝子発現の薬剤間の類似性をノンパラメトリック親近性尺度（GSEA)で評価
  - この類似性のもとに薬剤ネットワーク構築
    近隣解析によりDR
  - 例：抗マラリア剤をクローン病に適応



合理的DRへのアプローチ

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





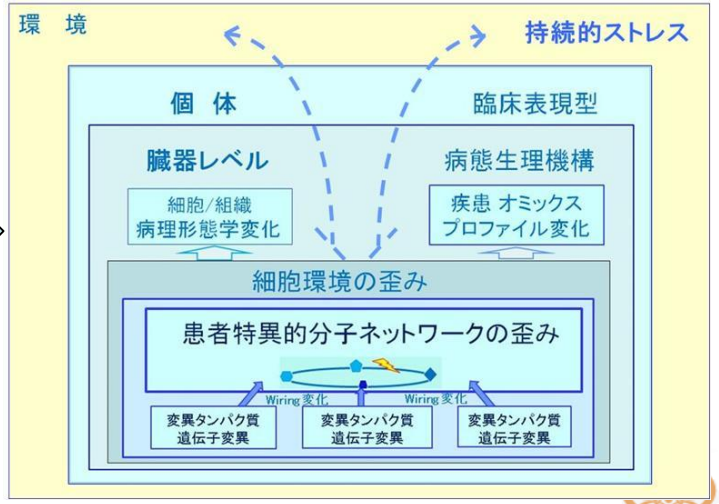
システム分子医学（2010～）

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
疾患のシステム的理解に基づく医療

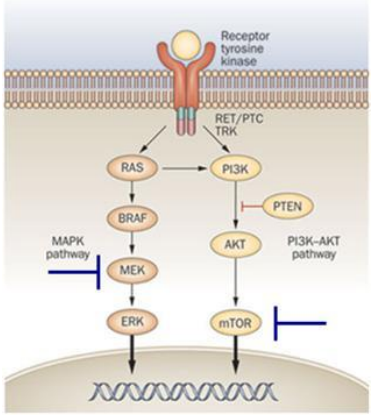
疾患オミックス
(molecular phenome)
成り立たせる基底としての
「細胞分子ネットワークの
構造変化」
疾病の理解における
第３のパラダイム
現在のゲノム医療では
多因子疾患に対して無力

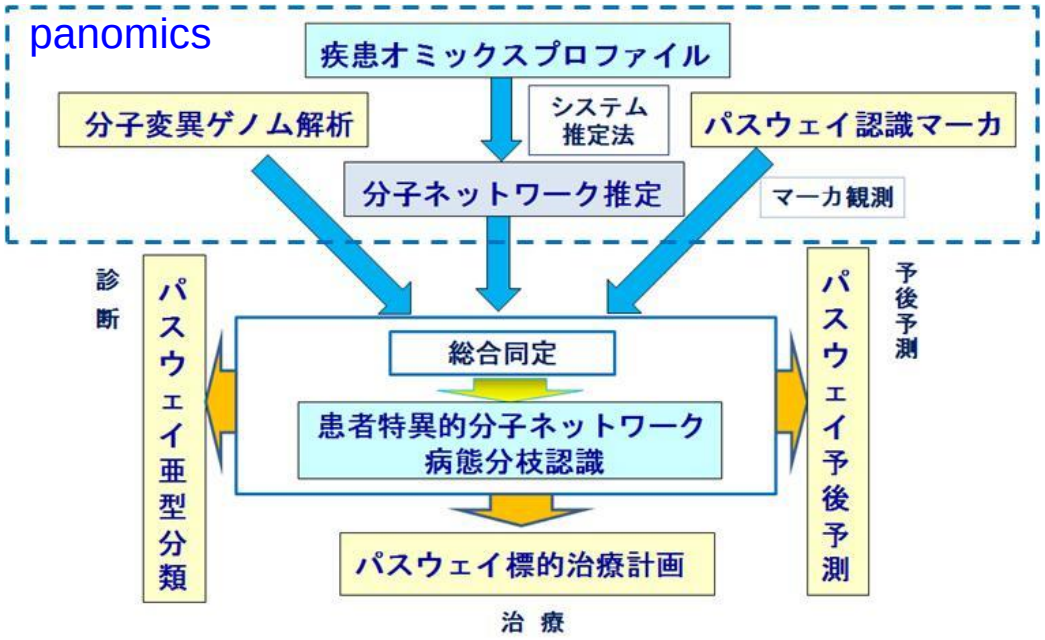


TMDU

個別化医療の展開

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





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

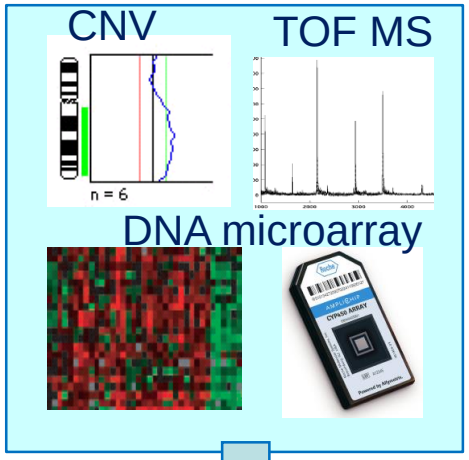
2. パンオミックスによる臨床的実践の戦略

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

Precision Medicineとは
分子システム医学のことである

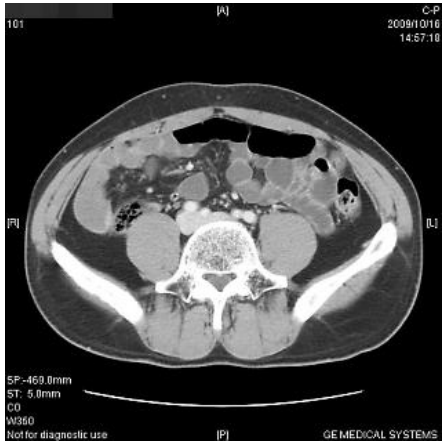
疾患
多層オミックス・プロファイル

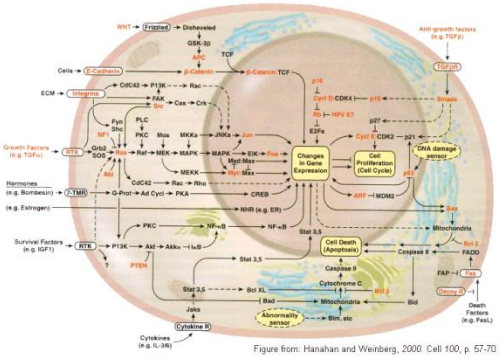
断層像
再構成技術



スーパーコンピュータ

逆計算





患者特異的分子ネットワーク
調節不全分枝同定

合理的な
診断・治療
予後予測

合理的な
診断・治療
予後予測

合理的な
診断・治療
予後予測

AI

BioBank

New Knowledge

ご清聴ありがとうございました