

# 生体ビッグデータとPrecision Medicine

東京医科歯科大学 名誉教授・特任教授  
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

田中 博

# 略歴

- 1974年 東京大学工学部計数工学科 卒業
- 1981年 東京大学医学系大学院博士課程修了 医学博士
- 1982年 東京大学 医学部 講師
- 1983年 東京大学工学系大学院より 工学博士
- 1982～1983年 スウェーデン ウプサラ・リンシェーピング大学客員研究員
- 1987年 浜松医科大学 医学部附属病院 医療情報部 助教授
- 1990年 米国マサチューセッツ工科大学 客員研究員
- 1991年 東京医科歯科大学 難治疾患研究所 生命情報学 教授
- 1995年～2009年 東京医科歯科大学 情報医科学センター センター長 併任
- 2003年～ 東京医科歯科大学大学院 疾患生命科学部教授へ異動
- 2006年～2010年 東京医科歯科大学大学院 生命情報科学教育部 教育部長・大学評議員 併任
- 2015年 東京医科歯科大学 名誉教授、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
- 2017年 東京医科歯科大学 医療データ科学推進室 室長・特任教授

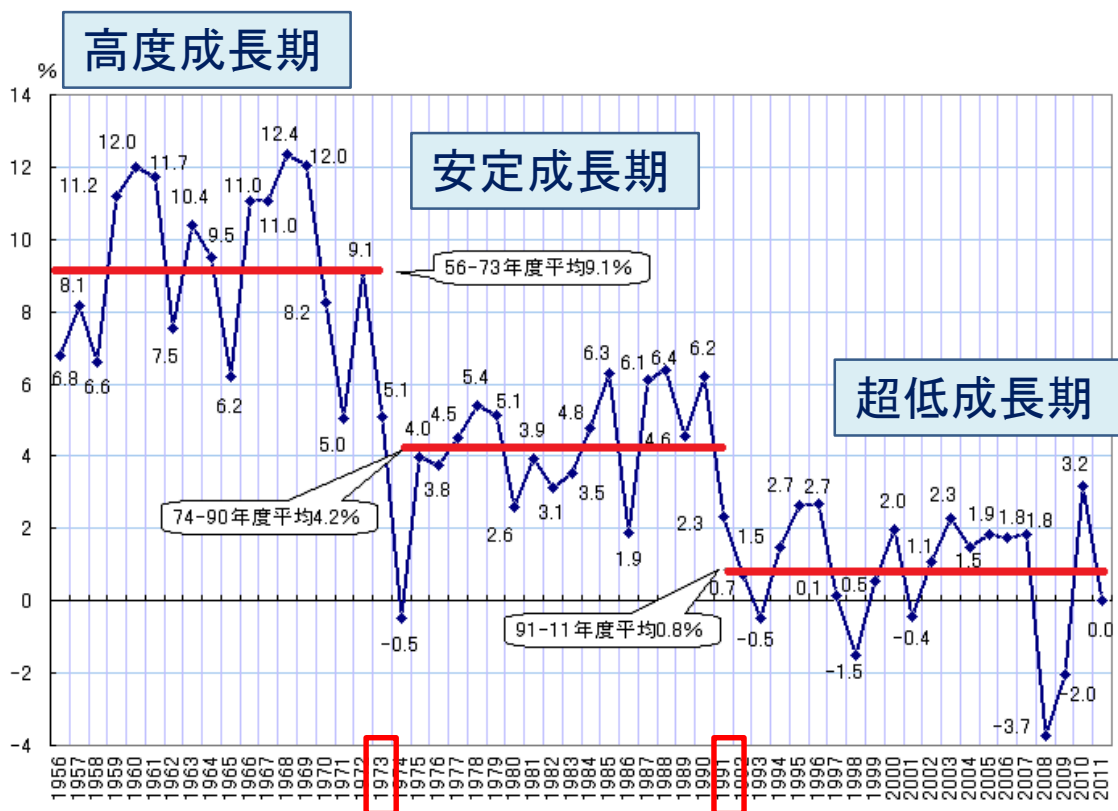
# 我が国の医療・介護・福祉を 取り巻く状況と経緯

# 我が国の社会経済的な推移

## 超低成長(super-low growth)社会へ

戦後日本の  
経済社会世代

1. 終戦期  
(1945～1954)
2. 高度成長期  
(1954～1973)
3. 安定成長期  
(1973～1991)
4. 停滞期  
(1991～現在)



高度成長期  
経済成長率約10%期

安定成長期  
経済成長率約4%期

超低成長期  
経済成長率約0～1%期

# これまでの医療・医療政策の体制の変遷

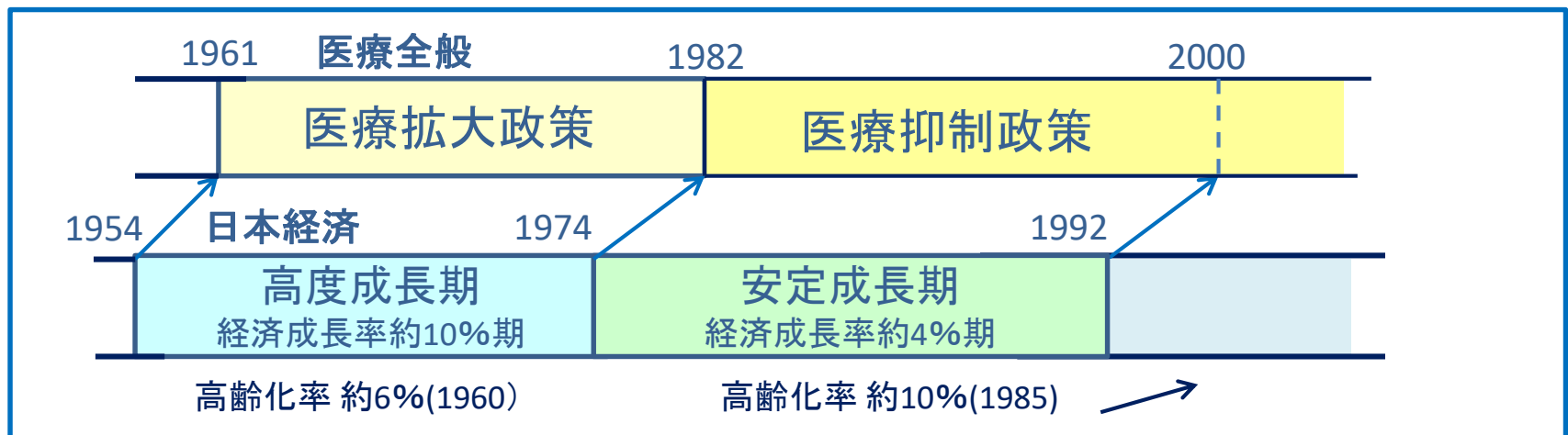
社会経済的推移と7～8年遅れて推移

## 高度成長期医療政策(医療拡大政策:1961～1982年)

- 国民皆保険制度導入から、1982まで。病床数2倍
- 老人医療の無料化(1972-1982)、新設医大(1972-1979)

## 抑制期医療政策(医療抑制政策:1982～2010年まで)

- 医療費・医師養成数抑制政策、「医療費亡国論」(1983)
- 後半は「崩壊期」と重複



# 従来の日本型医療体制の確立と 成長期社会

## 成長型若齡社会

高度成長期・安定成長期(54~91)

「成長期にある若い人中心の国」\*1

急性期疾患中心の医療

「病院完結型医療」の  
無関連な集まり



## 日本型医療体制\*2

フリーアクセス制、自由開業制(民間病院多在)

現物(医療)給付・出来高払い制

\*1 80年代 高齢化10%以下 1970 7% →1994 14%

\*2 国民会議「1970年代モデル」 1985年平均寿命世界一

# 超低成長期における医療の崩壊と再生

1961 国民皆保険制

高度成長・安定成長期

1991

日本型医療  
「病院完結型医療」の無関連な集まり

超低成長期(停滞期)

2005~

地域医療の崩壊

医療費抑制

小泉政権診療報酬低減

医師不足

新臨床研修医制度  
医師数抑制政策  
戦中戦後医師退職

超高齢化

慢性疾患増大

医療再生へ

# 地域医療の崩壊

## ■ 絶対的医師不足とその加速

- ✓ 長年に亘る「医師数抑制政策」(1986～98, 08)による絶対的な医師不足
- ✓ 戦中・戦後育成医師世代の高齢化による退職
- ✓ 「新医師臨床研修制度」(2004～): 都市集中, 診療科偏在

## ■ 医療費の長期にわたる抑制政策

- ✓ 小泉政権下(2001/6)の医療費削減政策(OECD24位)

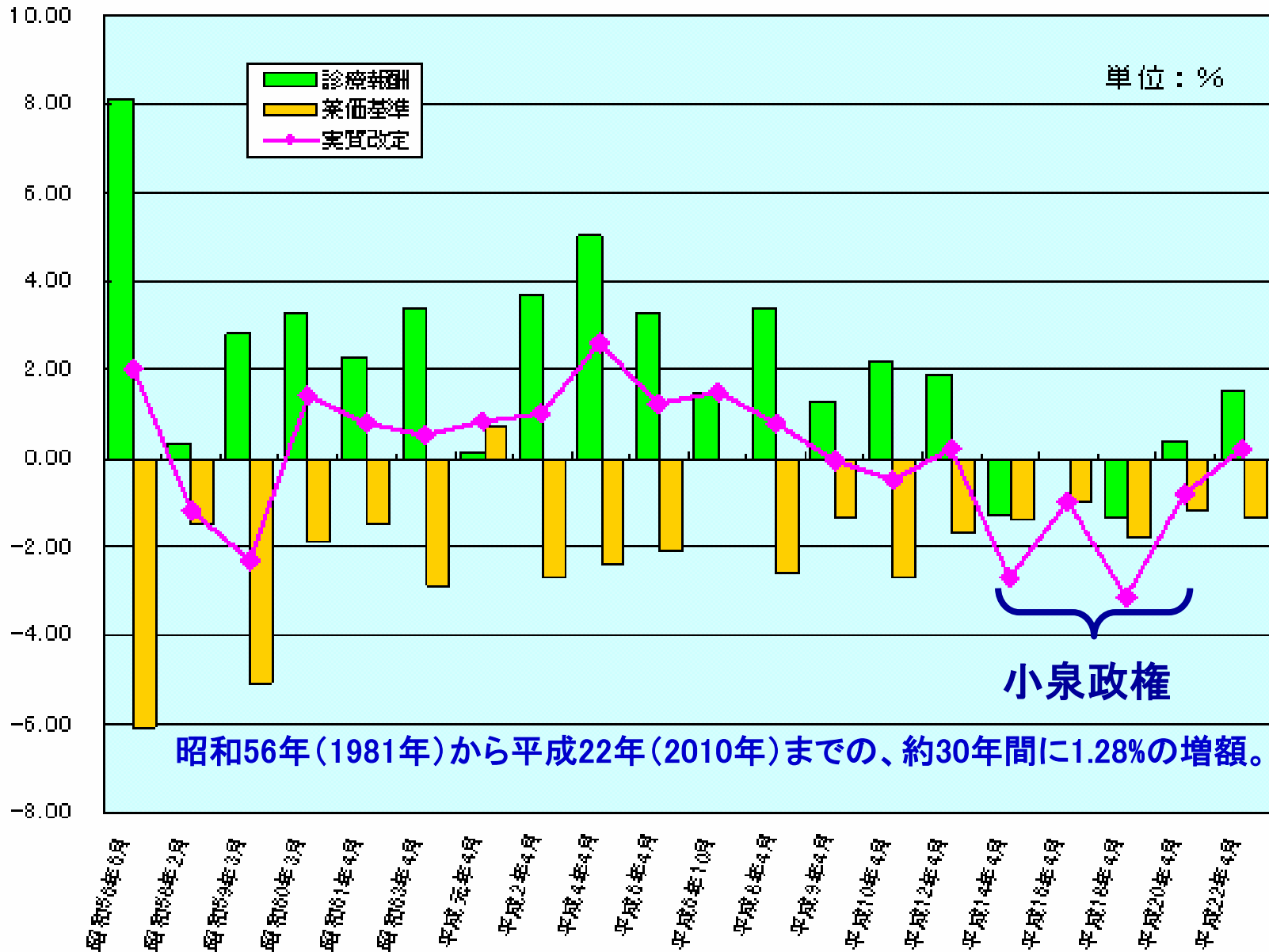
## ■ 国民の医療権利意識の増大

- ✓ 医療過誤訴訟の増加 産科/小児科/外科医の減少

病院勤務医の負担増大と負のスパイラル  
立去り型サボタージュ / 病院勤務医から開業医へ

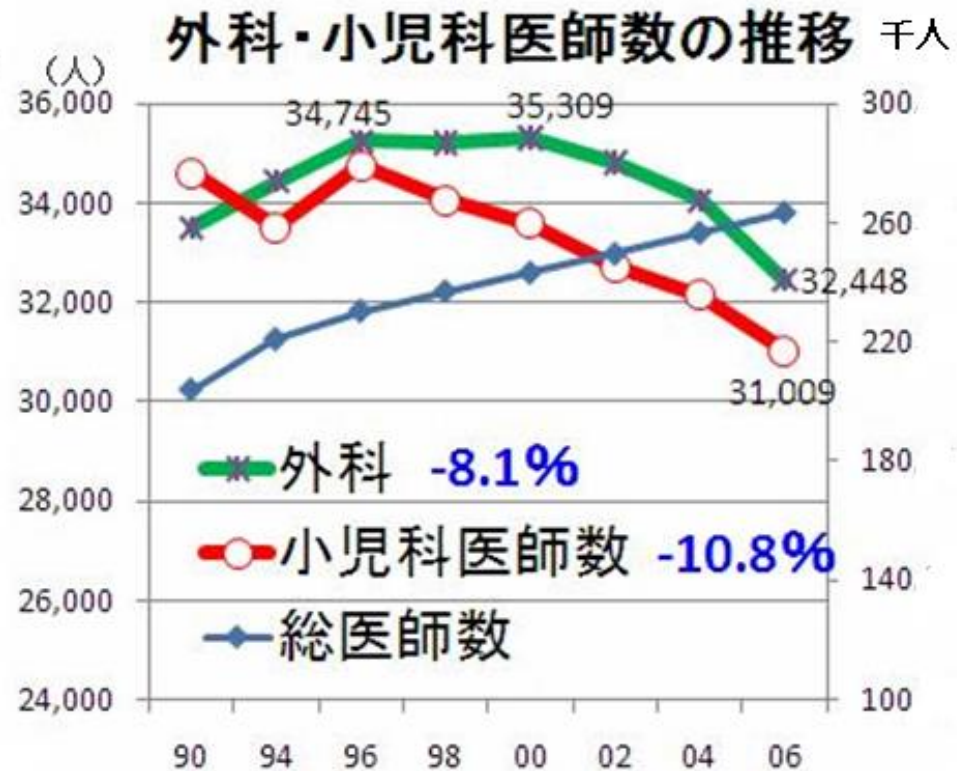
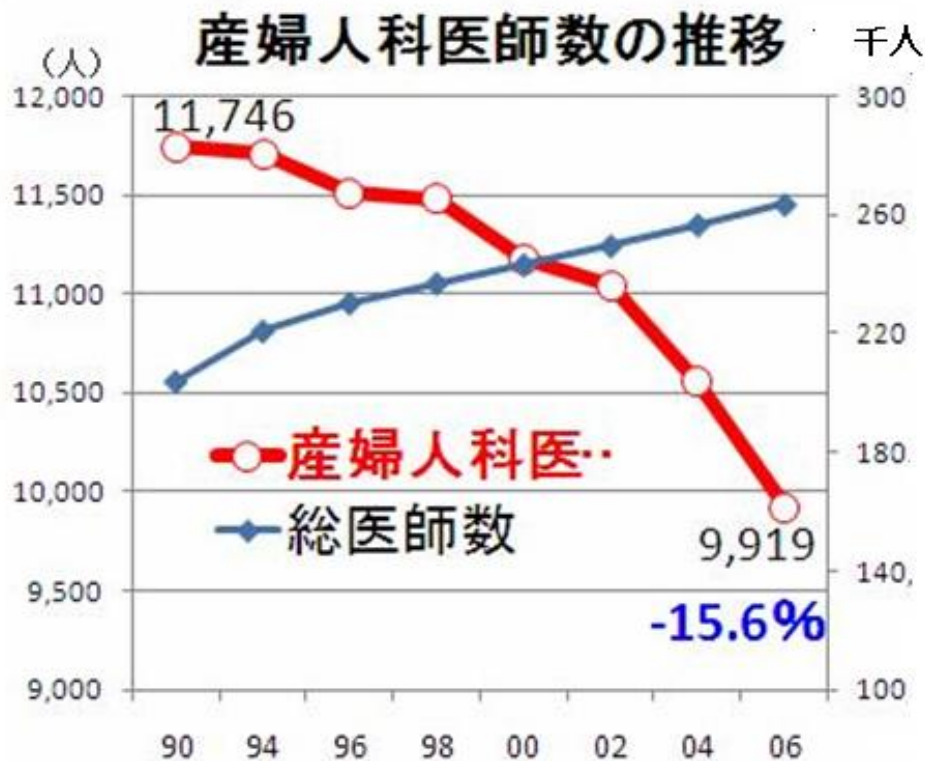


# 診療報酬の抑制



# 医師数抑制政策

医師不足を診療科別に見た場合、それが顕著なのは産婦人科である。  
1990年から2006年までの16年間のあいだに、総医師数が増加しているなかで産婦人科の  
医師数はピークからの減少幅が15.6%にも達している。  
次いで多いのが、小児科10.8%、外科8.1%の順である。



# 高齢化がもたらす疾病構造の変化

## ■ 高齢化に伴う慢性疾患の増大

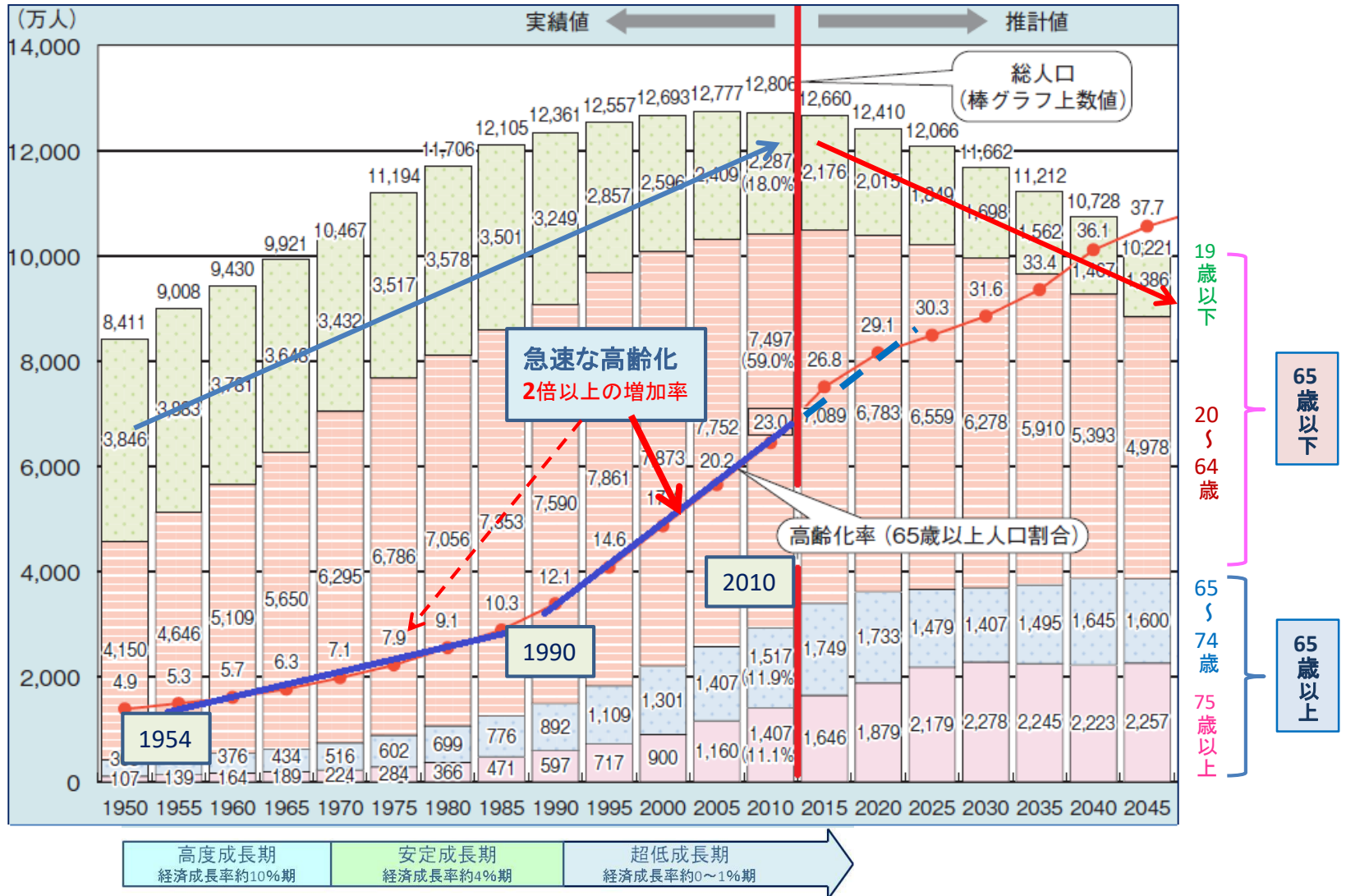
- ✓ 国民医療費に占める65歳以上の医療負担50%超
- ✓ 慢性疾患患者の増加とその負担
- ✓ 糖尿病の増大と重症化患者の加速増加
  - 人工透析患者増加による国の負担増 30万人
  - 国の負担 1兆5000億円
  - 人口当たりの人工透析患者が世界一多い
- ✓ 脳卒中:急性期後の長い維持期・再発による重症化
  - 維持期における長期にわたる疾病管理
  - 高い再発率、寝たきり 在宅患者の約4割・医療費2兆円

## ■ 治療から疾病管理への転換

「急性期疾患」から「慢性疾患」へ

従来の急性期医療に主眼を置いた医療制度だけでは  
支えきれない社会構造になった

# 少子超高齢化社会へ

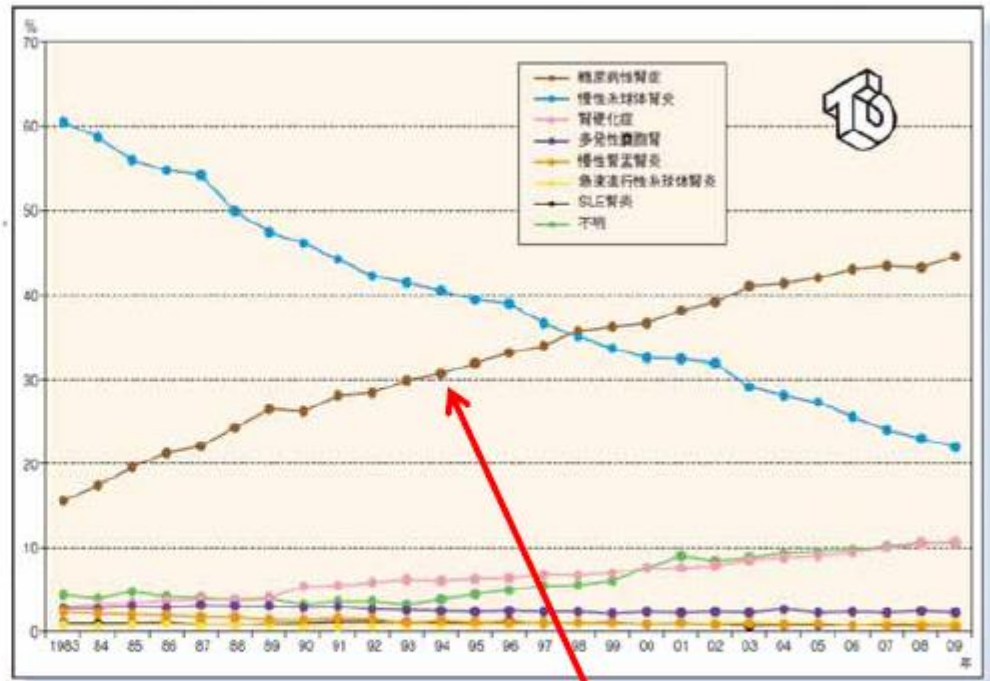
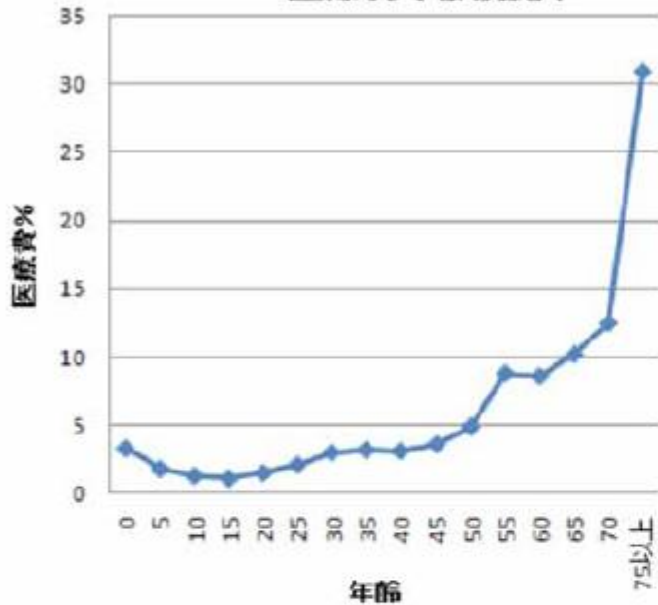


高齢化率約6%(1960) 高齢化率約10%(1985) ↗ 高齢化率23%(2010) 30～40%(2025以降)

# 慢性疾患患者の増加

## 医療費の年齢別比率と人工透析原因疾患

医療費年齢別比率



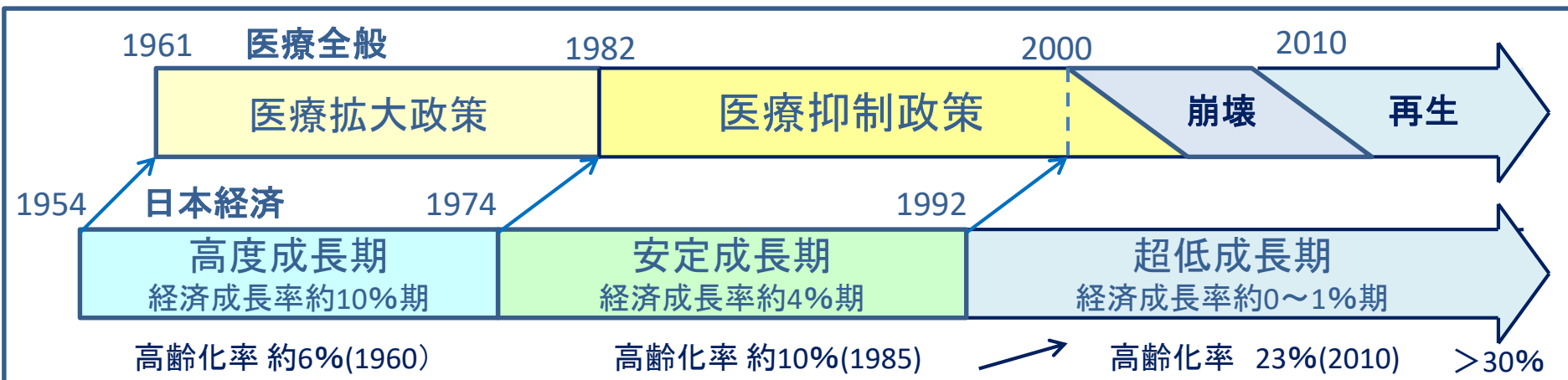
国民医療費:  
65歳以上が  
50%以上

日本透析医学会  
統計調査委員会  
2009資料より

糖尿病性腎症  
2009年 44.5%

# 医療・医療政策の体制の変遷

- **医療拡大政策**(1961年から1982年)
  - 国民皆保険制度の導入から1982まで続いた高度成長期の医療
  - 老人医療の無料化(1972-1982)、新設医大(1972-1979)病床数2倍
- **医療抑制政策**(1982年から2010年まで)
  - 1982から始まった医療費・医師養成数抑制政策の医療
  - 「医療費亡国論」(1983)、2000年初頭小泉政権下の激烈な診療報酬削減、
- **医療崩壊**(2005年頃から現在も)
  - 2000年代に顕現した絶対的医師不足と地域医療の崩壊
  - 2010年から再生期と重複、2006年第5次医療制度改革(地域医療連携)
- **医療再生政策**(2010年から始まる)
  - 2009年の政権交代とともに医療費抑制政策が一旦停止
  - 地域医療再生(基金09,10年)と勤務医負担軽減(10,12診療報酬改定)
  - 2025年、超高齢化社会の到来への対応を模索。



# 病院完結型医療



# 持続可能な新たな医療・ケア体制へ

国民皆保険1961  
高度成長期医療

医療費抑制政策1982~2009  
とくに小泉改革等 2001~09

新設医大設置  
1970~1979

医師数抑制  
1986~98

戦中戦後養成医師  
退職2000代から

医療費・医師数抑制

疾患構造変化

新研修医制度  
2004~

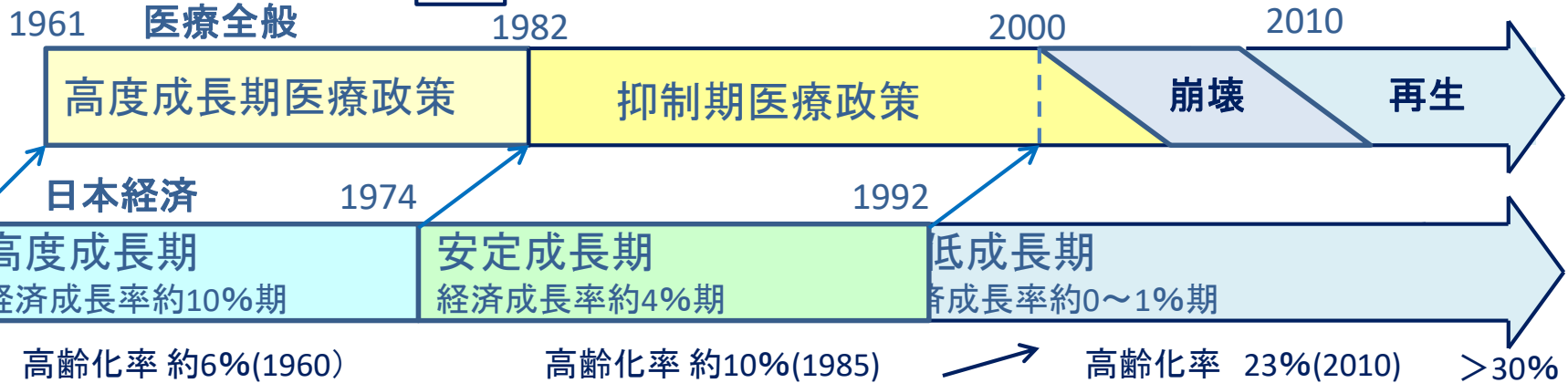
医療訴訟増加  
産科・小児科医減少

超高齢化社会の到来  
慢性疾患治療長期化

日本型医療の破綻

再生過渡期

2025年問題







# 我が国が直面する医療・ケアの課題

## ■ 地域医療の崩壊

社会経済的要因（超低成長・医療政策など）

疾患構造的要因（慢性疾患増加・維持期など）

## ■ 介護福祉の崩壊の到来

高齢者の将来の急激な増大（2025年問題）

# 我が国の医療・福祉の再生

我が国はこの現状に対応して  
どのようにして  
持続可能な医療・ケア体制をつくるのか

# 「地域で連携する」医療・ケアへ

- 「病院完結型医療」から「地域連携（完結）型医療」へ
  - 医師不足/慢性疾患増加 希少医療資源の共有
  - 「急性期病院の無関連な集合」では対応できない
  - 「地域医療」が単位、「病院」「診療所」は要素
  - 総合病院指向ではなく専門病院化・分担機能化
  - 地域医療崩壊を防ぐ病院・診療所の地域連携
  - 地域へ拡張する診療連携パス（病院負荷集中防止）



# 「生涯を通じた」医療・ケア

- 「生涯継続性」 lifelong-continuity, life-log
  - 急性期治療中心から生涯継続的な健康/疾病管理への転換
    - 慢性疾患増大/超高齢化社会での医療の役目
    - カルテの保存期間に見られる急性期中心主義
    - 国民の疾病観の転換の必要性
    - 「生涯健康医療記録」(EHR)の不可欠性
- 重症化・再発予防と医療費の適正化
  - むしろ重症化の予防(悪化抑制)→医療費の適正化
  - 「治す」から「重症化させない」疾患管理へ
  - 糖尿病・脳卒中の長い維持期(再発予防)「支える医療」



生涯継続的な健康/疾病管理

# 日常生活圏域を基点としたケア

- 「地域（日常生活圏）包括ケア」
- 在宅医療の基点化 *daily life basis*
  - 維持期医療の在宅化
    - 在宅療養の重点化・往診の重装備化
  - 医療と介護のシームレスな連携
    - 退院病院と往診医・訪問介護との連携
      - 退院時コンファレンス等
- 日常生活圏での健康/疾患管理
  - 慢性疾患管理と自己測定→「情報薬」
    - 生理計測モニターなどモバイル医療(mHealth)
  - 健康リスク管理と日常測定—健康長寿政策
  - 地域医療連携との連携



施設中心医療から生活圏中心ケアへ

# わが国の健康医療ケア戦略の パラダイム変換

医療政策からは「連携医療・包括ケア」政策であるが、患者・生活者サイドからは

## 「日常生活圏」での健康/疾患管理

- 健康リスク管理と生理変量モニター(mHealth)
- 疾患管理としての「情報薬」
- 在宅医療・地域包括ケア

## 「生涯にわたる」健康/疾病管理

- 疾患罹患可能性に対する「生涯的」perspective
- life-long healthcare
- 生涯型EHRの国規模での実験

# これからの健康医療戦略

この2つの概念をまとめると

医療施設に全面的に頼らない

「患者参加型医療 (participatory medicine)」

## 自らの健康は自らが守る

- 定量化自己(Quantified Self)
  - 自らの健康状態を把握する
- Intelligent Consumer:
  - 医療情報・知識を自ら収集する
- 先制医療:
  - 継続的な自己健康状態把握に基づき疾患状態への過渡を認識して、疾患前に治療する
- 3次予防:
  - 疾患に罹患しても重症化しないように疾患管理に努める

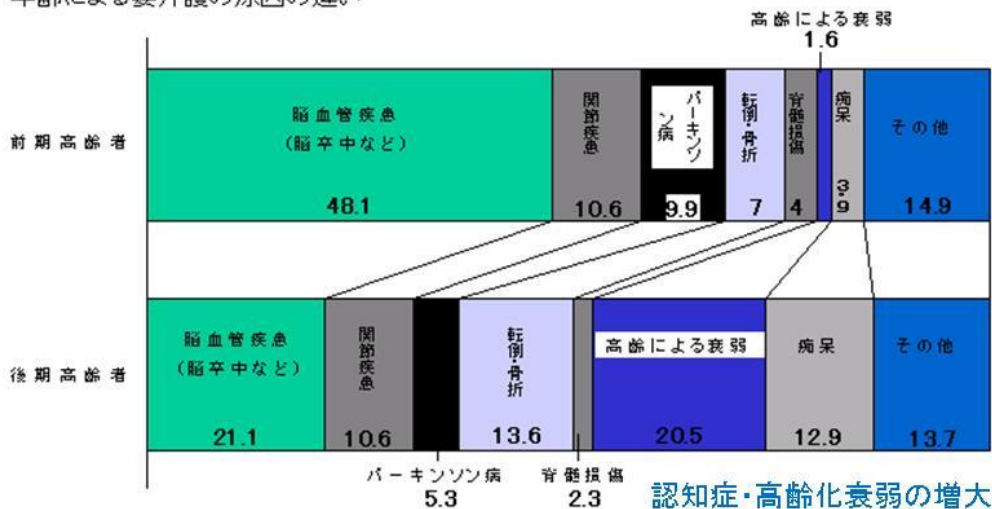
「生涯にわたる日常生活圏における健康・疾病管理」

# 介護原因疾患から知る高齢化 による罹患可能性の高い疾患の時系列

- 高齡化に従い主要罹患疾患推移 **がん→脳卒中→認知症**
- 要介護原因疾患：男性 **脳卒中→認知症**
- 要介護原因疾患：女性 **関節・骨折(ロコモティブ症候群)→認知症**



年齢による要介護の原因の違い





# 生涯的Perspectiveへの 健康・疾病概念のパラダイム転換

- <若い人中心－急性期疾患中心>の疾病概念から
- 慢性期疾患中心－「生涯にわたる」健康疾病概念へ  
Life-long (course) healthcareの概念
- 医療施設中心(施設医療)→  
患者中心・日常生活圏・患者参加型(participatory medicine)

Reactive (対応的)で  
Occasional (機会的)な  
医療・ケア



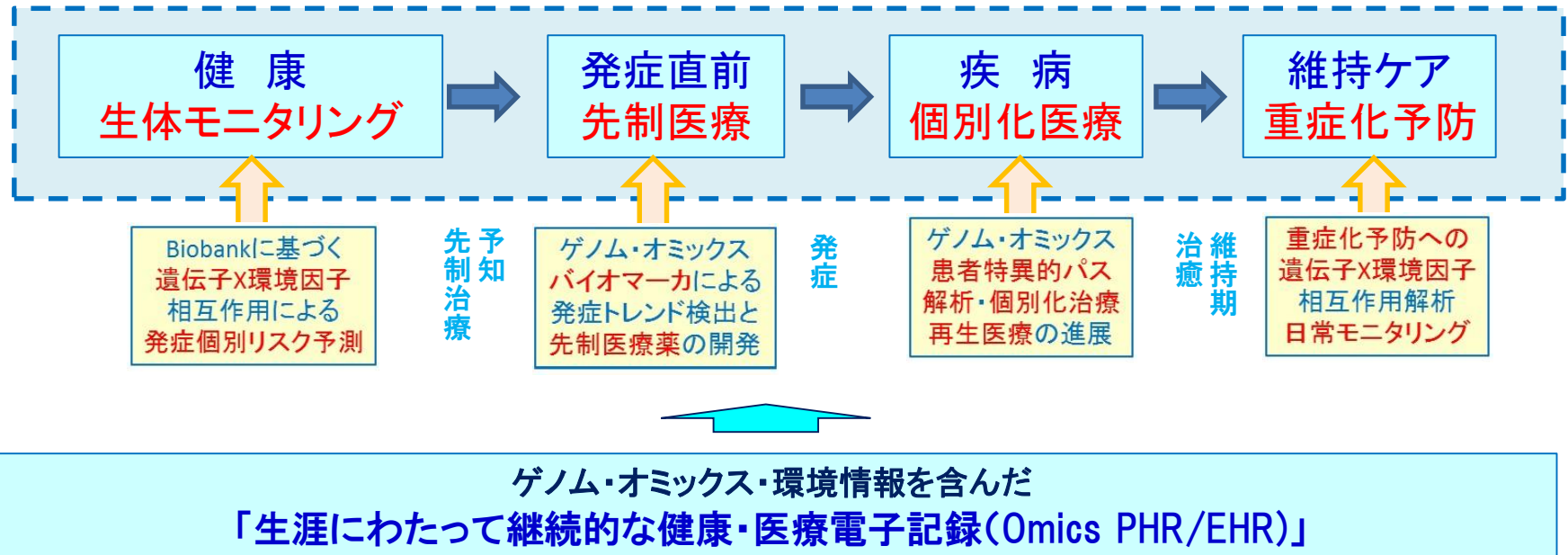
Proactive (予見的)で  
Life-long (生涯的)な  
医療・ケア

生涯型医療・先制医療への転換

# ライフコースと医療

これまでの認識をまとめると健康医療ライフコースは以下ようになる

## life-long healthcare



# ライフコースの予測 に基づいた健康医療戦略

- 慢性疾患の増加
  - 超高齢化による慢性疾患（一般のがん、脳卒中、認知症）の重要性の増加。この順で発症頻度の年齢分布ピークが継起する。
- 健康モニタリング
  - 日常的生体モニタリングによる健康状態観測による発症前管理。疾病へのトレンドの予測
- 発症リスク予測
  - ゲノム・オミックス医療の進展による発症リスクの（発症前）個別化予測の確度の上昇
  - <遺伝的素因と環境・生活習慣要因>との相互作用の解明
- オミックス・バイオマーカの進展
  - バイオマーカによる発症トレンド認識の上昇・先制医療の実現
- 疾病の個別化治療
  - ゲノム・オミックス医療の進展による発症罹患後の個別化医療・予後予測医療の実施
  - <患者特異的ネットワーク調節不全>の同定を基礎にシステム医学
- オミックス・バイオマーカによる3次予防
  - 回復・維持期における再発・重症化予防オミックス日常モニタリング

生涯にわたって継続的に電子記録するOmics EHR/PHR

# 継続的モニタリングの階層

## 第1種 表現型情報による生涯継続的な生理量モニタリング

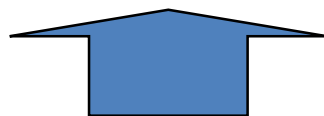
- Quantified Self (そとからの生体モニタリング)
  - モバイル・ウェアラブル生理モニタリング
  - 参加型医療・proactive professional consumer

## 第2種 分子情報も含めたパンオミックス・バイオマーカの継続的モニタ(内からの生体モニタリング)

- 疾患発症を予知するオミックスバイオマーカを継続的に計測
- SnyderらのiPOP (integrated Personal Omics Profiling)概念
- 疾患予知オミックスは先制医療・先制医療薬へ
- 液性バイオマーカ(Liquid Biomarker)
- 分子システム医学的なマーカ
  - Dynamic Network Biomarker (L.Chen)

# mHealth/ healthcare IoTの概念

「どこでも」「いつでも」最良の医療  
健康管理を受けられる情報環境



Wireless communication 技術に基づいた  
診療現場や日常生活圏における  
健康・医療の知能化環境

モバイル健康医療社会の実現

# モバイル・ユビキタス, IoT情報の概念

Ubiquitous, IoT(Internet of Things)

ユビキタス:「(神のごとく)遍在する」(Xerox社MarkWeiser1986)  
大型コンピュータやパソコンと違った  
第3のコンピュータ利用形態

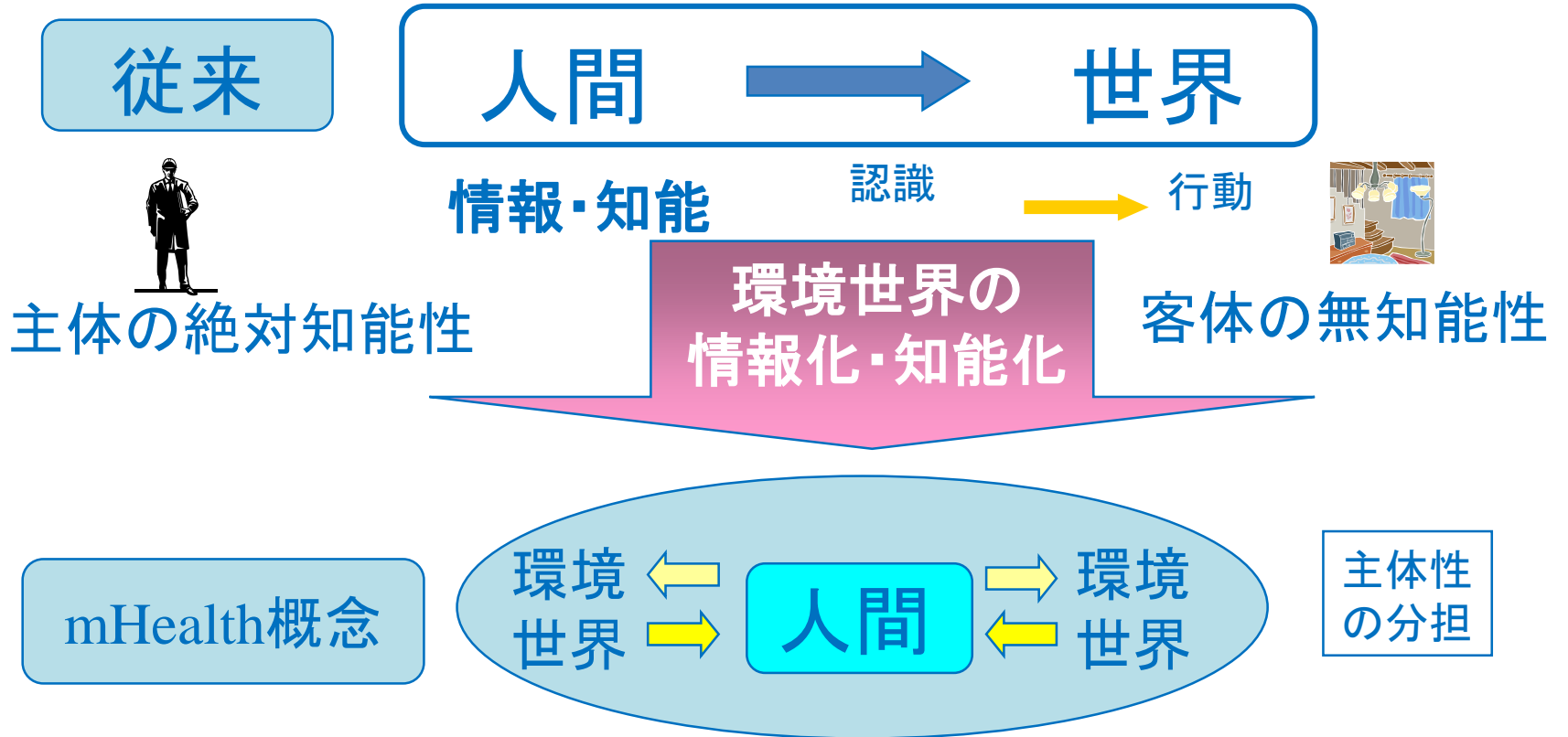
生活環境の中にコンピュータチップ・計測装置と  
ネットワークが組み込まれ、ユーザーはその場所  
や存在を意識することなく利用できる情報環境

- 見えない (invisible)
- 区別がつかないほど日常生活に織り込まれている
- しかし、観測され知的アドバイスを提供する

Network/Intelligenceの環境世界化

# モバイル健康医療(mHealth)の概念

## 環境世界と人間の関わり方の変革



## 機械知能と人間とのSymbiosis

Conscious-freeな生活 確実・安心・安全な社会

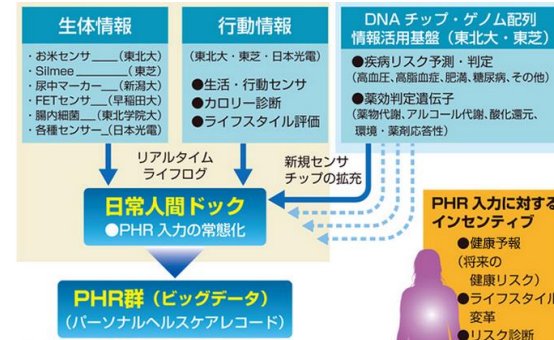
# 第1段階 生体生理量センサリング

- **日常生体モニタリング**
  - 生理変量(血圧、心拍、運動、酸素飽和度、血糖値)などの**表現型**情報の測定
- **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
- **東北大学-東芝COI**
  - 「さりげないセンシングと日常人間ドックで実現する理想自己」
- **そのほか Wearable sensor**
  - コンティニューアなど幾つかmHealthのプロジェクト
- **生涯型電子カルテ(EHR/PHR)**
  - EHR: 医療施設の診療記録の共有
  - PHR: 自己入力のアレルギーなどのリスク情報を追加



ECG; EEG; Skin Conductivity; EVG

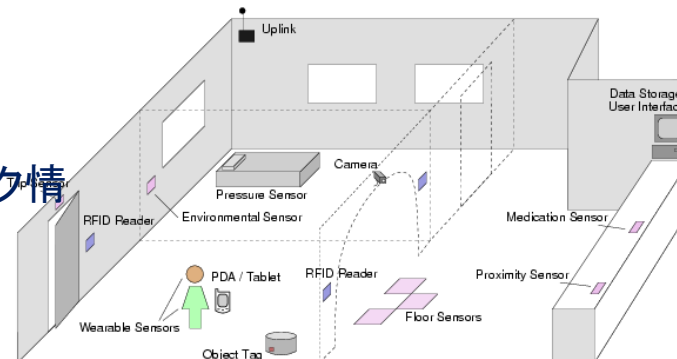
## 東北大学COI「さりげないセンシング」



Apple watch



Sony 健康時計





# Dr. John Halamka

Professor at Harvard Medical School

- PGP#2

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



# 第2段階 分子情報による日常 モニターリング

- パンオミックス・バイオマーカなどの分子情報の継続的モニターリング
- 疾患発症を予知するオミックスバイオマーカを継続的に計測
- SnyderらのiPOP (integrated Personal Omics Profiling)概念
- 液性バイオマーカ (Liquid Biomarker)
- 分子システム医学的なマーカ
  - Dynamic Network Biomarker (L.Chen)

# ゲノム・オミックス医療の基軸概念

## 3世代区分

### 第1 (生得的)ゲノム医療

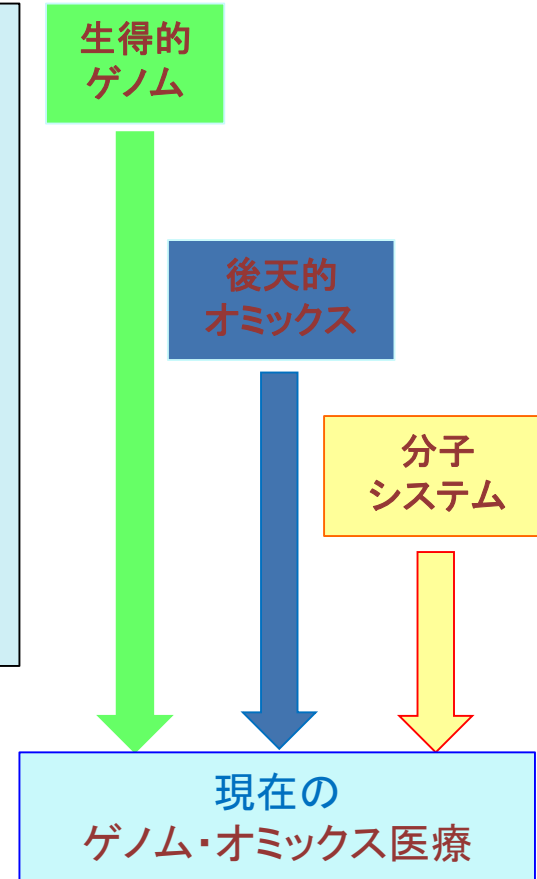
- 生得的ゲノム情報の変異・多型性

### 第2 (後天的)オミックス医療

- 後天的体細胞の網羅的分子プロフィール

### 第3 システム分子(分子システム)医学

- 細胞分子ネットワークの歪み/構造変化



H.Tanaka:

”Genome and Omics Medicine

—Principles, Clinical Implementation and BigData Approach—

Springer 2016

# 網羅的分子医学 第1パラダイム ゲノム医療(1990~現在)

Manolio (2009)

- 「生得的な(germ-line)」ゲノム変異や多型性に基づいた個別化医療
  - 生得的ゲノム
    - 全細胞で生涯を通じて同一99.5%は共通

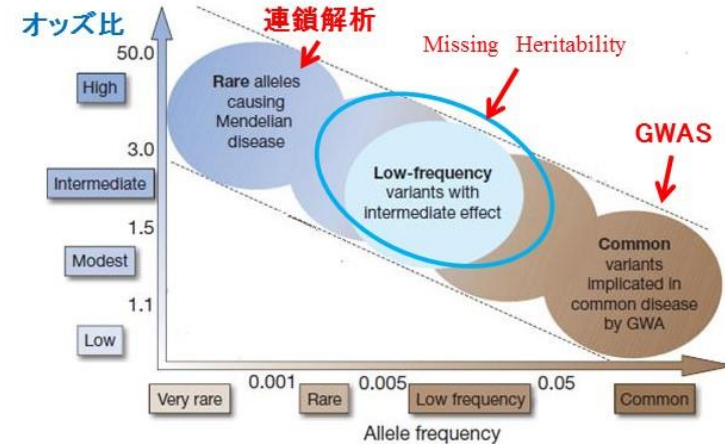
- 疾患原因遺伝子(Disease causative gene)

## 家系調査/Linkage解析

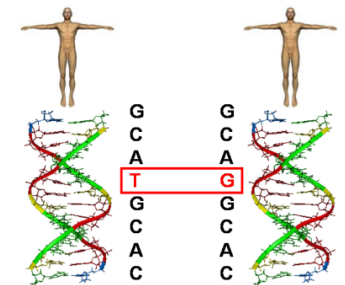
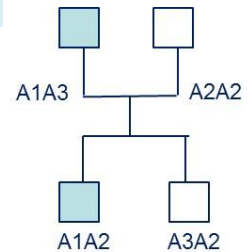
- 1980年代から: ハンチントン病
  - CAGリピート, HTT遺伝子, 99%浸透率
- その他にデシャンヌ型筋ジストロフィー, 嚢胞性線維症
- 当時400程度のDNAマーカヒトゲノム解読計画へ

- 疾患感受性遺伝子(Disease susceptibility gene)

- 多型性: 一塩基多型(SNP), 3000万位,
  - そのほかにマイクロサテライト、CNVなど
- 全ゲノム関連解析(GWAS)患者対照分析
- 薬剤代謝酵素の多型性(PGx)



単一分子へ起因する方法の限界



1塩基の変異が疾患や薬剤への応答性に関与する場合がある

# 網羅的分子医学 第2パラダイム オミックス医療(2000~)

- 「後天的・疾病依存的に変化する網羅的分子病態像」

後天的体細胞ゲノム変化+ゲノム以外の疾患オミックスに基づく医療

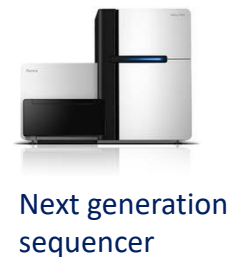
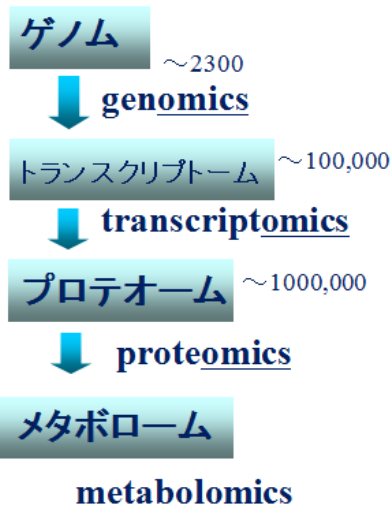
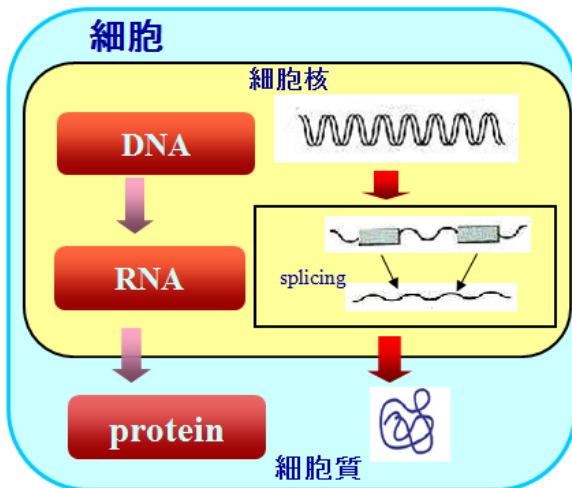
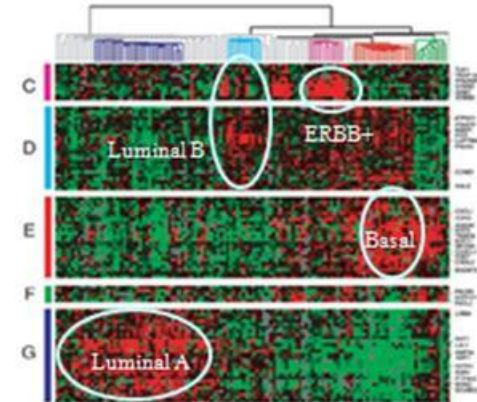
– 疾患 オミックスプロファイルによる分子的病態像

- 後天的ゲノム変異・遺伝子発現プロファイル・プロテオームなど、病態進行の段階に依存

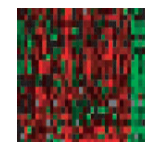
- 疾患の進行度の評価 → 予測医療・先制医療

- Ongoing state of disease progressionを表現
- 臨床症状や病理変化より早く変化 Predictive Omics

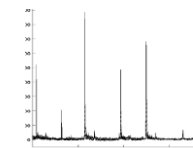
乳がんの亜型と治療・予後



Next generation sequencer



DNA microarray



mass spectrometry



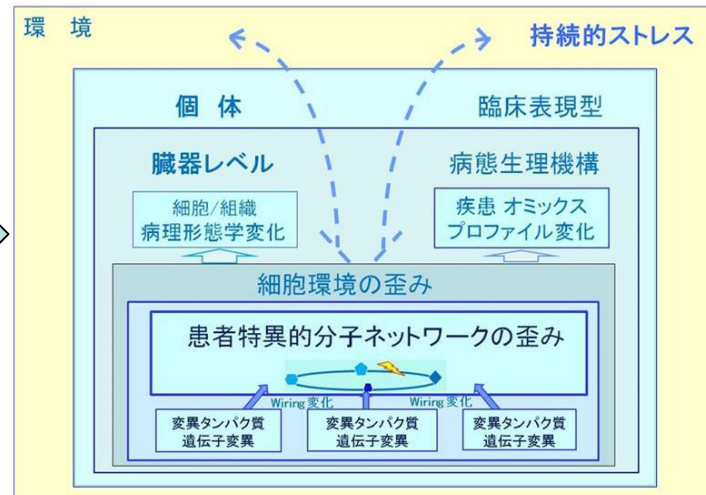
# 網羅的分子医学 第3パラダイム システム分子（分子システム）医学（2010～）

システム生物学の疾患への応用  
「疾患をシステムとして理解する」

単因子性の疾患を除いて、大半の疾患は1個や2個の遺伝子の変異ではなく多数の遺伝子の変異やタンパク質の機能異常による分子パスウェイ/ネットワークの調節機能不全や歪み  
**distortion (dysregulation) of molecular network**

個別化医療・予測医療・先制医療・創薬のための疾患のシステムの理解に基づく医療

**疾患オミックス**  
(網羅的分子表現型  
molecular phenome)  
成り立たせる基底としての  
「**細胞分子ネットワークの**  
(分子システム) **構造変化**」  
疾病の理解における  
第3のパラダイム



# ゲノム・オミックス医療の現状

バイオテクノロジーの  
急速な進展による

「ビッグデータ医療」時代の到来

# 医療・創薬への超大なインパクト ビッグデータ時代の到来

- (1) 次世代シーケンサ (Clinical Sequencing)を始めとする「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)



コストレスで良質なデータが大量に収集可能



治療医学の的確性の飛躍的進展: 「精密医療」  
医療の国民レベル・生涯ヘルスケアの進展



# ゲノム・オミックス医療の2つの流れ

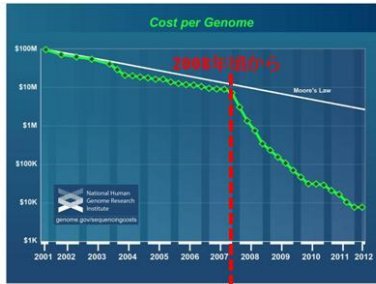
## 米国でのゲノム医療

- 「シーケンス革命」(2007)からの怒濤の展開(2010から)
- 個々の患者の「治療医学」レベル質的向上: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

## 欧州でのゲノム医療

- 「集合的遺伝情報」の価値⇒ゲノム・バイオバンクへ流れ
- 国民医療(医療の国民レベル)の向上:社会福祉国家の理念
- 「予防医学」レベル質的向上のためにゲノム情報導入
  - 大規模前向きpopulation型バイオバンク/ゲノム・コホートの確立
  - 遺伝的素因と環境要因(生活習慣)との相互作用に基づいた「多因子疾患」の発症予測を通じた「国民医療の向上」
  - 生涯的健康/疾病管理へ

# 米国ゲノム・オミックス医療の流れ



DNA Sequencing Cost: the National Human Genome Research Institute

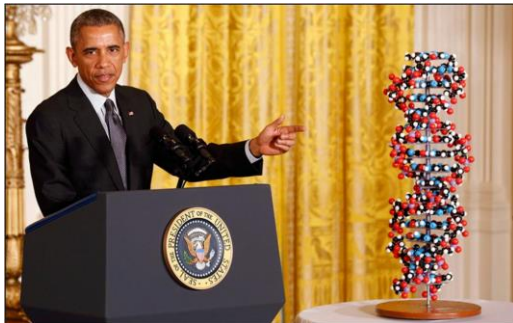
シーケンス革命 2007/8

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	ラビッドラン
解析時間	11日	27時間
リード長 (bp)	2 x 100	2 x 150
データ産出量 (Gb)	約600	約120
試薬コスト (ヒト1人全ゲノム)	数十万円	不可 エクスプロームのみ

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



オバマ大前大統領 Precision Medicine Initiativeを開始、2015年1月 大統領一般年頭教書演説

先陣争いの時代

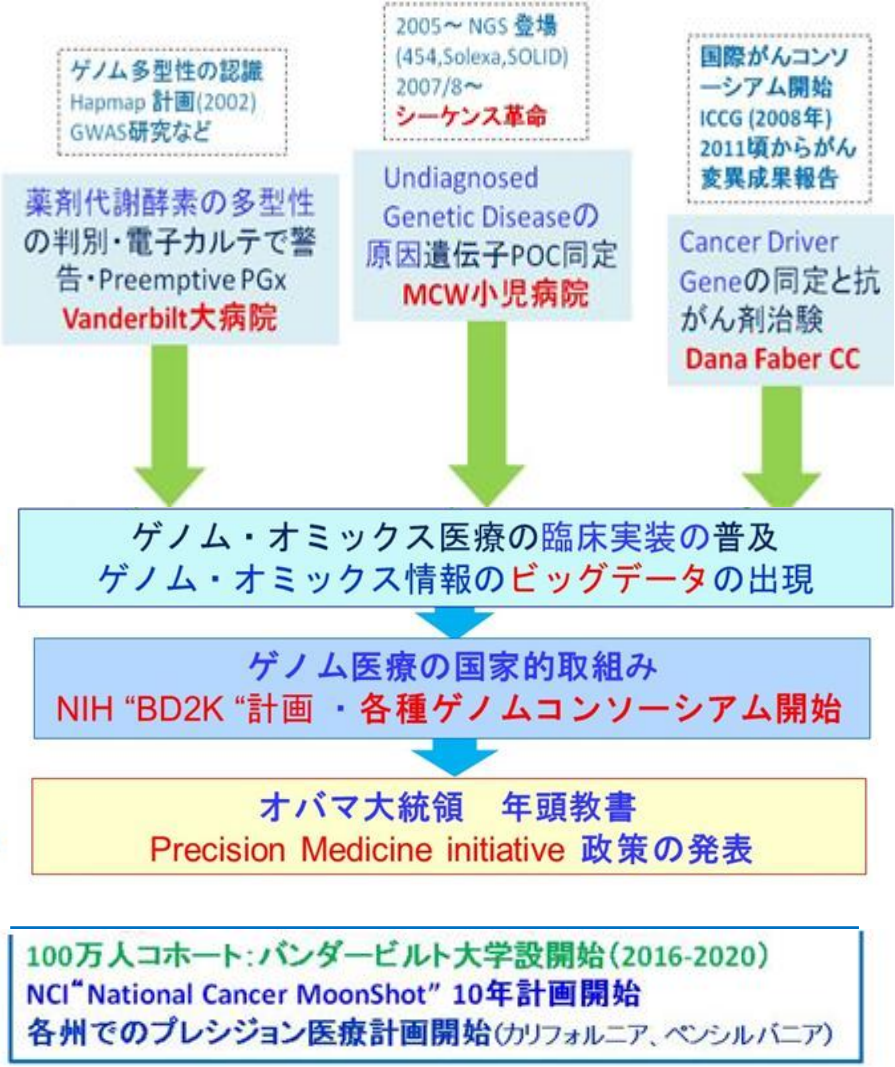
国家政策の時代

精密医療普及期

第一期

第二期

第三期



2007年  
2009年  
**2010年**  
2011年  
2012年  
2013年  
2014年  
2015年  
2016年  
2017年

# 個別化医療から Precision Medicine

個人の遺伝素因・環境要因に合わせた (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  
診断/治療の個人化ではなく層別化を明確化。ゲノム一元論からの脱却

概念の革新：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

遺伝要因(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と明示的に強調。電子カルテの臨床表現型情報  
(Clinical Phenome)情報の重要性認識。

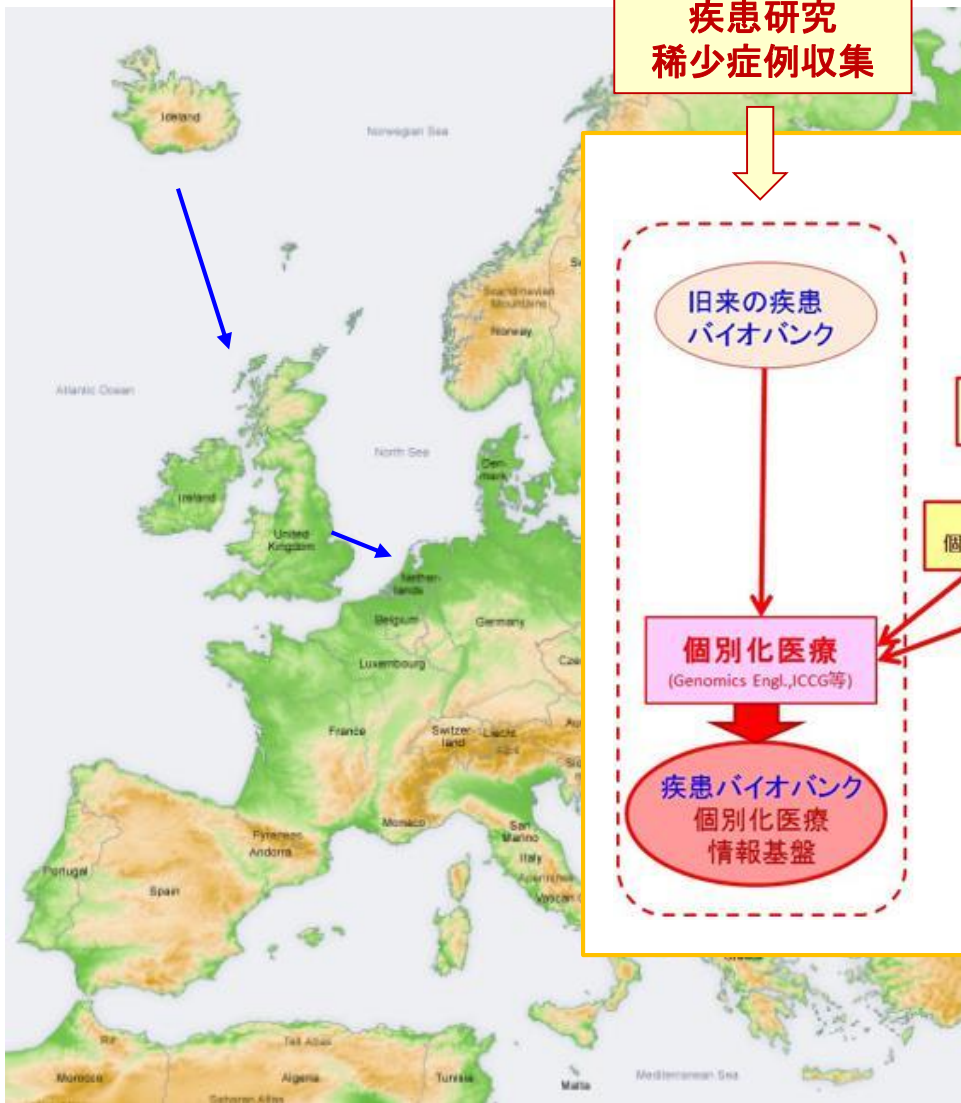
### (2) ゲノムコホート・Biobankの重視

Precision Medicineを実現する「情報基盤」として、ゲノムコホート/Biobankが  
不可欠であることを認識。

### (3) 日常生理モニタリング情報の包摂

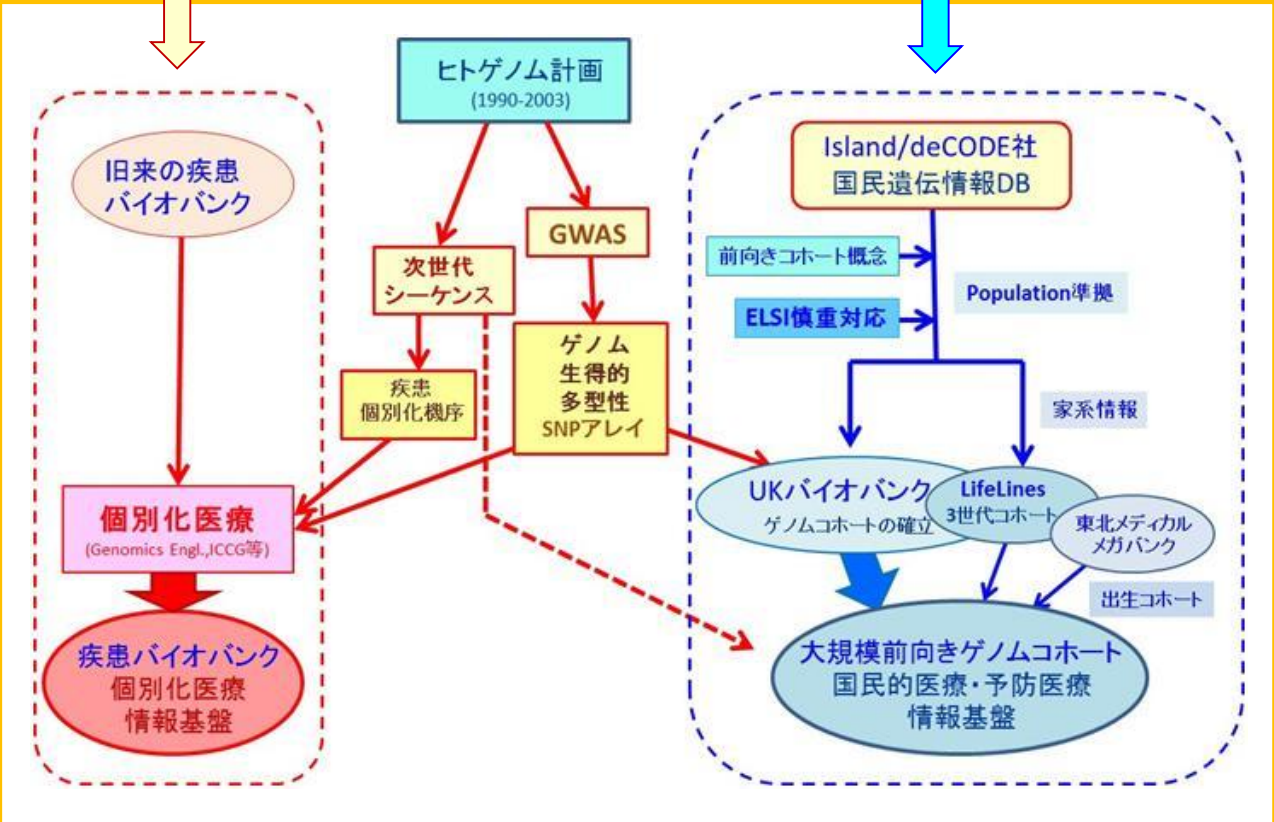
モバイルヘルス(mHealth)・wearableセンサーによる大量継続生理情報の価値認識

# 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



# ビッグデータ医学/医療の第2の流れ

## Biobankとゲノムコホートの世界的興隆

### バイオバンクの目的・機能の変化

- 従来は：稀少疾患組織標本や臨床研究の資料保存のため
- 最近では：ビッグデータ時代での変換：ゲノム医療の情報基盤としての役割・世界的に普及
- 疾患BioBank：ゲノム・オミックス個別化医療／創薬の情報基盤：
  - 従来の疾患バイオバンクが個別化医療の概念により変革、個別化医療の情報基盤としての役割
  - 疾患罹患患者の網羅的分子情報とそれに対応する臨床表現型情報の収集。
  - 疾病の個別化分子機序解明や治療戦略構築、予測医学・創薬科学への貢献
- Population型BioBank：「集合的遺伝情報」国民医療レベルの向上、予防医学の情報基盤
  - 「健常者」前向きコホート。網羅的分子情報（genome）と環境情報（exposome）収集  
長期間（生涯）を追跡するゲノム・コホート
  - 主に遺伝子素因情報も含めた「多因子疾患」の疾患の発症リスク予測、重症化予測

### 欧米のBiobank

- 英国 UK biobank
  - 50万人の健常者。40～69歳（2006-2010, 62Mポンド）、追加調査（2011-16, 25Mポンド）
  - 健診データ（血液・尿・唾液サンプル、生活情報）とゲノム情報（SNPアレイを集め、健康医療状況を追跡する。アイスランド、deCODE社の「国民遺伝子情報データベース」の概念を新たに実現
- 英国 Genomics England,
  - 2013開始、2017年までに10万人のゲノム配列収集。全ゲノム次世代シーケンス
  - 最初の対象は稀少疾患（患者・家族）、がん患者、最初はEnglandのみ。企業との協同重視
- 欧州 BBMRI (Biobanking and Biomole Research. Infrastructure.)
  - 250以上の欧州各国のBioBankを連携する
- オランダ Lifeline
  - 165000人北部オランダ 2006年開始 30年間の追跡、家系情報・3世代コホートを世界初で実現
- 米国 Precision Medicine Initiative, Genome Cohort：“All of Us”コホート
  - これまでのBiobank（例えばBioVUなど）を集めて100万人のゲノムを集める

# ビッグデータ医学/医療の2つの流れに起因する大規模な生命情報DB/KBの出現と利用

- ヒトゲノム解読計画以降急速に進展
  - Hapmapプロジェクト, 1000 genome, がんICGC, TCGA, TopMED
  - ゲノム変異・多様体
    - dbSNP, HGMD, **Clinvar**, **Clingen**, OMIM, GWAS catalog
    - 表現型との対応: dbGaP, EGA
  - 遺伝子発現プロファイル
    - 疾患特異的transcriptome: **GEO**, **ArrayExpress**,
    - 薬剤特異的transcriptome: **c-Map**, **LINCS**
  - タンパク質
    - 3次元構造: PDB, Swiss-Prot,
    - タンパク質間相互作用: **HPRD**, **STRING**, BIND
  - 分子ネットワーク、パスウェイ
    - KEGG, TRANSFAC, BioCyc, Reactome
- 各種バイオバンク症例ベース（制限アクセス）
  - UK biobank, BMBRI, 東北メディカル・メガバンク
- これらの大規模DB/KBを組合せてゲノム医療/創薬を推進

# 医学/医療へのビッグデータの衝撃

	HiSeq 2500	HiSeq Proton
本体価格	約1億円	約2500万円
モード/チップ	ハイブリッドブリック	カピッドラン
解読速度	118	2798
リード長 (bp)	2 x 100	2 x 150
データ量 (G)	8960	8320
設置コスト (ヒト1人ゲノム)	約1万円	約1万円

次世代シーケンサの登場  
**シーケンス革命 (2007)**



コストレスで高精度な網羅的分子情報の出現

## 1. ゲノム・オミックス医学/医療の進展

- **Clinical Sequencing**によるゲノム・オミックス医療の**臨床実装**の急速な進展

## 2. Biobank/ゲノムコホートの世界的普及

- 個別化医療/予防の**情報基盤**として普及

## 3. 大規模な生命情報DB/KBの出現

- ゲノム・オミックスによる**DB/KBの膨大化**

# わが国での現状「ゲノム医療元年 (2016)」

## ■「ゲノム医学実現推進協議会」(中間報告) 2015.7

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た

### ●例：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 (Top-gear、SCRUM-Japan)
  - 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
  - がんのゲノムパネル：来年先進医療 (7施設)
- 岡大,京大,北大,千葉大 病院併設型BB

### ●予定：2018年度より「がんゲノムパネル」先進医療 (7か所) 開始

## ■AMED (日本医療研究開発法人) がゲノム医療を推進に予算

### ●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。

### ●ゲノム医療実現推進プラットフォーム事業

### ●臨床ゲノム情報統合DB事業

ゲノム医療臨床実装では、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遅れてはいない。米国とは異なったBiobank準拠のゲノム医療/創薬を推進するべき



# ビッグデータ医療の課題

# 医療の「新しいビッグデータの革命性」

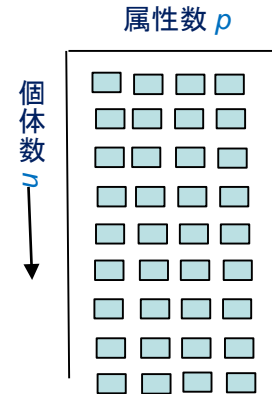
～ゲノム・オミックスデータの基軸的な特徴～

<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」( $n \gg p$ )

医療情報・疫学調査では 属性数：数十項目程度  
個体数：近年電子化の流れ⇒個体数：膨大

- 目的：Population（集合的）医学のBig Data  
⇒個別を集めて「集合的法則」を見る



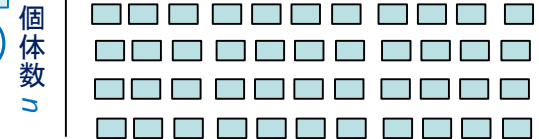
## 網羅的分子情報などのビッグデータ( $p \gg n$ )

1 個体のデータ属性数が膨大（SNP4000千万）  
ただし個体数は大規模biobankでも数十万

属性( $p$ ) $\gg$  個体数( $n$ ):従来の変量統計学が無効

「新  $np$  問題」：GWASは単変量解析の羅列

- 目的：医療の場合 個別化医療 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

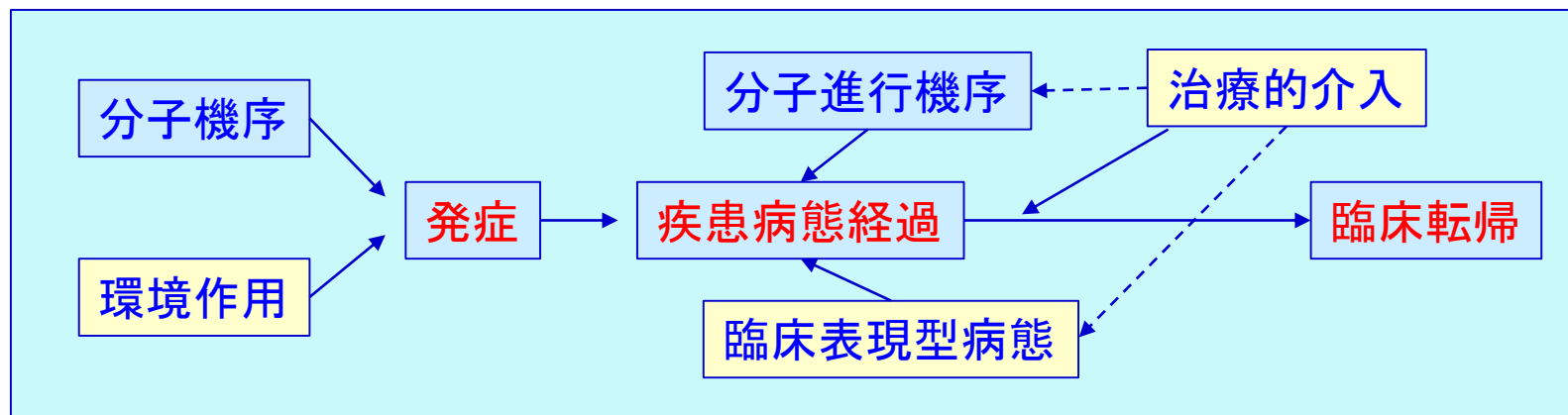


## 新しいデータ科学の必要性

# 医療の「ビッグデータ」革命は どんな既存のパラダイムに変革しているか

- Population（集会的）医学からパラダイム転換
  - <One size fits for all>の集会的医療はもはや成り立たない
  - 個別化医療“Personalized medicine”の概念
    - 個別化医療実現のために<個別化・層別化パターン>がどれだけ有るか  
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
  - 臨床研究の基礎：従来の範型RCTは、個別化概念を取扱えない
  - <EBM: (statistical) evidence based>の呪縛からの解放
  - 「標本」統計・「推測」統計学に制約されない臨床研究
  - Real World Data・ビッグリアルワールドデータからの知識生成
  - Learning Health System: 学習的医療実践

# 課題 1 対応する非ゲノム病態データの 検証的「情報化」



疾患経過のオントロジー

疾患分子発症進行機序（生体分子ネットワーク）

対応する非分子機序の明確化

環境発症要因 臨床表現型情報 治療介入効果



# 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 **M**edical **R**ecords + **G**enomics (NHRI-funded) **phase I** (2007-2011) EMR-basedゲノム研究の探求

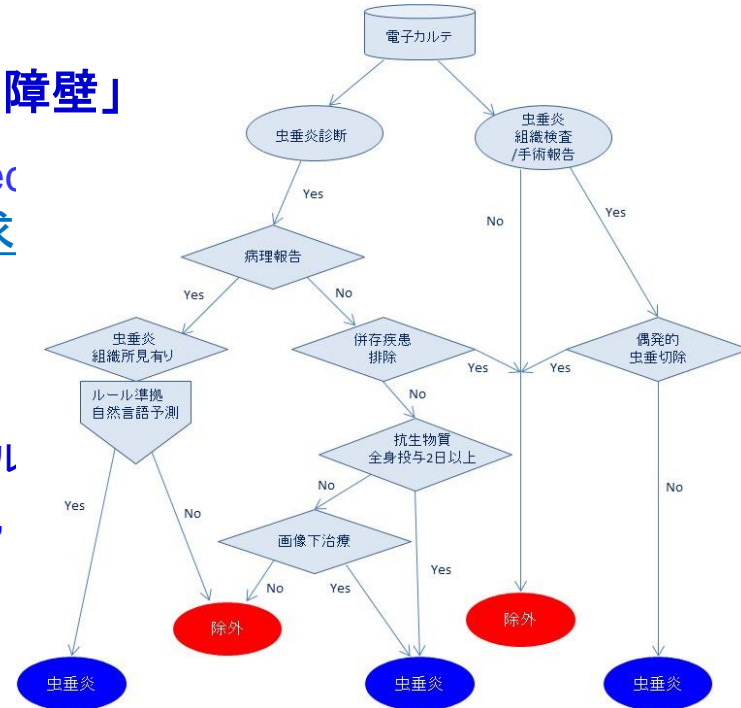
- EMR(臨床phenotyping)とbiorepositoryに基づくGWAS等 (EMR-based GWAS) が可能か (LHS)。
  - 開始時はGWAS全盛時代。ゲノム医療の臨床実装未着手
- 電子カルテより臨床表現型情報抽出 phenotypingルール
- 計画開始時参加施設：Mayo, Vanderbilt Univ, Marshfield, Univ. Washington, Northwestern Univ.など5施設,

### **phase II** (2011-2015) 臨床実装へ舵を切る

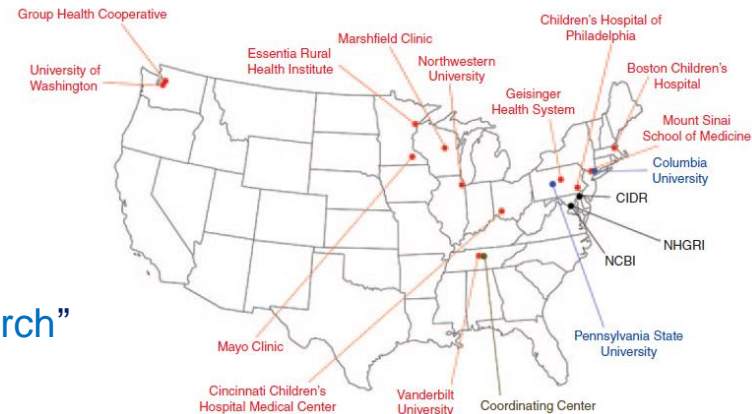
- MCWの臨床実装のインパクト, Vanderbiltの先制PG x
- 電子カルテと遺伝情報の統合
  - 電子カルテへのゲノム情報の統合
  - **PheKB** (Phenotype Knowledge Base)
  - ゲノム医療の実装、PGxの臨床応用
  - 結果回付 **Return of Result**, ELSI等
- 4つのサイトが新しく加わる
  - 小児病院グループとMount Sinai, Geisinger

### **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



# 課題 2 生命医療情報のビッグデータ化による 「革新的(innovative)知識」発見の困難性

- 臨床ゲノム医学
  - 全ゲノム配列の普及、多層オミックス情報の収集、分子画像の発展
  - ビッグデータ化：超多次元相関ネットワーク
  - 〈網羅的分子情報と臨床表現型情報〉の相関
- 予防ゲノム医学
  - バイオバンクの大規模化、国際連携によるバーチャル連携
  - 〈遺伝的素因と環境/生活様式要因〉の相互作用と発症の相関ネットワーク

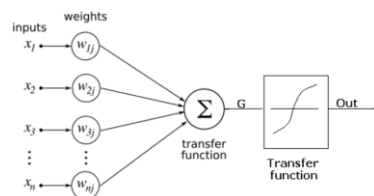
いずれも超多次元複雑ネットワークの縮約理論

# 人工知能 Deep Learning への期待

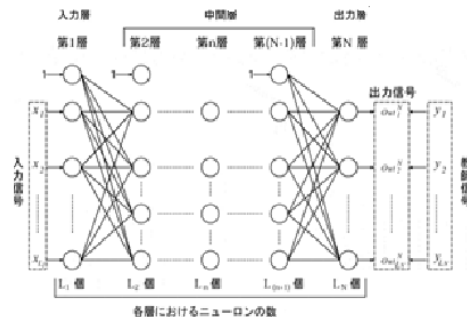
- 機械学習のこれまでの限界

- 「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



神経情報素子



多層ニューロネットワーク

- Deep Learningの革命性

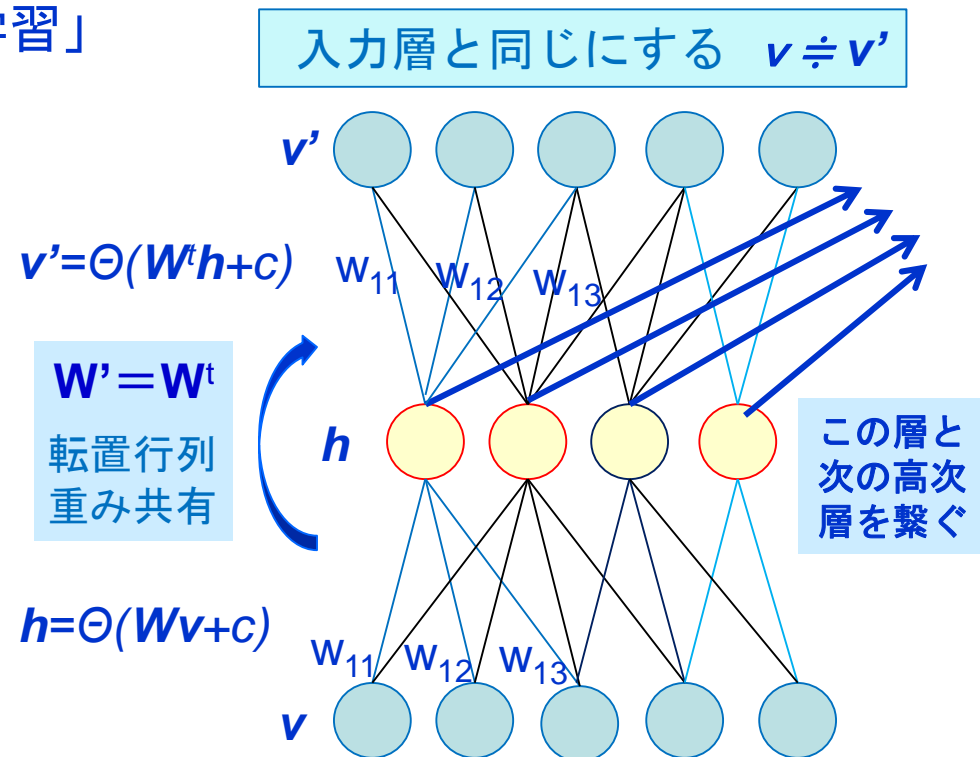
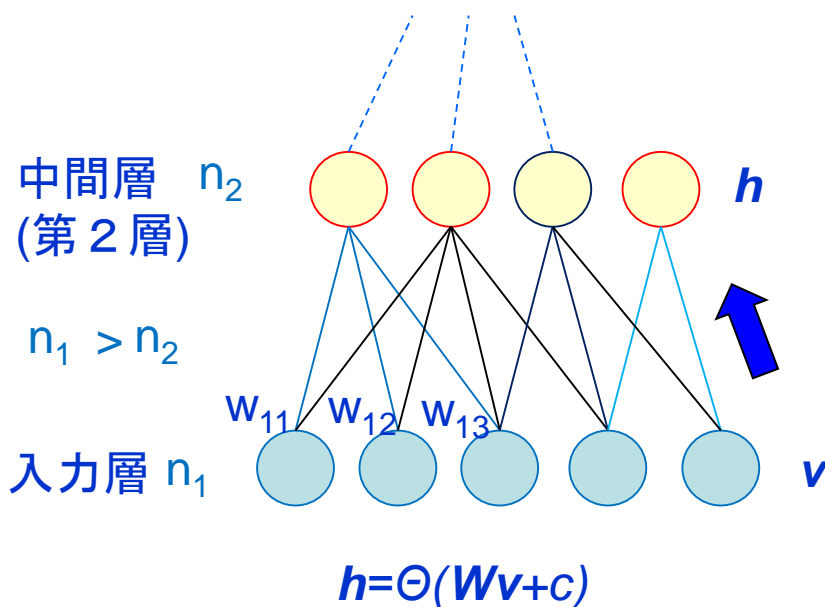
- 「教師なし学習」

- 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 Autoencoder

- 対象に固有な**内在的特徴**を学ぶ**自己符号化の原理**
- 格段ごとに入力の少ない中間層を入力へ逆投影して復元できるか
- 次元を圧縮され可及的に復元する ( $1000_{\text{nodes}} \Rightarrow 100_{\text{nodes}} = ? \Rightarrow 1000_{\text{nodes}}$ )
  - 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
- 最終層で人との対応「教師あり学習」



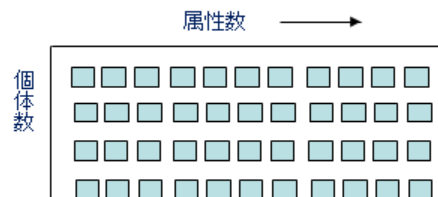


# 「ビッグデータ」のData 縮約原理

問題点 属性項目数(p) ≫ サンプル数(n)

p: 数億になる場合あり、n: 多くても数万

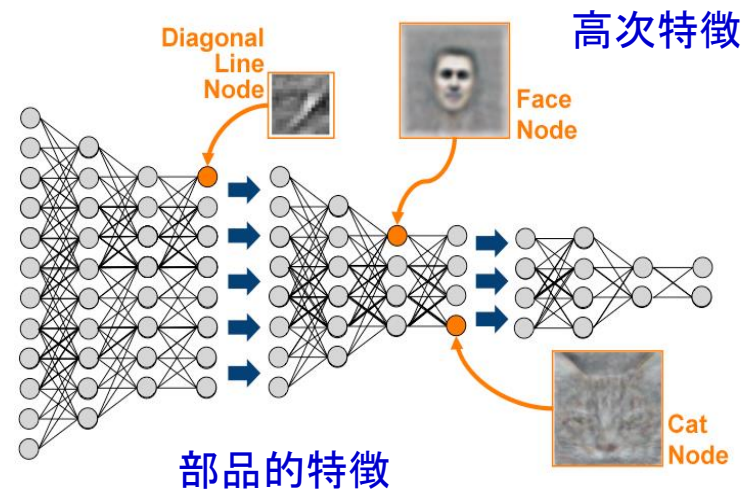
これら膨大な属性変数がすべて独立ならばビッグデータの構造解析は不可能。単変量解析の羅列 (GWASのManhattan Plot) しか可能でない



## 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。  
「構造」の存在：推定した結果で判明

データ構成性の原理  
principle of compositionality



構成性の推定 (Google猫の例)

# 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は高い
- 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<遺伝素因–環境要因(生活習慣)>
- 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
  - 線形分解（特異値分解：SVD）との比較

# 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:  $G \times E$

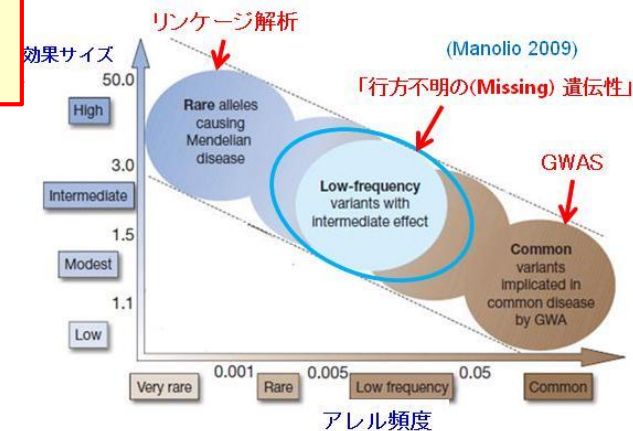
SNPの相対リスク  
低い(1.1~1.3)理由  
 $G \times E$  組合せ特異的効果  
を環境要因の平均



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



オランダ  
飢饉 (1944)

- 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 $\alpha/\gamma$ （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

環境因子



Epigenome変化



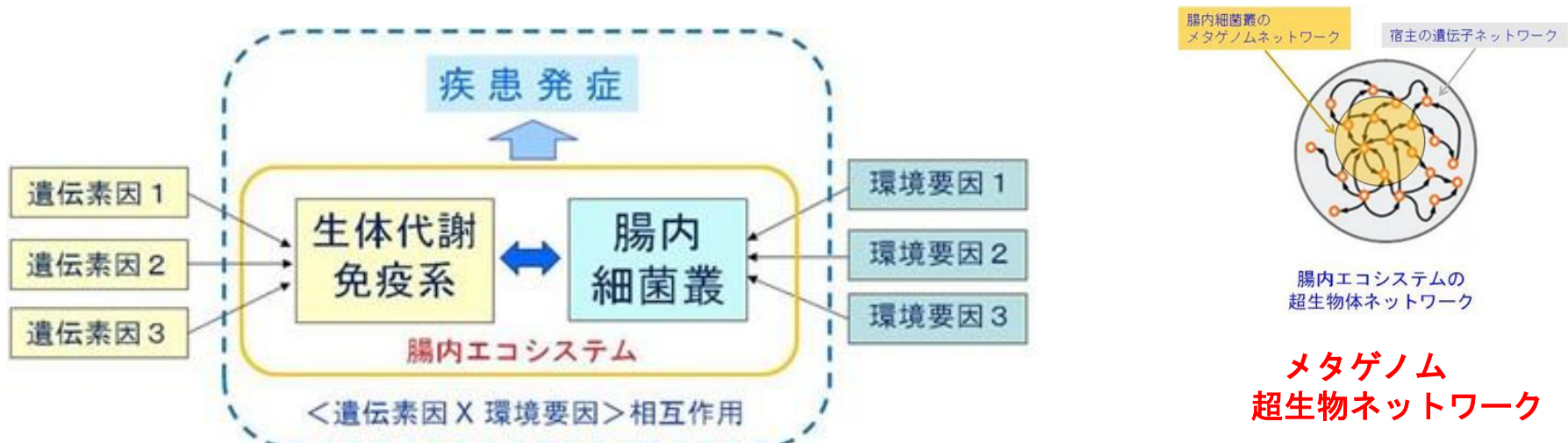
遺伝子発現調節



疾病発症

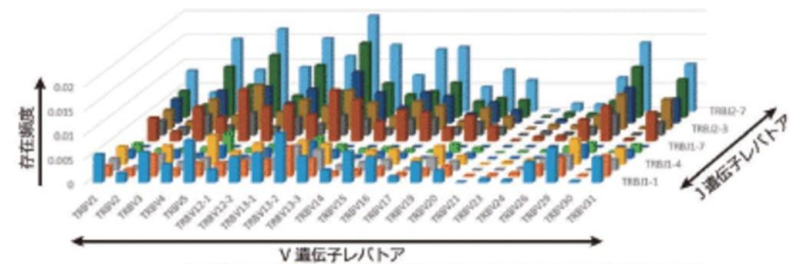
# 腸内細菌叢microbiome：メタゲノム

- **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)

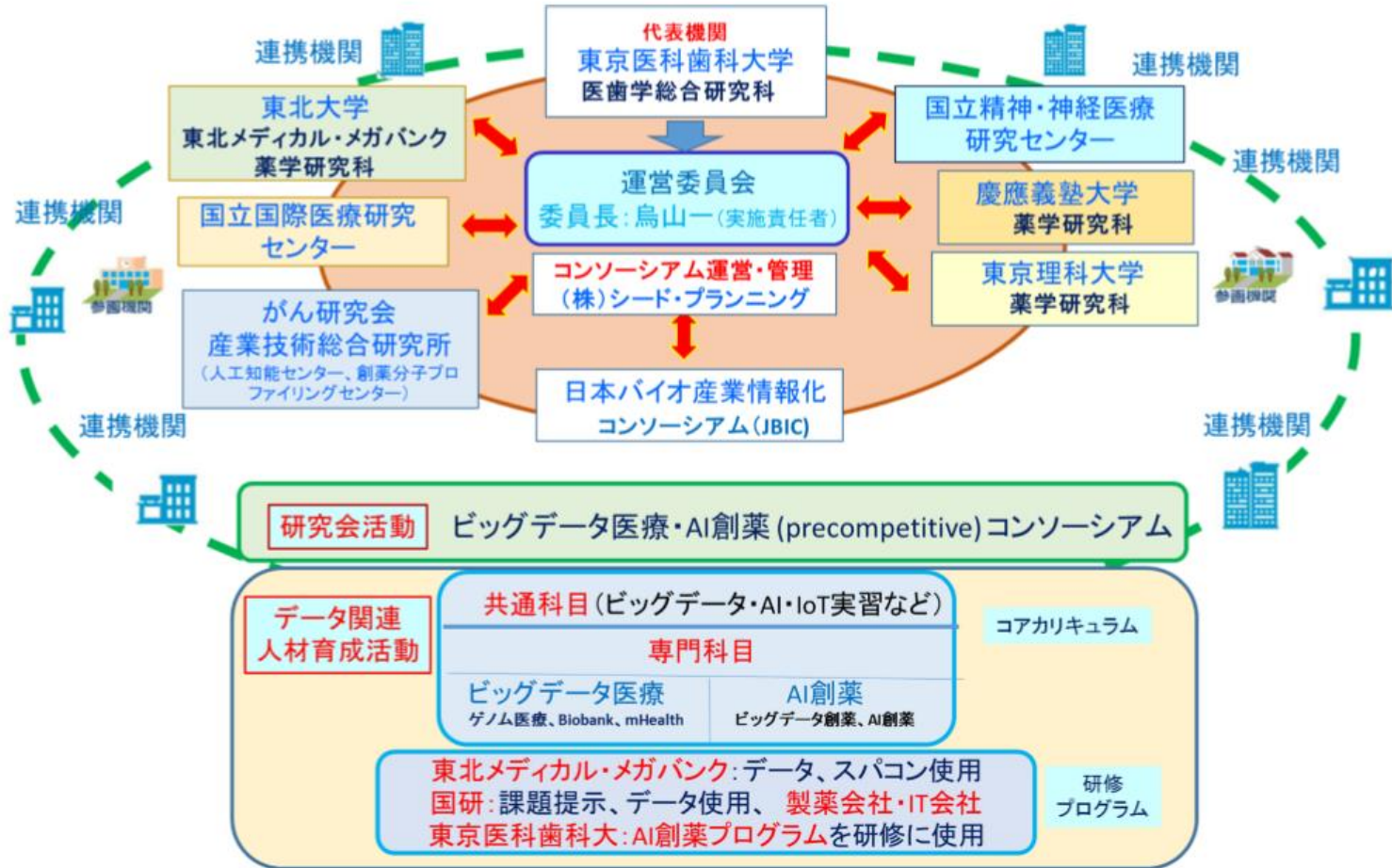


# 第2世代のゲノム・オミックス医療

- 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

**疾患メタ・オミックス修飾**

# 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

**ご清聴ありがとうございました**