医療ビッグデータの活用

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

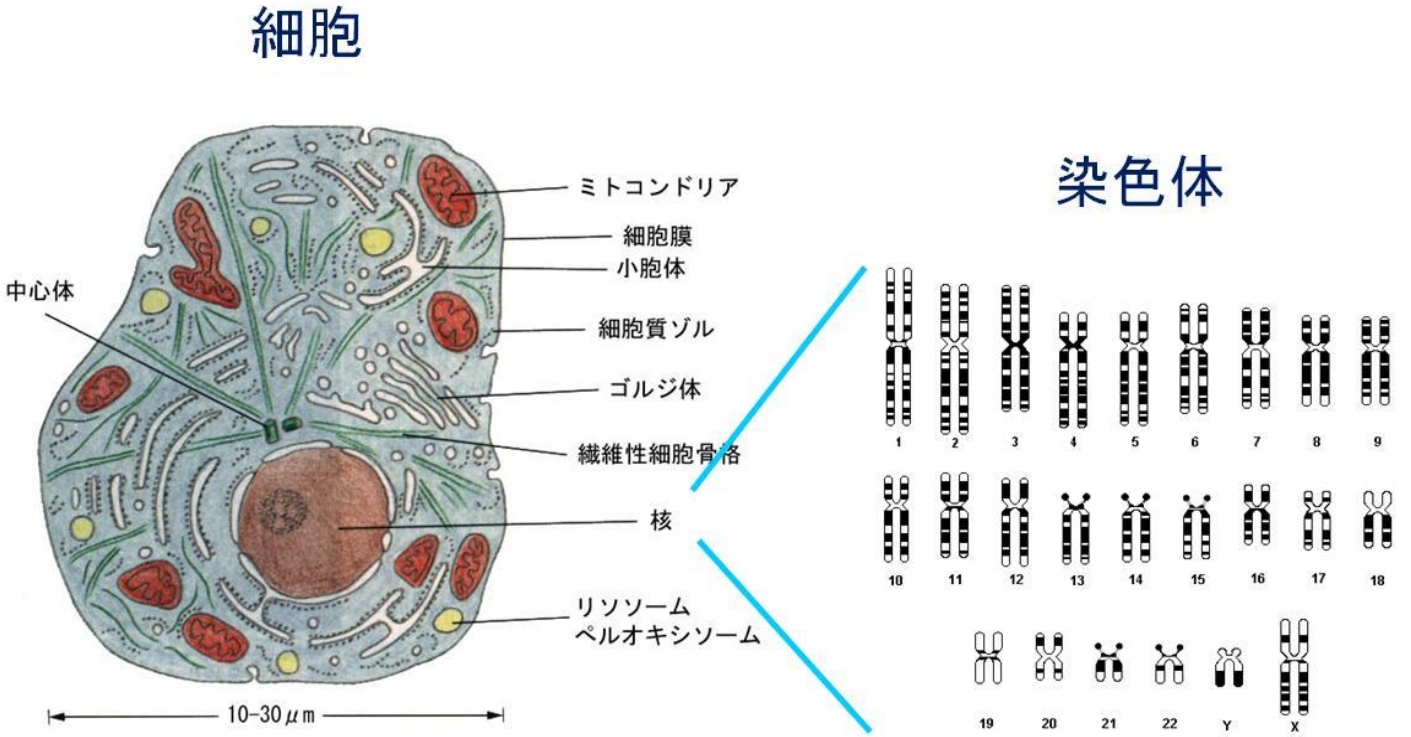
!!

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

ゲノムとは

- 遺伝情報の総体を表す言葉で、染色体上に存在する、ヒトの場合約32億のDNAの塩基成分の配列がもたらす全情報を指す
- 染色体の種類は23種あり、22種類の常染色体と、X,Y染色体の性染色体がある
- 遺伝子：約２万１千（20,687遺伝子ENCODE計画）の遺伝子が存在する
- コーディング領域、合計でゲノムの約1.5%



原核生物と比較したヒトを含む真核生物のゲノムの構成原理

i) DNA二重ラセン

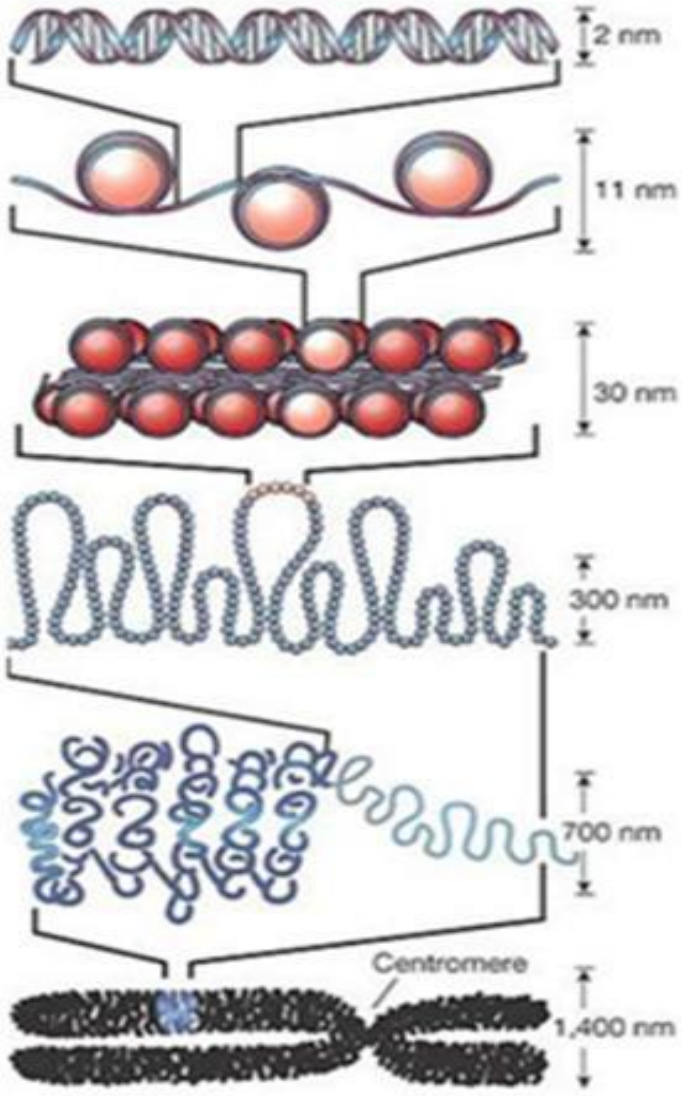
ii) 10 nmクロマチン繊維
（糸を通したビーズ状態）

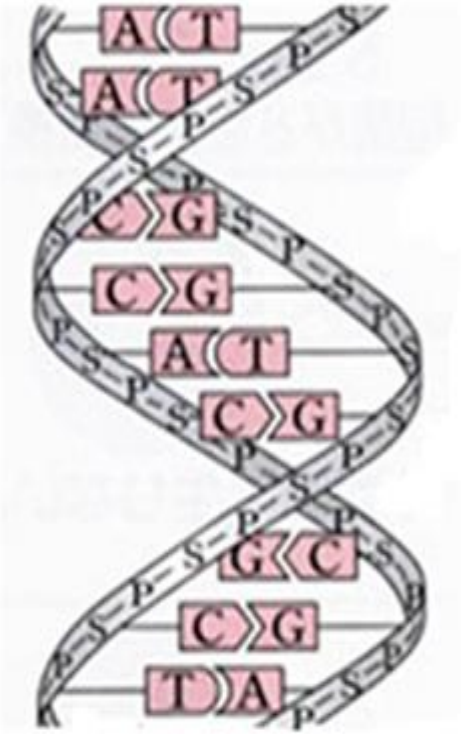
iii) 30nmクロマチン繊維
（10nm繊維をらせん状に折り畳む）

iv) 弛緩した構造の染色体
（分裂期以外の通常状態）

v) 分裂中期の染色体
（染色体凝縮）

vi) 中期染色体の全体像



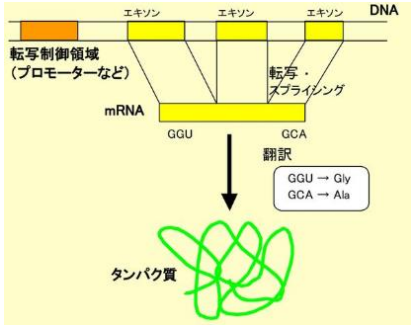


TMDU

ヒトゲノム解読計画

- ヒトの遺伝子の配列が解読できるようになった1980年代は、先天的な遺伝病の**疾患原因遺伝子**の探索が始まった
- 疾患家系図とともにDNAの地図の道標となるDNAマーカによって疾患遺伝子を同定した
- DNAマーカの付近に疾患原因遺伝子を同定するまで、多大な労力を有したことから、ヒトゲノム全体を解読することが必要ではないかという考えが生まれてきた。
- ヒトゲノム国際機構（HUGO）が結成され、ヒトゲノム計画が国際的にも1990年に開始され、米国、英国、日本、ドイツ、フランス、中国などが参加した。
- 生物科学の世界では初めての国際協力プロジェクト
- 1990年から2003年まで13年・
  30億ドル掛かった。
- 翌2001年2月には両者の粗配列でのヒトゲノムの解読結果が学術雑誌「ネイチャー」（国際コンソーシアム），「サイエンス」（セルラ社）両紙の特集号に公表された

ヒトゲノムの大域的構造



ヒトゲノム
3200Mb

遺伝子
関連領域
1200Mb

遺伝子
間領域
2000Mb

遺伝子
48Mb

関連配列
1152Mb

散在反復配列
1400Mb

他の遺伝子間配列
600Mb

1.5 %

LINE
640Mb

LTRエレメント
250Mb

マイクロサテライト
90Mb

その他
510Mb

偽遺伝子

遺伝子断片

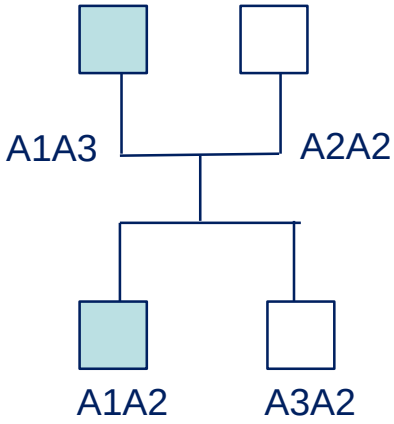
イントロン
UTR

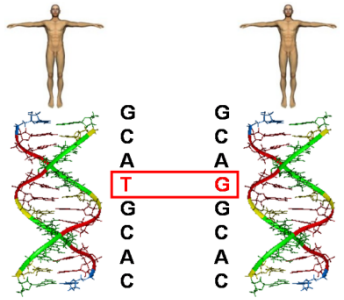
SINE
420Mb

DNAトランスポゾン
90Mb

ヒトゲノムの変異と多型性

- 「生得的な(germ-line)」**変異や多型性**
  - 生得的ゲノムは全細胞で生涯を通じて同一
- 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,デシャンヌ型筋ジストロフィー,嚢胞性線維症　非常に稀少
    - 当時400程度のDNAマーカ ヒトゲノム解読計画へ
- 疾患感受性遺伝子(Disease susceptibility gene)
  - 多型性：一塩基多型(SNP),数千万, 1%以上
    - 他にマイクロサテライト、CNVなど
  - 全ゲノム関連解析（GWAS）HAPMAP プロジェクト、1000ゲノムプロジェクト
  - 「ありふれた病気」（common disease）の発症リスクと関係が深く、SNPの多型は、疾患の発症の相対リスクの増減に関連



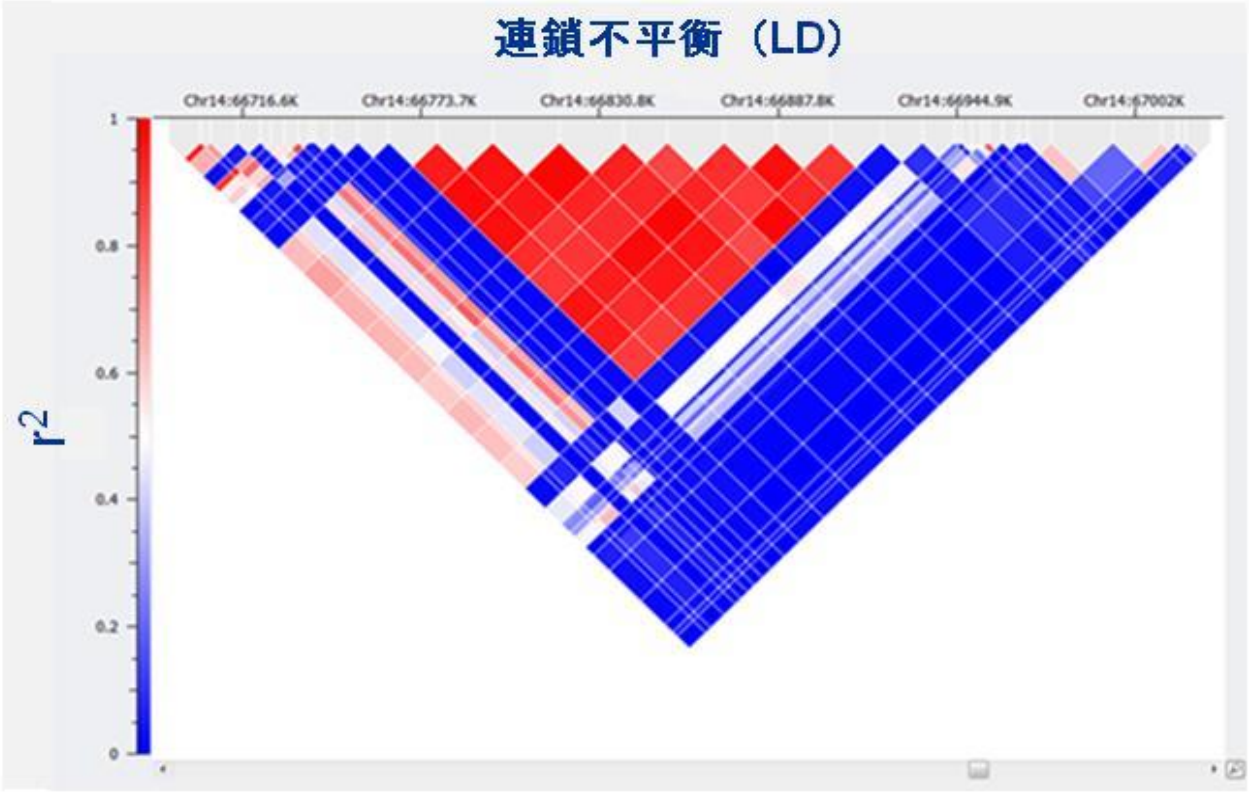


1塩基の変異が疾患や薬剤への応答性に関与する場合がある

TMDU

ゲノムの連鎖構造

- 対立遺伝子が連鎖して遺伝継承されるからである。ハプロタイプ・ブロックは、組み換えの頻発する**組み換えホットスポット**で切断される。ハプロタイプ・ブロックの内部では連鎖が起こり、主要なハプロタイプの種類も少ない。
- ハプロタイプ・ブロックの長さは、平均的には数十kb程度（20kbから40kbが多い）であるが、ときには100kbを超える長大なブロックも存在する



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

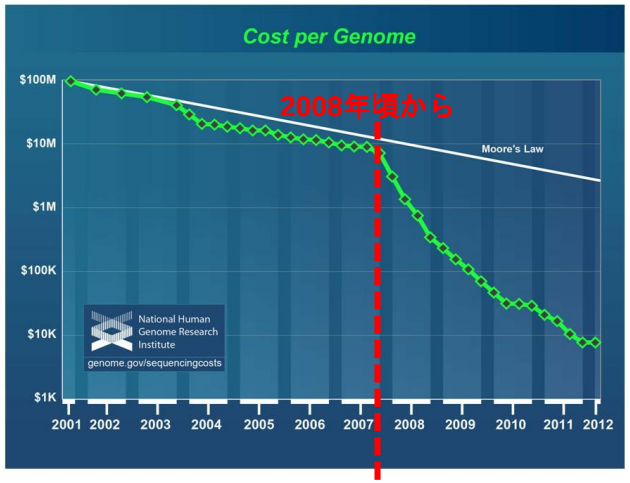
急速な高速化と廉価化　ヒトゲノム解読計画13年,3500億円⇒1日,10万円

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)

Nanopour型
シーケンサ



千ドル
ゲノム

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。

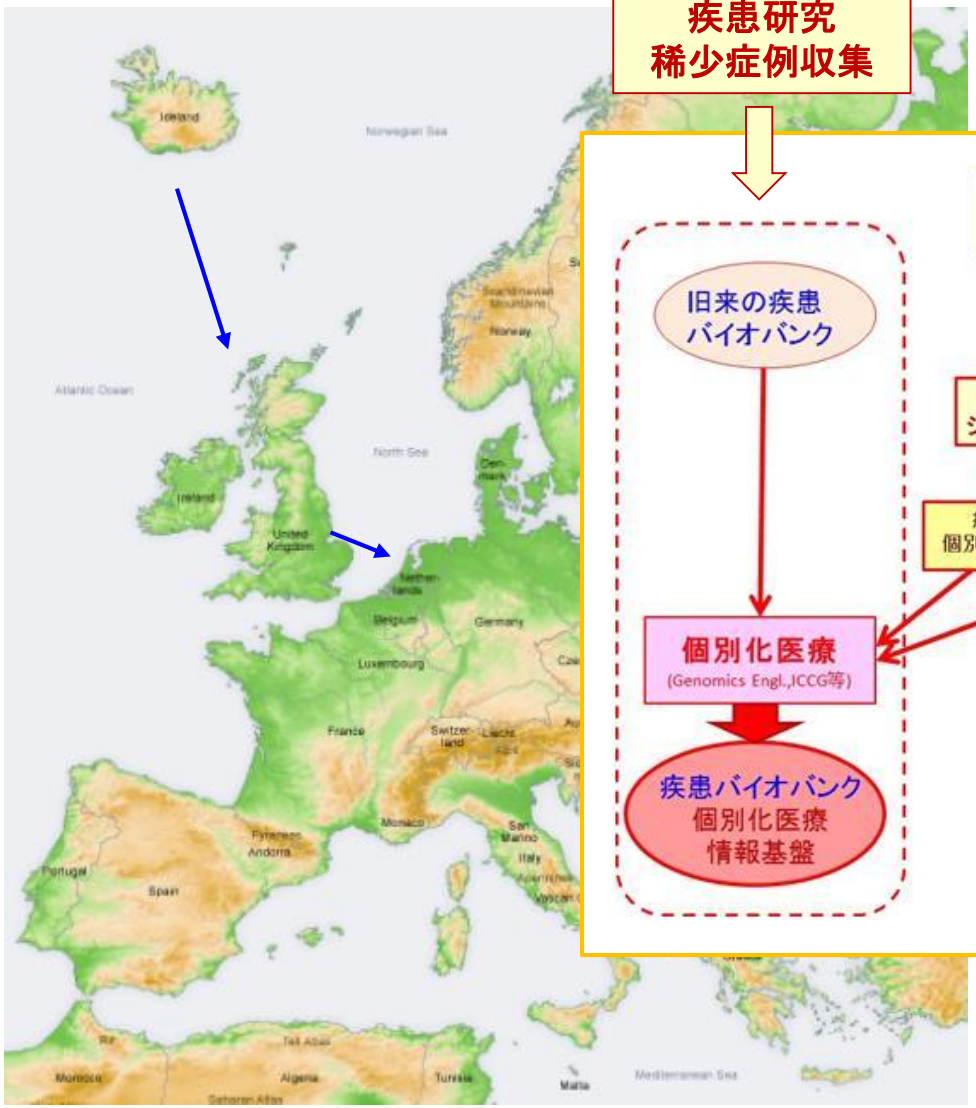
TMDU

ビッグデータ医療の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」からの怒涛の展開（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

TMDU

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

旧来の疾患
バイオバンク

ヒトゲノム計画
(1990-2003)

次世代
シーケンス

GWAS

疾患
個別化機序

ゲノム
生得的
多型性
SNPアレイ

個別化医療
(Genomics Engl.,ICCG等)

疾患バイオバンク
個別化医療
情報基盤

Island/deCODE社
国民遺伝情報DB

前向きコホート概念

ELSI慎重対応

Population準拠

家系情報

UKバイオバンク
ゲノムコホートの確立

LifeLines
3世代コホート

東北メディカル
メガバンク

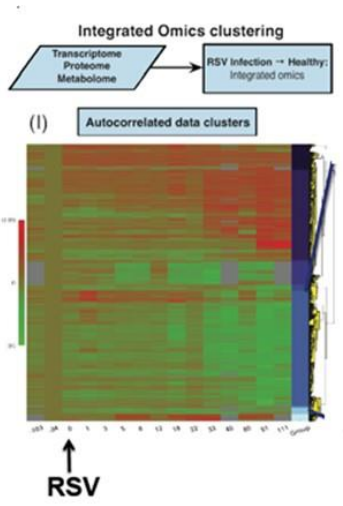
出生コホート

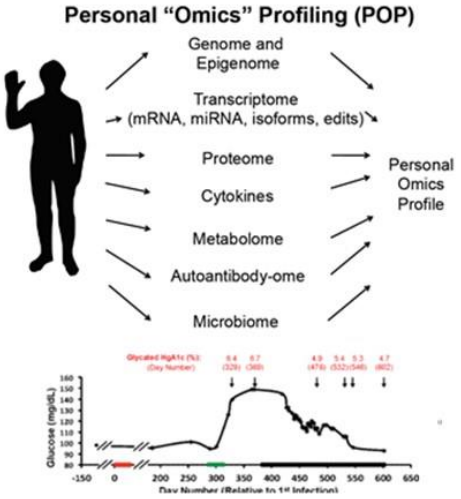
大規模前向きゲノムコホート
国民的医療・予防医療
情報基盤

もう一つの医療ビッグデータ
モバイルヘルス

- 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。サンフランシスコから広がる
  - 東北大学 - 東芝COI
    - 「さりげないセンシングと日常人間
      ドックで実現する理想自己」
- Dr. John Halamka
  - 埋め込み式マルチセンサー
- Dr.Snyder
  - Integrated personal omics profile (iPOP)

ECG; EEG; Skin Conductivity; EVG





TMDU

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

TMDU

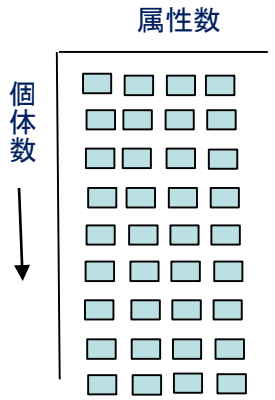
医療の「ビッグデータ革命」

～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(IV)の「ビッグデータ」

$\mathcal{N}$ー Big Data　(n≫p)



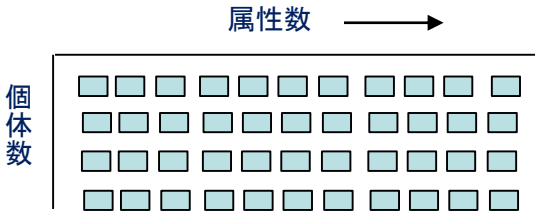
医療情報・疫学調査では 属性数：数10項目程度

– 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

$\mathcal{P}$ー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

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

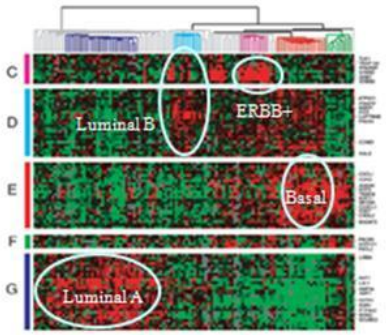
個別化医療
Personalized Medicine

先制医療
Preemptive Medicine

TMDU

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：先天的ゲノムの変異・多型性
    - 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



乳がんの内因性亜型

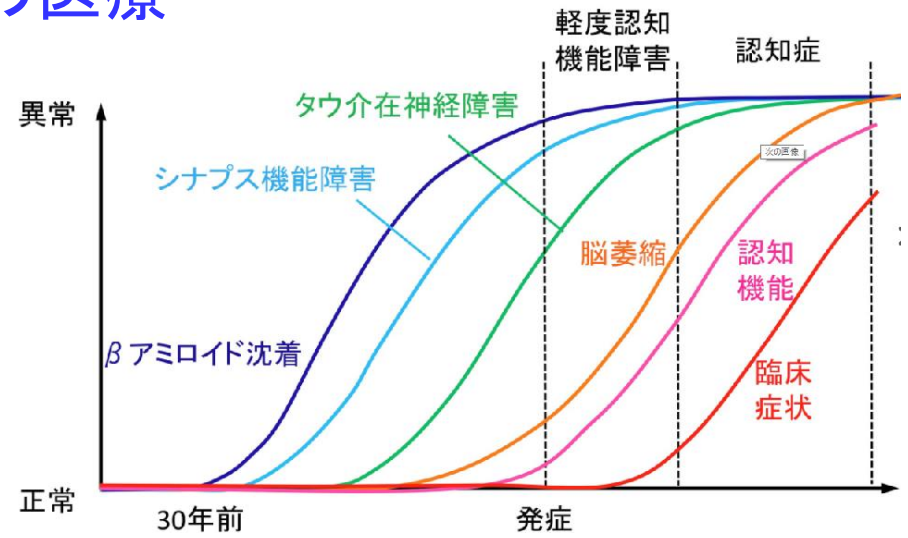
TMDU

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

Disease Big Data

個別化医療・先制医療

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

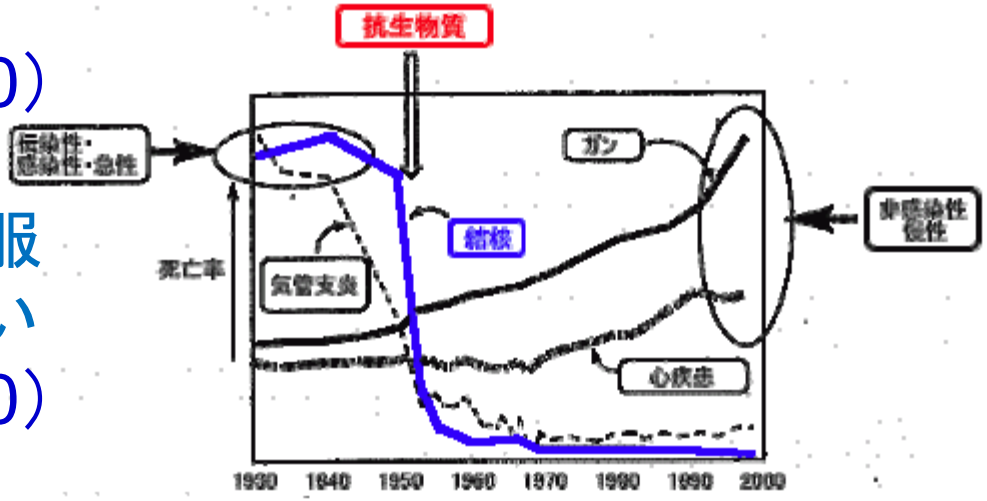
生涯医療・先制医療

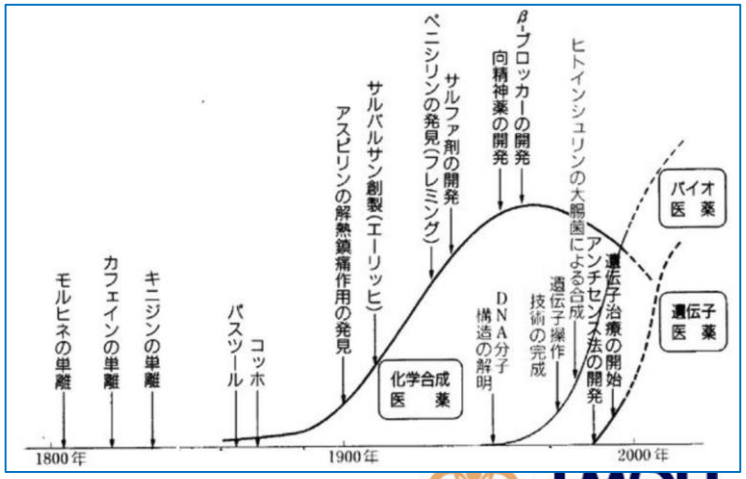
Life-long Big Data

TMDU

２１世紀医療の長期世代交代

- 第１世代(1930～1970)
  - 抗生物質の登場
    により細菌感染症の克服
  - 疾病の病原菌との闘い
- 第２世代(1970～2010)
  - 分子生物学の発展
  - 分子的機序による疾患との闘い
  - 分子標的薬・抗体医薬の登場
- 第３世代（2010～2040）
  - 網羅的分子情報・
    モバイルヘルスの発展
  - ビッグデータ・AIの登場
  - データ駆動型医療の登場





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)

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８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クエンサ決定
- 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)

TMDU

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

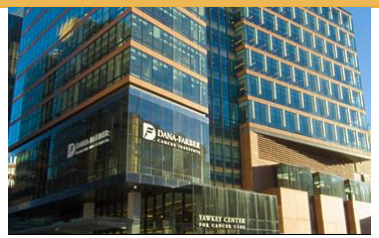
- Wisconsin医科大学　Genome sequencing program
  - Nic君に続いて（翌年3月まで6例）
  - 候補選択（nomination）
    - 従来の検査・診察で診断困難な症例
  - Multidisciplinary 患者選択委員会でレビュー
  - 6-8時間のアセスメントとカウンセリング
  - 32 全ゲノム, 550 全エキソーム（2015年4月まで）
  - アメリカ病理学会（CAP)およびClinical Laboratory Improvement Amendments(CLIA:CMS) 基準：最初外注 Froedtert 病院
  - 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）

Baylor医科大学

ゲノム医療の第２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

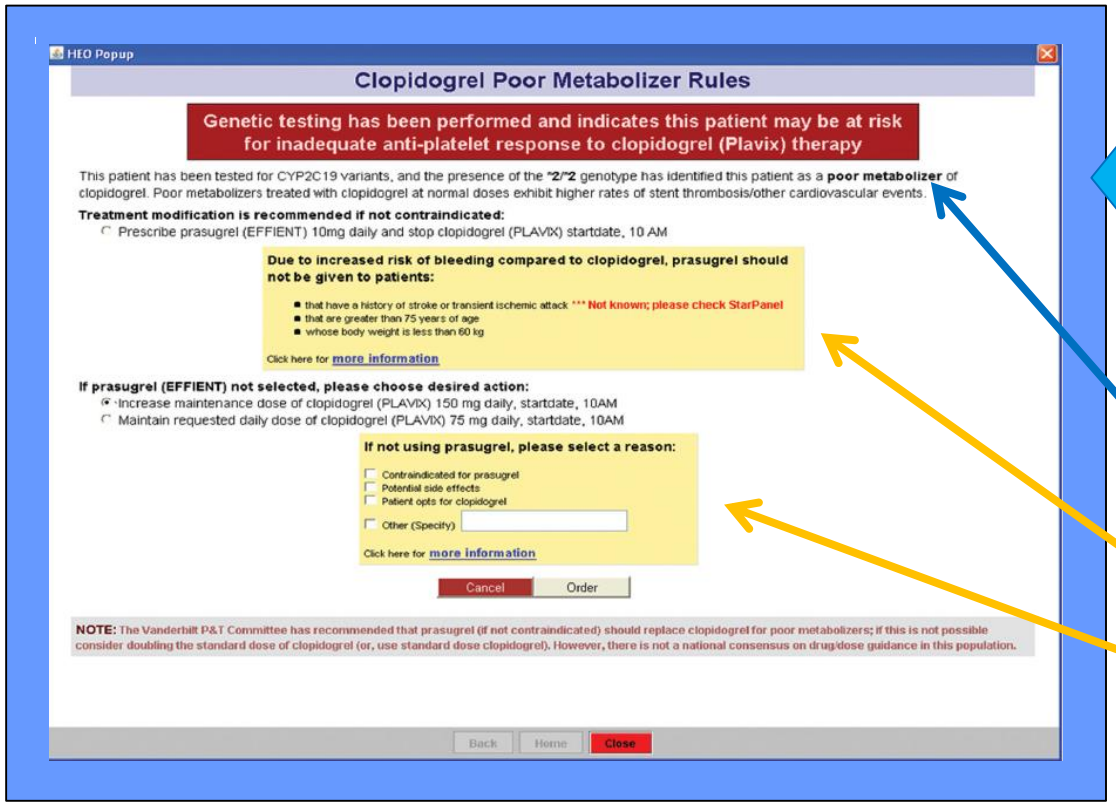
ゲノム医療の第３の流れ

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

Pharmacogenomic **R**esource for
Enhanced
Decisions **i**n **C**are and **T**reatment

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



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)プラスグレル
(商品名エフィエント)に替えるか

分量を2倍にしろと警告している

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

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

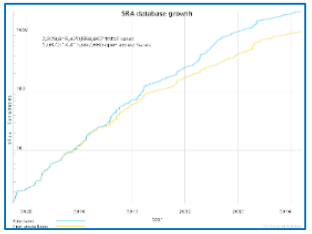
次世代シーケンシングの臨床普及(2010～)

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

Baylor**医科大学**
Mayo Clinicなど
後続病院多数

第１世代

Early adopter 時期

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

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

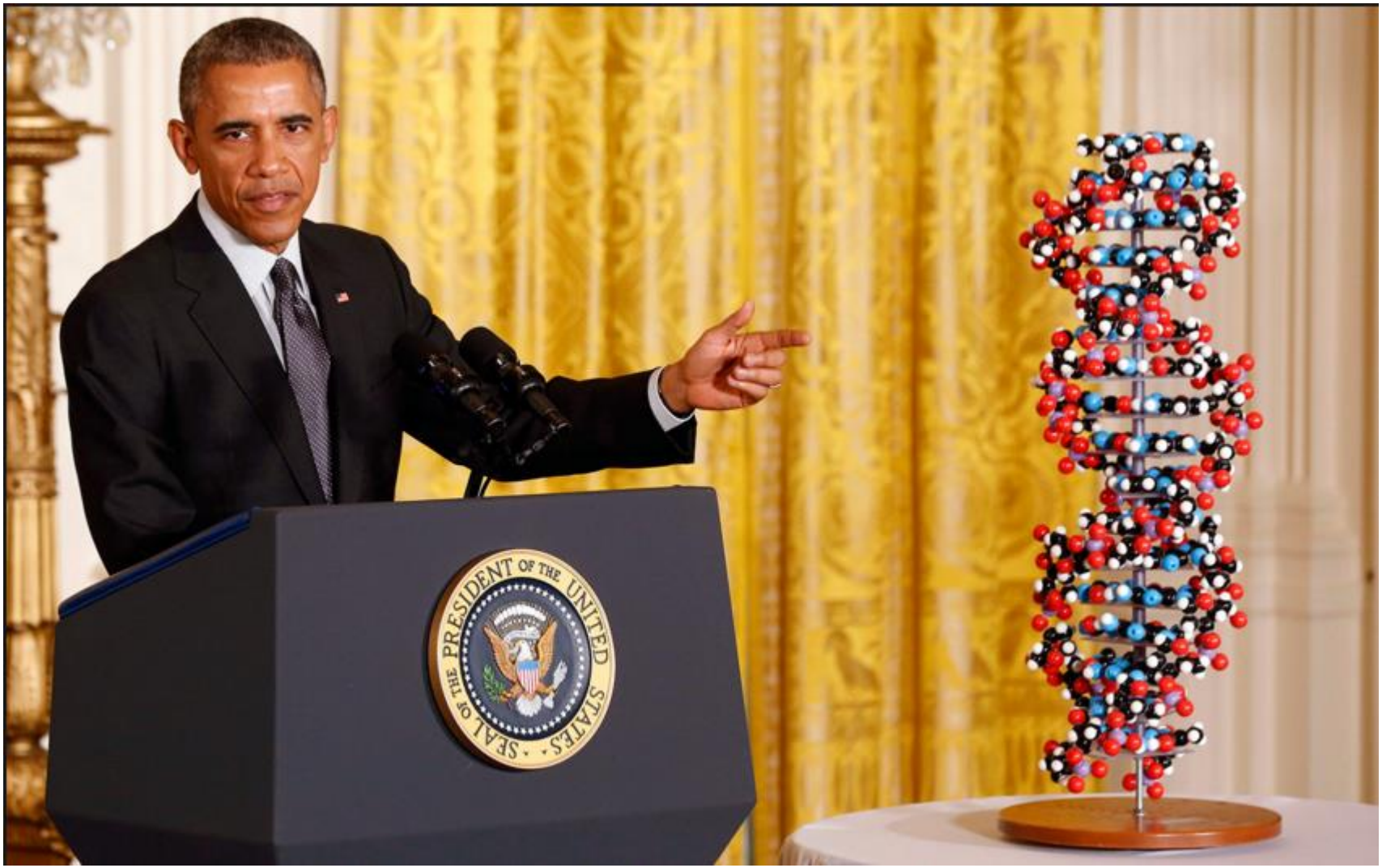
NIH “BD2K ”COE in Data Science, DDI (2014)
ASCO “CancerLinQ”, Cancer Common
“Precision Medicine (Obama) ” 1 M genomic cohort

国家戦略としての「医療ビッグデータ」
Big Data to Knowledge

- 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

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

旧来の疾患
バイオバンク

ヒトゲノム計画
(1990-2003)

次世代
シーケンス

GWAS

疾患
個別化機序

ゲノム
生得的
多型性
SNPアレイ

個別化医療
(Genomics Engl.,ICCG等)

疾患バイオバンク
個別化医療
情報基盤

Island/deCODE社
国民遺伝情報DB

前向きコホート概念

ELSI慎重対応

Population準拠

家系情報

UKバイオバンク
ゲノムコホートの確立

LifeLines
3世代コホート

東北メディカル
メガバンク

出生コホート

大規模前向きゲノムコホート
国民的医療・予防医療
情報基盤

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）の病因に関係する**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

TMDU

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）」我が国では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

研究費を用いた「臨床応用を実施している部門は12施設」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

◎先天的疾患 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)

◎がんのドライバー遺伝子診断：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複数遺伝子パネル・１１がん医療拠点病院
  - 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
- 京大腫瘍内科（OncoPrime)、岡大、北大、千葉大 診療施設併設型BB

◎大規模前向きpopulation 準拠ゲノム・コホート　東北メガバンク計画

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では大きく遅れていはいない

TMDU

最近のビッグデータ医療の状況

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**（治癒行為が**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

ACMG POLICY STATEMENT | Genetics inMedicine

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[a]	Variants to report[b]
Hereditary breast and ovarian cancer	604370 612555	20301425	Adult	*BRCA1* *BRCA2*	113705 600185	AD	KP and EP
Li–Fraumeni syndrome	151623	20301488	Child/adult	*TP53*	191170	AD	KP and EP
Peutz–Jeghers syndrome	175200	20301443	Child/adult	*STK11*	602216	AD	KP and EP
Lynch syndrome	120435	20301390	Adult	*MLH1* *MSH2* *MSH6* *PMS2*	120436 609309 600678 600259	AD	KP and EP
Familial adenomatous polyposis	175100	20301519	Child/adult	*APC*	611731	AD	KP and EP
MYH-associated polyposis; adenomas, multiple colorectal, *FAP* type 2; colorectal adenomatous polyposis, autosomal recessive, with pilomatricomas	608456 132600	23035301	Adult	*MUTYH*	604933	AR[c]	KP and EP
Von Hippel–Lindau syndrome	193300	20301636	C				
Multiple endocrine neoplasia type 1	131100	20301710					
[illegible]crine neoplasia type 2	171400	[illegible]					

100万人コホート“ALL of Us”

- プレシ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・プレシジョン医療コンソーシアムや地域のプレシ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しい成果を上げているにもかかわらす、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も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

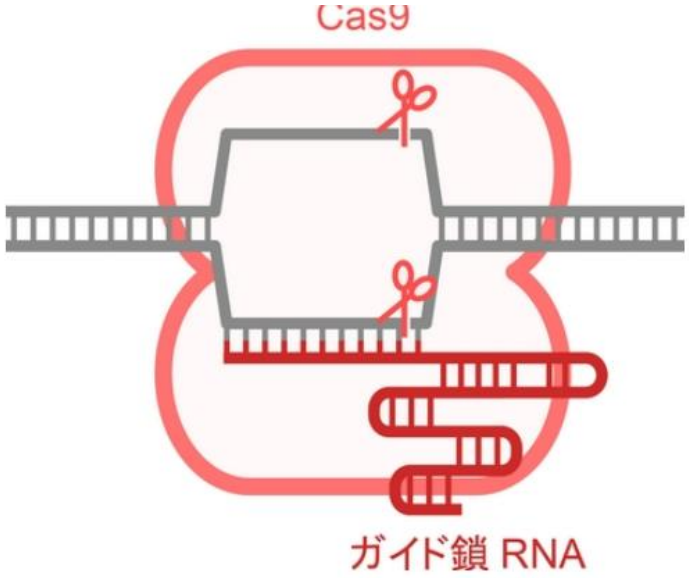
がん細胞と微小環境との相互作用を捉える

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。

TMDU

Precision治療学としてのゲノム編集



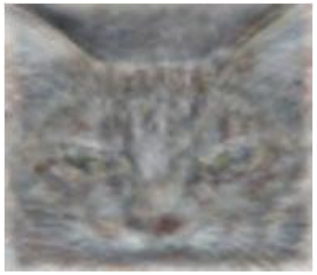
- **Crisper-Cas9**
  - もともとは細菌の免疫
    - ウィルス配列を記憶切断
  - 従来の遺伝子組み換えと違って切断配列を特定
  - オフ・ターゲット効果が問題
- **Ex vivo療法とIn vivo療法**
  - HIV療法やがん治療
    - CAR-T療法と急性リンパ性白血病の少女（レイラ3か月）治癒
- **体細胞・生殖細胞ゲノム編集**
  - 生殖細胞のゲノム編集は基礎研究のみ
    - 筋ジストロピー　変異エキソンを読み飛ばす
  - それでよいのか。「デザイン・ベービー」の危険
  - 国際サミット2015年での発言：6日死亡
  - 中国でのヒト生殖細胞のゲノム編集・2015
    - 広州中山大学ベータサラセミア遺伝子改変

レイラ

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

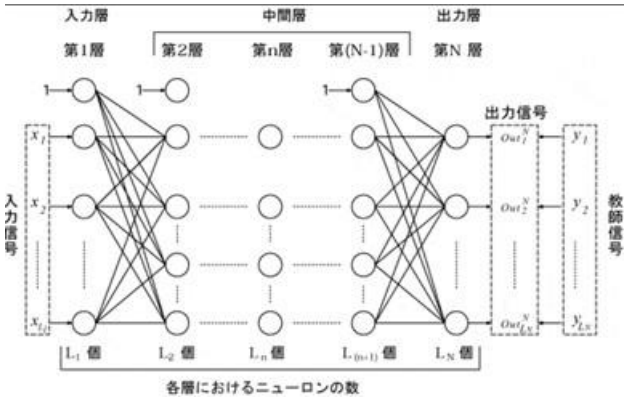
- **「アルファ碁」**（Google DeepMindによるコンピュータ囲碁プログラム）が2016年3月に数多くの世界戦優勝経験のあるプロ棋士李世乭（Lee Sedol：九段）に挑戦し、**4勝1敗と勝ち越した**
  - チェス：IBM「Deep Blue」が1997年に当時の世界champion, カスパロフ氏（ロシア）に勝利
  - 将棋：ボンクラーズ, 2012年米長永世棋聖に勝利
  - **「アルファ碁」**には**ニューラルネットワーク**（Deep Learning)が使われた。**評価経験則が人間によってコードされていない**
  - **最初、**棋譜に記録された**熟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分自身と多数の対戦（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

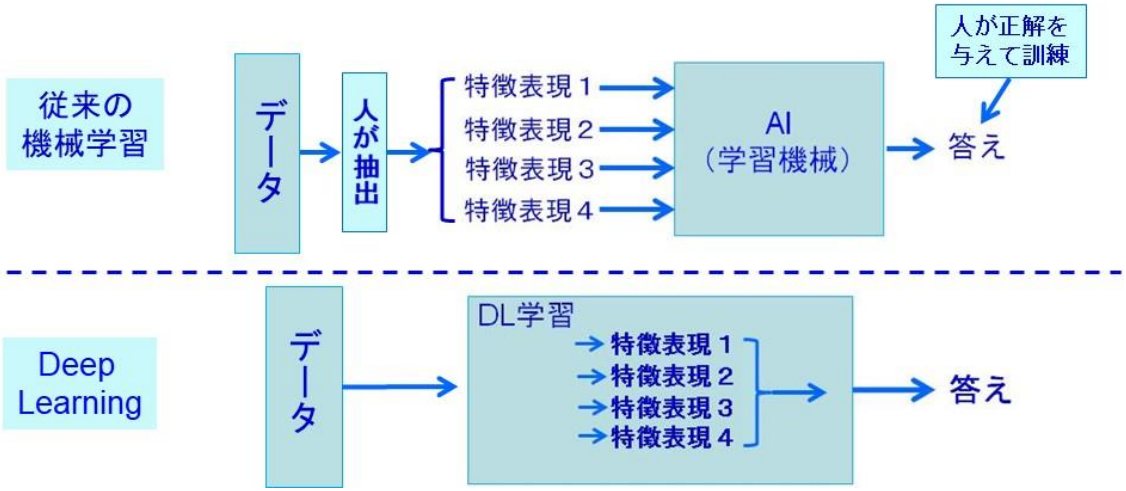


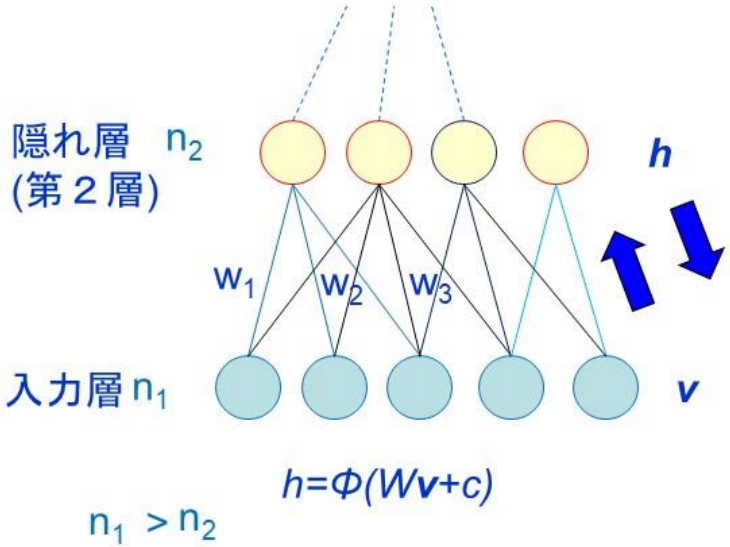
TMDU

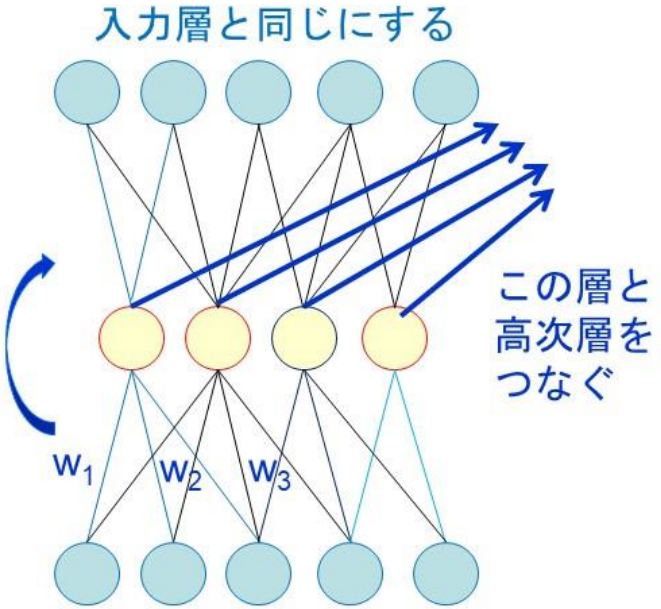
Deep Learning 人工知能革命

ニューラルネットワーク



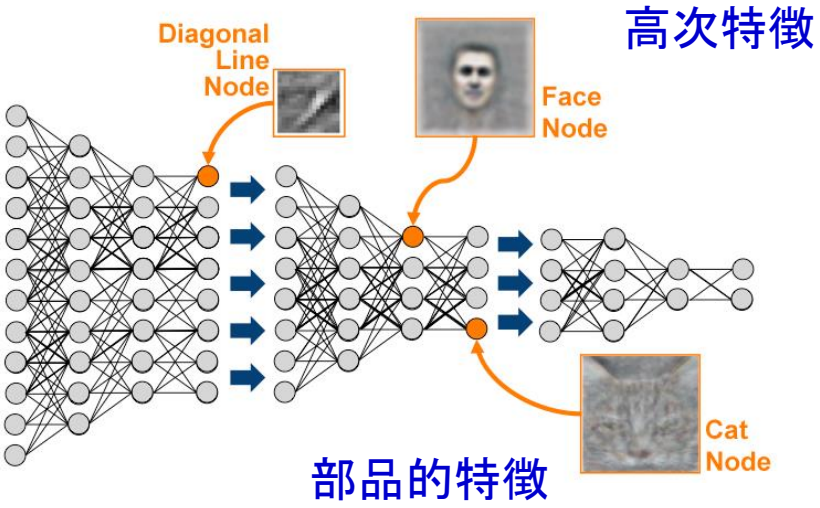






DLの革命点 Autoencode２

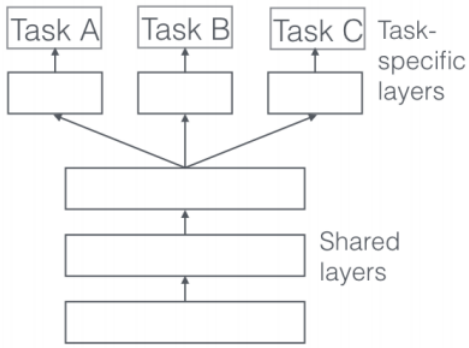
- 各層ごとに自己符号化を行うので何層でも組める
  → **各層間で「自己符号化」の積上げ**（autoencoder stack)
- 第一層で学習した特徴量を使って次の階層を作るので高次の特徴量が作られる
- 特徴的表現と概念を結びつけるため「教師あり学習」が最後に必要。
- 自動特徴抽出によってこれまでの学習手法の限界を克服した
  → 内在的な特徴量による構造的理解
- 人間の「思考の枠組み」を超えた正解の低次
  – **「アルファGo」が定石にない手で碁の名人に勝つ**



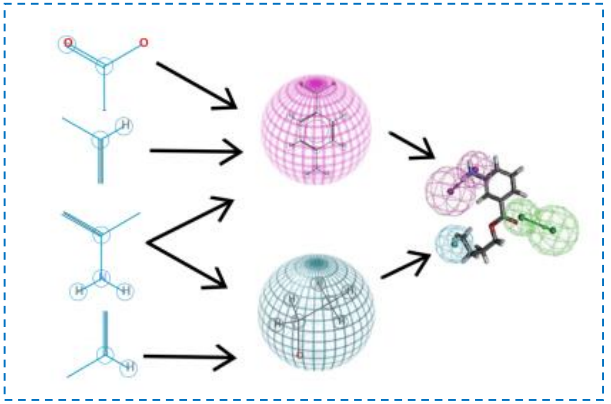
TMDU

Deep learning: 創薬からの注目

- **Kaggle** (データサイエンス競技会)に**Merck社**が出題
  Molecular Activity Challenge (2012).
  - 15種類の標的分子に対する化合物データセットから異なった構造活性相関のデータを学習して構造から分子の生物学的活性を予測するモデルの開発コンテスト
    勝利したモデルはdeep learning を用いた



- Unterthinerの大規模な構造活性相関（QSAR)研究
  - ChEMBLに対するdeep learning
  - 13 M 化合物特徴量（ECFP12), 1.3M 化合物, 5k 薬剤標的
  - Ligand-based 標的予測,7種の予測法とAUC比較
  - Deep learningがSVM, k-最近隣法, logistic回帰より有効
  - 特徴量の抽出、薬理機序への理解
- Google in collaboration with Stanford (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"



AI創薬の方法

- **Virtual Screeningへの人工知能・機械学習の応用**
  - Ligand-based AIバーチャルスクリーニング
  - Structure-based AIバーチャルスクリーニング
- **標的分子探索に人工知能を用いた方法**
  - Hase－Tanakaの多層Deep AutoEncoderを用いた標的分子探索法
- **その他**
  - 化合物の人工知能を用いた自動設計
  - 合成経路設計
  - AI毒性学

薬剤標的分子探索への応用

対象疾患決定後、治癒に有効な生体側の標的分子を探索
　標的分子探索の範囲を限定
⇒ヒト・タンパク質相互作用ネットワーク(PPIN)

学習的アプローチ

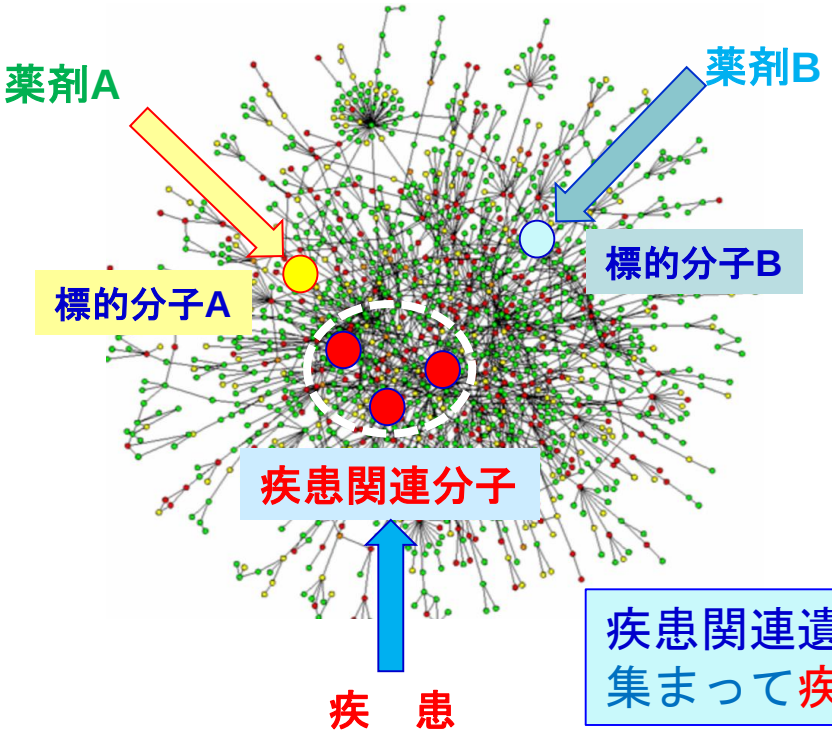
その疾患にこれまで有効な標的分子が既知
既知の標的分子がPPIN上でどのような位置にあるのか
帰納学習する

PPINはHPRDでは＜1万タンパク質 X 一万タンパク質＞の
超多次元ネットワークで通常の機械学習では困難

Deep learning による
＜ネットワーク埋め込みNetwork Embedding＞

標的分子や疾患関連分子の
タンパク質相互作用ネットワーク（PPIN）

- 薬剤ネットワークと疾患ネットワークの基盤：生体分子ネットワーク
- タンパク質相互作用ネットワーク（PPIN）での創薬/DR戦略
- PPIネットワーク場を基礎にして距離（近接性）を検討
- **薬　剤**：**薬剤の標的分子**（タンパク質）によって　PPI場と繋がる
- **疾　患**：**疾患特異的発現遺伝子**を疾患関連分子（**タンパク質**）へ翻訳、
- PPIN場内での**薬剤（標的分子）と疾患（疾患関連遺伝子）の「代理人」**の**距離・近接性**を基準に、薬理作用のインパクト力を評価

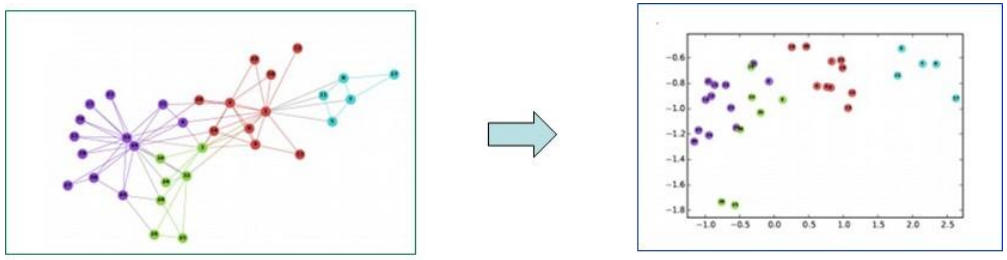


タンパク質相互作用
ネットワーク（PPIN）

疾患関連遺伝子はネットワーク上の近傍に集まって疾患モジュールを形成する

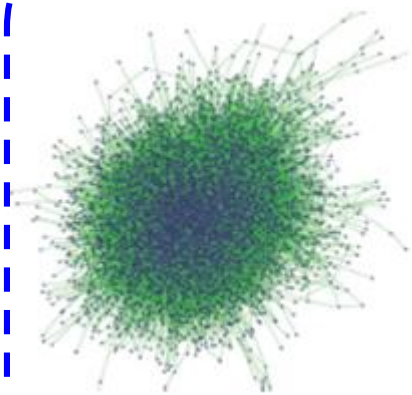
Deep Learningによる Network Embedding

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



Original network　　Latent space

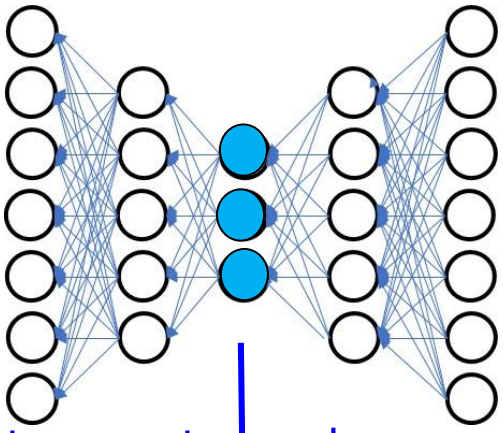
Structual Deep Network Embedding



大規模
ネットワーク

全節点の近接ベクトル

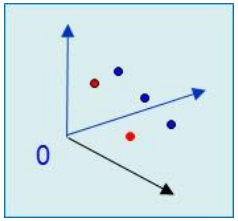
入力層



encoder　　dencoder

対称出力層

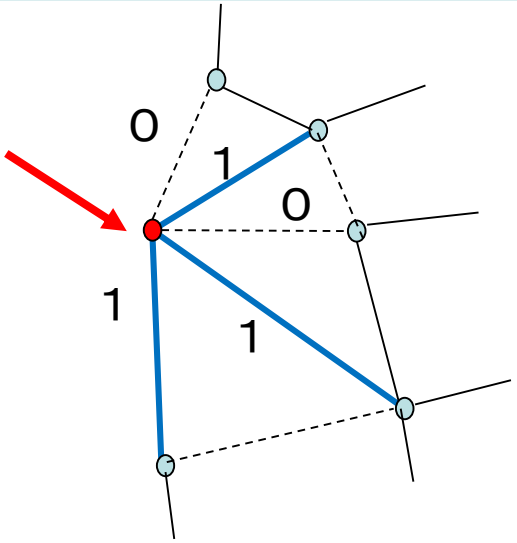
潜在空間
Latent space

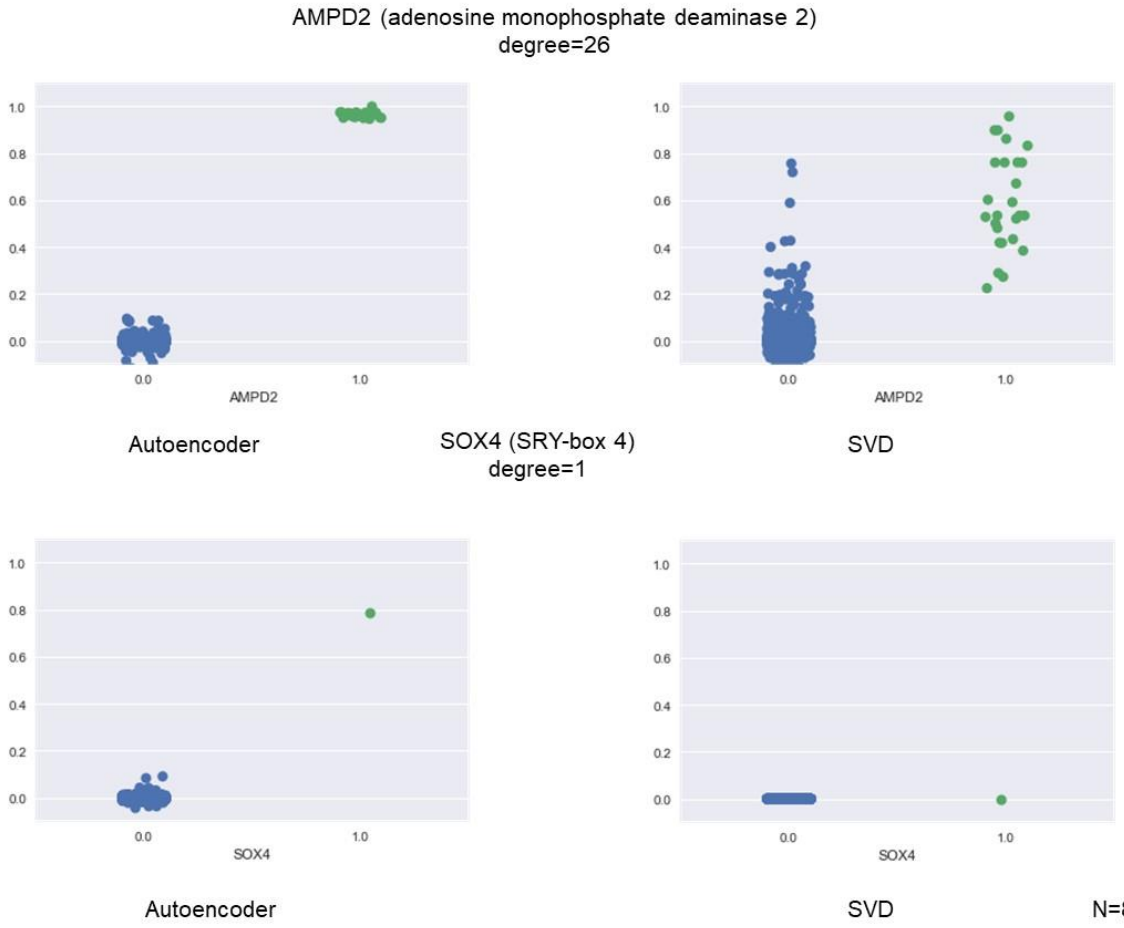


Deep Learning と SVD (singular value decomposition)の精度の違い

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,……..)



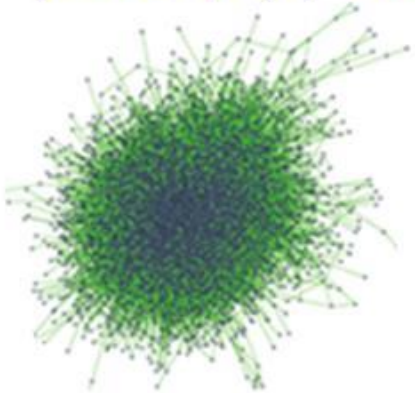


Deep Learningによる創薬・DR

入力

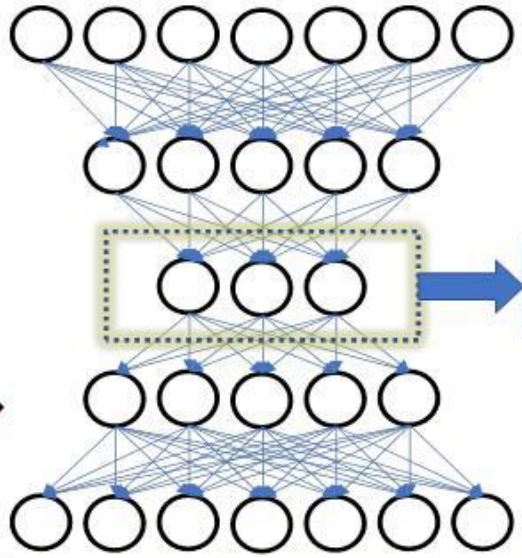
特徴量産出

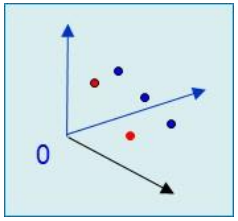
ゲノムワイドPIN



薬剤および標的
データベース

特徴量出力“deep autoencoder”と
特徴抽出アルゴリズム
（DLによる次元縮約）





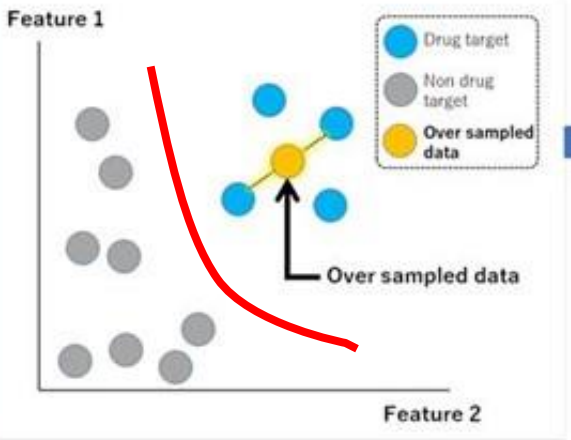
Latent space

Feature selection
by Relief algorithm

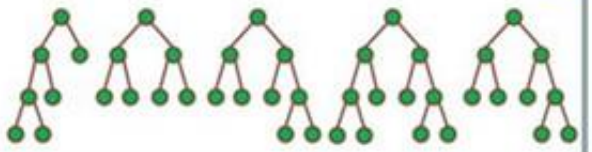
Deep Learningによる創薬・DR

分類モデル

２群分類と標的判定
最新のアルゴリズム



標的性判定 algorithm to build a binary classfier



標的選定

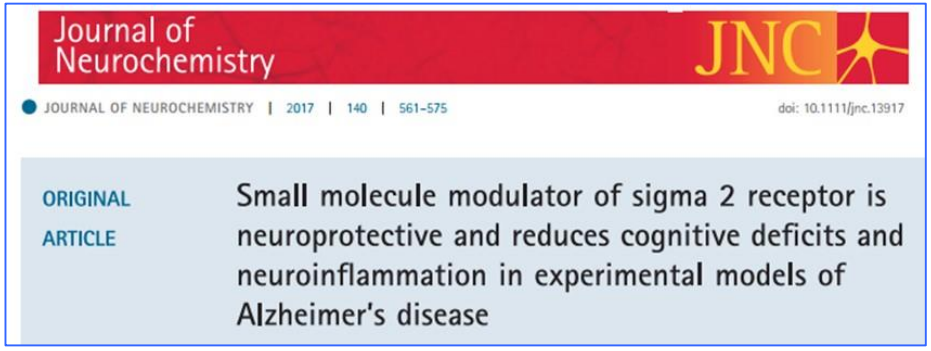
標的性判定

遺伝子	標的確率
GRASP	0.982971
PGRMC1	0.982345
GPM6A	0.982345
NRP2	0.975194
PFKM	0.972128
DLGAP2	0.953659
CD81	0.941095
IQGAP1	0.926867
TROVE2	0.916886

TMDU

実験的研究との付合 1

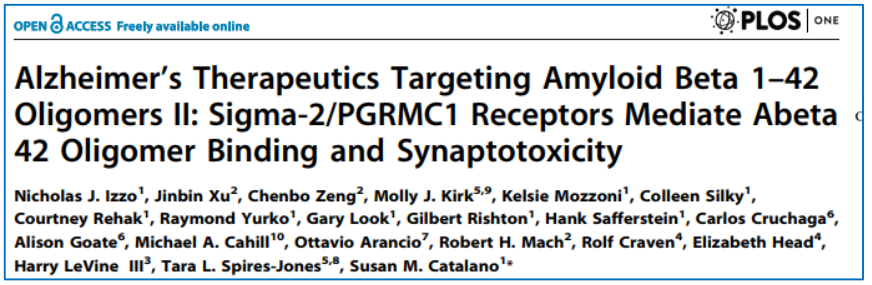
PGCM1：progesterone receptor membrane 1


Journal of Neurochemistry JNC

JOURNAL OF NEUROCHEMISTRY | 2017 | 140 | 561–575 doi: 10.1111/jnc.13917

ORIGINAL ARTICLE

Small molecule modulator of sigma 2 receptor is neuroprotective and reduces cognitive deficits and neuroinflammation in experimental models of Alzheimer's disease

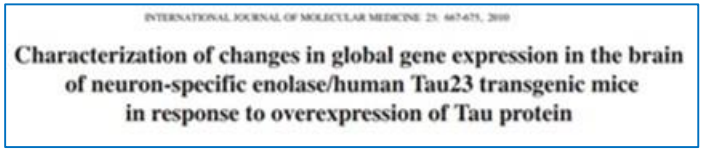

OPEN ACCESS Freely available online PLOS ONE

Alzheimer's Therapeutics Targeting Amyloid Beta 1–42 Oligomers II: Sigma-2/PGRMC1 Receptors Mediate Abeta 42 Oligomer Binding and Synaptotoxicity

Nicholas J. Izzo[1], Jinbin Xu[2], Chenbo Zeng[2], Molly J. Kirk[5,9], Kelsie Mozzoni[1], Colleen Silky[1], Courtney Rehak[1], Raymond Yurko[1], Gary Look[1], Gilbert Rishton[1], Hank Safferstein[1], Carlos Cruchaga[6], Alison Goate[6], Michael A. Cahill[10], Ottavio Arancio[7], Robert H. Mach[2], Rolf Craven[4], Elizabeth Head[4], Harry LeVine III[3], Tara L. Spires-Jones[5,8], Susan M. Catalano[1]*

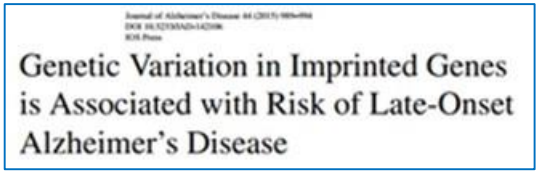
神経保護的効果（neuroprotective)認知不全・炎症に治療効果

GPM6A：Glycoprotein M6A

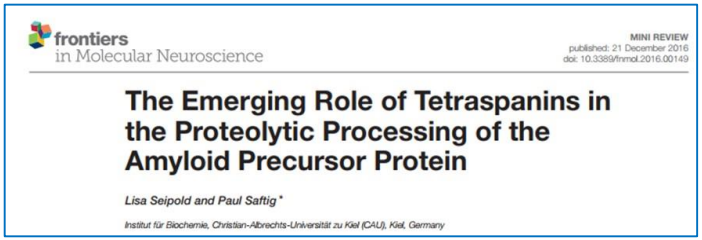

INTERNATIONAL JOURNAL OF MOLECULAR MEDICINE 25: 667-675, 2010

Characterization of changes in global gene expression in the brain of neuron-specific enolase/human Tau23 transgenic mice in response to overexpression of Tau protein

DLGAP2：DLG-Associated Protein 2


Genetic Variation in Imprinted Genes is Associated with Risk of Late-Onset Alzheimer's Disease

CD81:Tetraspanins family


frontiers in Molecular Neuroscience

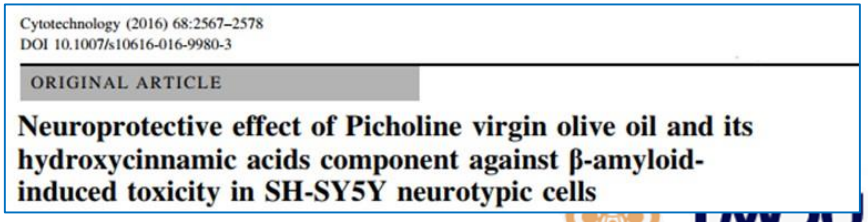
MINI REVIEW published: 21 December 2016 doi: 10.3389/fnmol.2016.00149

The Emerging Role of Tetraspanins in the Proteolytic Processing of the Amyloid Precursor Protein

Lisa Seipold and Paul Saftig*

Institut für Biochemie, Christian-Albrechts-Universität zu Kiel (CAU), Kiel, Germany

PFKM: Phospofructokinase

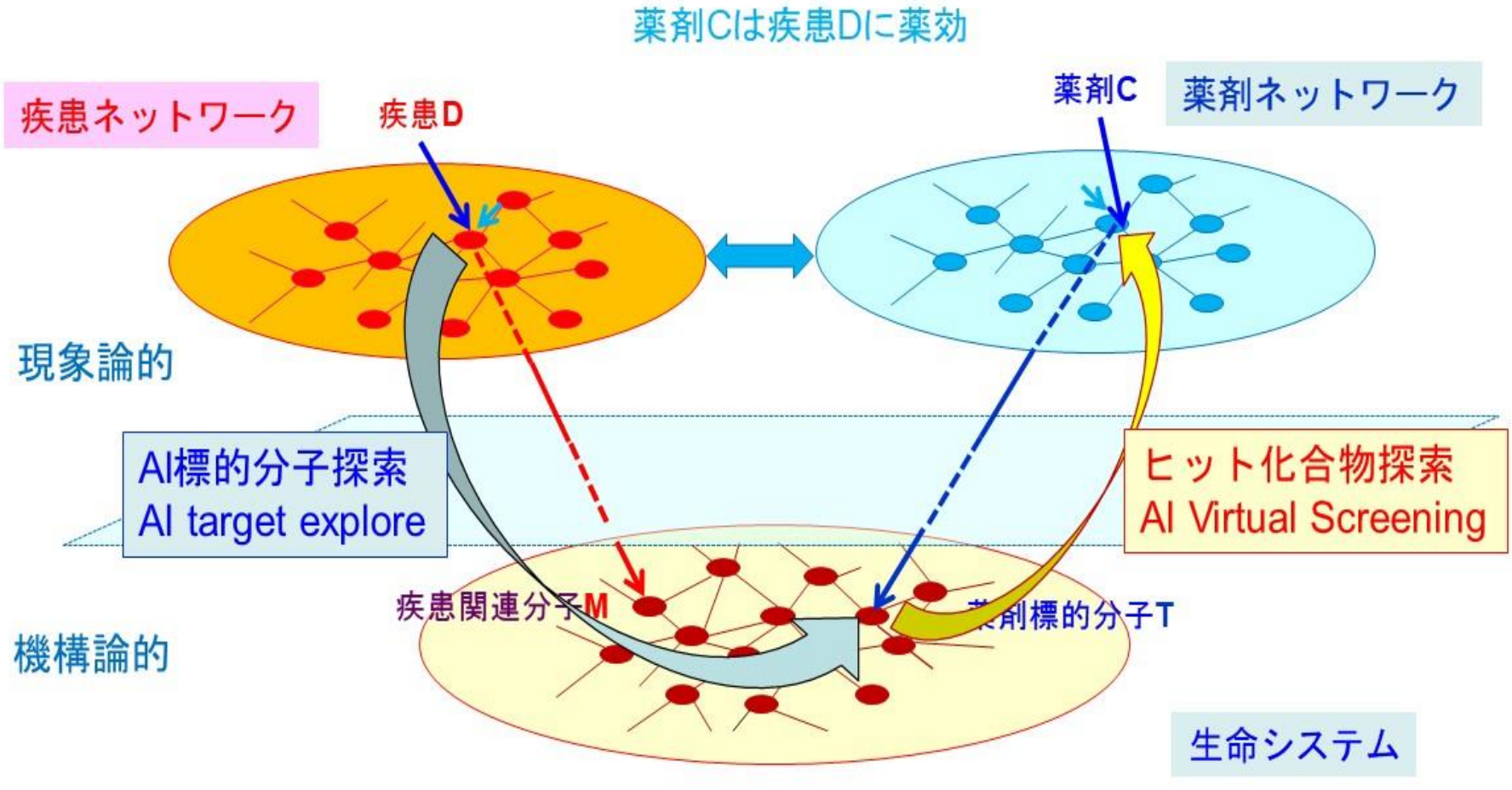

Cytotechnology (2016) 68:2567–2578
DOI 10.1007/s10616-016-9980-3

ORIGINAL ARTICLE

Neuroprotective effect of Picholine virgin olive oil and its hydroxycinnamic acids component against β-amyloid-induced toxicity in SH-SY5Y neurotypic cells

AI創薬の実現

3層生体・薬剤ネットワークによるAI創薬の過程



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

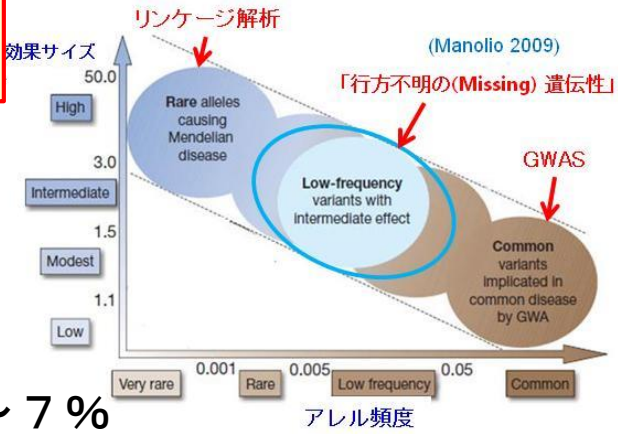
多因子疾患の機序/発症予測は無着手である

- 「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 x G
  - 環境と遺伝子の相互作用が：G x E

SNPの相対リスク
低い(1.1～1.3)理由
GxE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



統合失調症（50～70%遺伝）しかし108個のSNP合計で５～７％

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組合せ特異的効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)

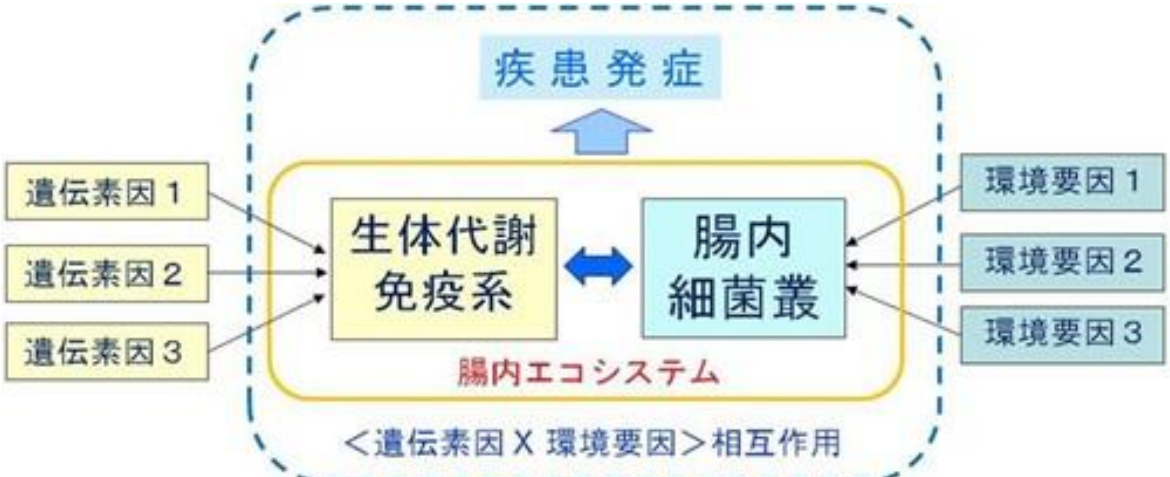
- オランダ飢饉
  - 第2次大戦末期、ナチスの封鎖、約半年間酷い飢饉
  - 飢饉の期間に胎児、戦後30年
  - 成人期:肥満,糖尿病,心筋梗塞,統合失調
- Baker仮説：英国心筋梗塞増加
- エピジェネティク機構
  - 過度な低栄養：肝臓のPPARα/γ（倹約遺伝子）メチル化低下・遺伝子発現がオン
  - エピジェネティック変化は**可変**：短期的変化、**長期的「記憶」**次の世代も

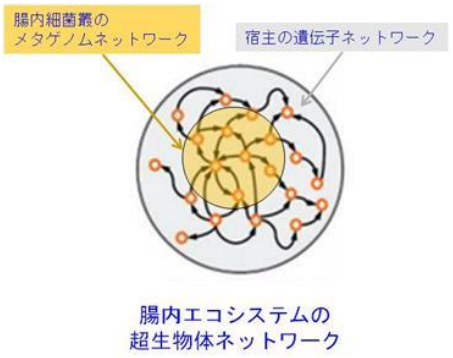
オランダ
飢饉 (1944)

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気的に代謝、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

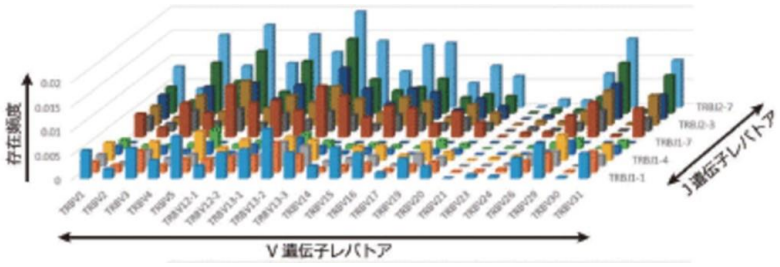




メタゲノム
超生物ネットワーク

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

TMDU

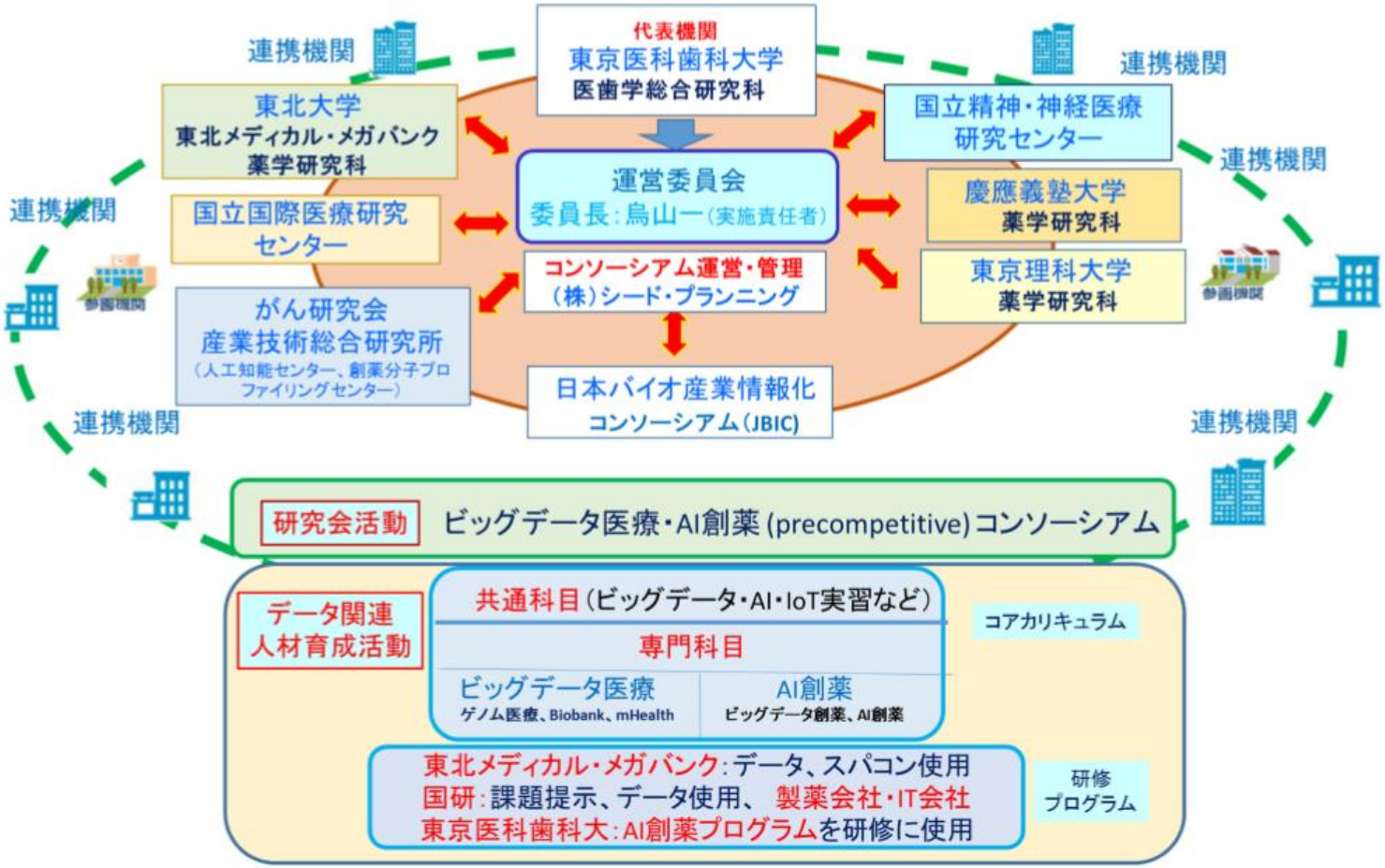
今後の医学＜知＞の展望

- 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

TMDU

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

TMDU

ご清聴有難うございます