

研究領域「イノベーション創発に資する人工知能基盤技術の創出と統合化」  
(研究総括: 栄藤 稔、H28年度発足)

# 人工知能を用いた統合的ながん 医療システムの開発

国立研究開発法人・国立がん研究センター研究所

がん分子修飾制御学分野・分野長

浜本 隆二



# 国立がん研究センターの理念と使命

- **理念**

社会と協働し、

全ての国民に最適ながん医療を提供する

- **使命**

1. がんの本態解明と早期発見・予防
2. 高度先駆的医療の開発
3. 標準医療の確立と普及
4. がんサバイバーシップ研究と啓発・支援
5. 情報の収集と提供
6. 人材の育成
7. 政策の提言
8. 国際貢献

中釜斉理事長・所信説明会資料より

# 新体制において特に強化する課題

- **ゲノム情報に基づく個々人に最適化された医療  
（“最適医療（Precision Medicine）”）提供体制の整備**
- **アンメットメディカルニーズ（希少がん・難治がん等）  
の課題解決のための研究・診療体制の強化**

中釜齊理事長・所信説明会資料より

# がん領域におけるPrecision Medicine実現は世界の潮流であり、 国策として取り組むべき課題

塩崎恭久厚生労働相とシルビア・パーウェル米厚生長官、鄭鎮※韓国保健福祉相は19日午後（日本時間20日未明）、米ニューヨークで会談し、がん治療研究への支援や早期発見技術での連携を強化する方針などを盛り込んだ共同声明を採択した。声明は「がんに終止符を打つことを共通の目的」と位置付け、日米韓3カ国でがん研究をリードし、抜本的ながん対策の推進によってがん患者の苦痛を軽減することを目指す。（※は火ヘンに華）

声明は、日本の国立がん研究センターなど3カ国の最先端のセンターが参加して共同研究などに取り組む国際的組織の設立や、それらの研究データと分析結果の世界への公開のほか、個々の患者に合わせたがん治療法や抗がん剤を特定したり、がん細胞を含むたんぱく質の遺伝情報の解析を進めることによって治療の方向性を定めたりする研究を進めることなどを盛り込んだ。

オバマ米大統領は今年1月の一般教書演説で、がん治療を推進する決意を表明。「ムーンショットイニシアチブ」として、昨年長男を脳腫瘍で亡くしたバイデン副大統領に指揮するよう指示した。さらに、3月の日米韓首脳会談で、がん治療研究のための取り組みで3カ国が協力することに合意していた。【阿部亮介】

毎日新聞 平成28年9月12日（火）

# 本CREST Projectの目的

Precision Medicineを指向した人工知能を用いた統合的ながん医療システムの開発を国家プロジェクトとして進め、我が国独自で開発した医療システムを国民の皆様を提供する。



# 本プロジェクト の概念図

## NCC がんOmic研究 成果 (基本データ)

ゲノム・エピゲノム

クリニカルシーケンス

画像情報 (CT, MRI など)

microRNA・血液

株化がん細胞・PDXマウス



NCCがんゲノム医療情報  
統合データベース

### 人工知能技術

(Deep Learning: 深層学習)



新規がん診断システム



バイオインフォマティクス、  
テンソル情報の機械学習における第一人者



深層学習の世界トップレベルの技術を有する日本の企業



個別化医療  
実現支援システム

## Precision Medicine *CANCER*



新規創薬設計システム

NCC Precision Medicine Catapult  
研究開発から産業化への動きを  
推進する新機構の設立



在宅医療への導入

ヘルスケア産業への展開



# 本プロジェクトの体制図

国立研究開発法人・  
国立がん研究センター  
(我が国におけるがん対策拠点・  
国立高度専門医療研究センター・  
臨床研究中核病院)



研究代表者・共同研究者  
は若手中心

国立研究開発法人・  
産業技術総合研究所  
人工知能研究センター



共同研究者  
研究チーム長：瀬々潤

研究参加者  
研究員：林浩平

統合的推進・創薬・DB構築  
[研究所]

AIを用いた統合  
的解析

研究代表者  
分野長：浜本隆二

研究アドバイザー  
理事長：中釜斉  
研究所所長：間野博行

データベース(DB)構築  
[研究所基盤的臨床開発研究コアセンター]

研究参加者  
センター長：吉田輝彦

microRNA解析  
[研究所]

研究参加者  
主任分野長：落谷孝広

クリニカルシーケンス  
[研究所]

研究参加者  
分野長：柴田龍弘

画像解析・治療  
[中央病院]

研究参加者  
科長：伊丹純  
がん修練専門医：  
小林和馬

クリニカルシーケンス  
[先端開発医療センター]

研究参加者  
分野長：土原一哉

画像解析・診断  
[中央病院]

研究参加者  
科長：荒井保明  
医員：三宅基隆

血液・臨床情報解析  
[中央病院]

研究参加者  
科長：山本昇

Preferred Networks



共同研究者  
取締役副社長：岡野原大輔

研究参加者  
最高執行責任者：大田信行

研究参加者  
ビジネス開発：ハムザウィ・カリーム

研究参加者  
エンジニア：大野健太

研究参加者  
知財担当：堂田文明



実用化

# AI技術の活用: Deep Learningの特徴

---

- マルチモーダル学習:
  - 異なる種類のデータを統合して入力として扱うことが可能
  - これらのデータをどのように組み合わせるかは  
Deep Learningが学習の過程で獲得可能。
- マルチタスク学習:
  - 複数の異なるタスクをモデルの一部を共有して同時に学習することで、  
学習の効果をあげることが可能
- 表現学習及び半教師あり学習:
  - 大量のラベル無しデータからデータの表現方法を獲得し、  
これにより少量のラベル有りデータから学習可能
- 階層的な特徴の自動獲得:
  - 入力の高次の相関を捉えることができる。
  - 従来の解析手法ではこのような特徴を全て兼ね備えることは難しかったが、  
Deep Learningはこれら4つの特徴を兼ね備えており、がんデータ解析に適している。

# 個々の課題に対する詳細な戦略

課題名	データの種類*：準備状況	データ数	現況及び今後の方針
がんの診断/microRNA及び血液検査による早期がん診断システムの開発	microRNA, 血液検査データ：多くの患者さんのデータが、すでに電子データとして解析可能な状態である。	1万～10万	乳がん診断のPOCでは画期的な成果をあげつつあるが、13種類のがん診断にはさらなる検証と臨床試験を行い、実用化に向けて改良をしていく必要がある。 <b>平成31年度-平成33年度の期間で実用化を目指す。</b>
医用画像解析	画像データ（CT・MRI）：多くの患者さんのデータが、すでに電子データとして解析可能な状態である。	数万～数十万	画像処理技術によって医用画像から大量の特徴量を抽出し、臨床データとの関連を系統的に推論する為のアプローチであるRadiomicsの確立を行っている。将来的には、コンピュータ診断支援及び従来の人間の識別能力では困難であった治療効果予測や予後推定、病理や遺伝的因子の推定など行うことをを目標とする。 <b>平成33年度までに実用化の目処を立てる。</b>
AI技術を用いたヒストン修飾解析（データベースの整備/創薬への応用）	ChIP-Seq data: 次世代型ChIP-Seq法を用いて、深層学習に使用可能なデータを取得予定。	数千～数万	非常に精度の高いChIP-Seq dataに基づく、がん組織におけるヒストン修飾状況の解析及び診断・創薬への応用 <b>平成31年度-平成33年度の期間で次世代型ChIP-Seq法として実用化を目指す。</b>
Precision Medicineを指向した人工知能によるがん治療の最適化/クリニカルシーケンスデータ (TOP-GEAR/SCRUM-J)	NGSを用いたDNAシーケンスデータ：多くの患者さんのシーケンスデータが、電子データとして解析可能な状態である。	数千～数万	ゲノム分析によるがん治療の最適化は、正確な患者データと紐づけられたNGSシーケンスデータがまだまだ少なく、大量データを必要とするDeep Learningのこれまでの手法では困難である。よって、少ないデータからも学習を可能とする、半教師あり学習などの新手法を取り入れて検証・新規開発する必要がある。 <b>平成33年度までに実用化の目処を立てる。</b>

# CREST:人工知能を用いた統合的ながん医療システムの開発イメージ図

AI対応多次元データベース構築  
(NCCがんゲノム医療情報統データベース)

- ・ゲノム
- ・エピゲノム
- ・画像情報
- ・microRNA/血液
- ・クリニカルNGS
- ・改正個人情報対応  
同意・登録・長期追跡

Big Data

NCC



深層学習技術を中心としたAIによる解析

科学的発見



解析システム (NCCに設置)

- \*GPUクラスター
- \*サーバー: 4ノード
- \*GPUカード: Nvidia Titan X\*4枚/ノード
- \*ストレージ: 100TBデータ解析用の高性能GPUクラスターの準備と運用

密接な連携



産総研  
人工知能研究センター



データ解析は基本  
NCC内において産  
総研及びPFNより  
派遣されたスタッ  
フが行う。

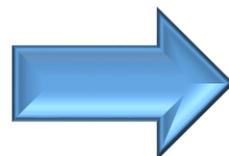
PFN



密接な連携



スモールフェーズ期間 (2年4ヵ月) POCを取得する。



加速フェーズにおいては、前向き試験を含めた研究の検証及び製品化を目指す。



実用化を  
推進

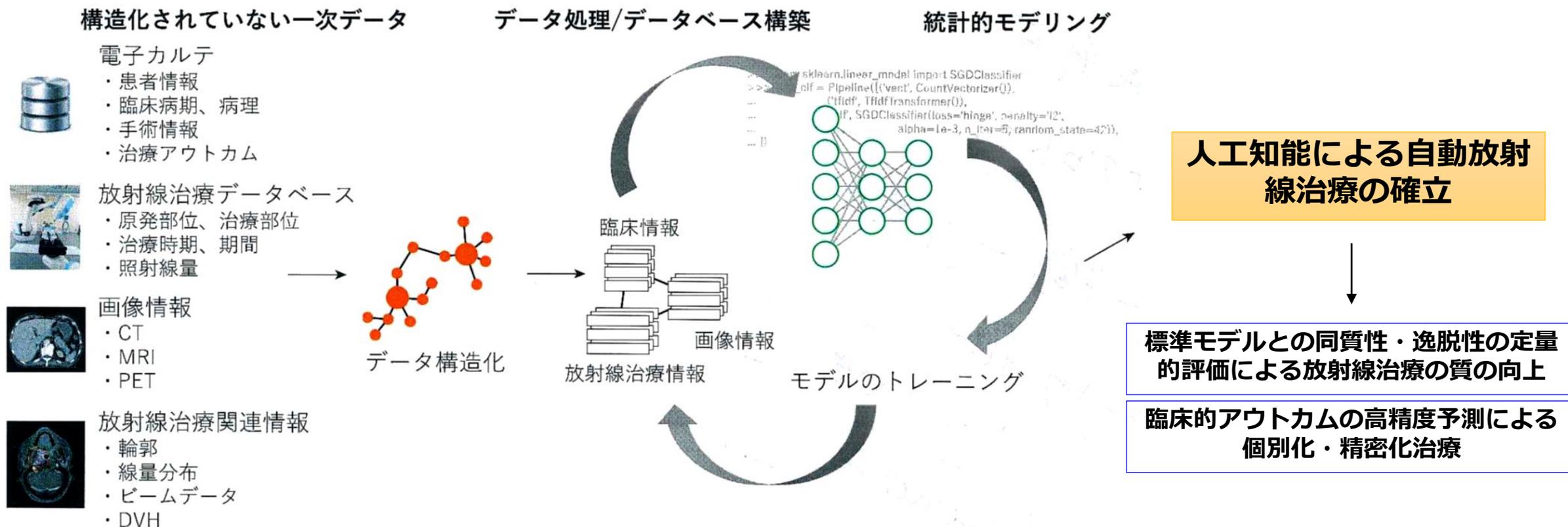
実用化  
(5年以内)

# 自動放射線治療を目指した包括的放射線治療データベースの構築

(目的) がん放射線治療に関わる膨大なデジタルデータを構造化されたビッグデータとして再構築し、個別化・精密化されたがん放射線治療を可能にするための人工知能による自動放射線治療を確立する。

(必要性) 放射線治療では、医用画像とそれを処理する為の高度な画像解析技術が重要な役割を果たし、診療過程において質的、量的に膨大なデジタルデータが蓄積されてきた。しかしながら、現状においては、その多くが不均一で構造化されておらず、価値創造の機会が大きく棄損されている。それらを、コンピュータによる解析、学習、推論の対象となるようなビッグデータとして蘇らせることがデジタル革命として急務である。

(概要) 院内に蓄積された膨大な一次データに対して構造化処理を施し、深層学習に基づいた自己進化型のデータベースを構築する。



# AI技術を用いたヒストン修飾解析及び創薬・診断への応用

## 深層学習を用いた次世代型ChIP-Seq解析法の確立

### このシステムの新規性・独創性

- \*FFPEサンプルからも解析が可能
- \*非常に再現性の高いデータを深層学習を用いて解析

➡ **ヒストン修飾情報の創薬・診断への応用が可能**

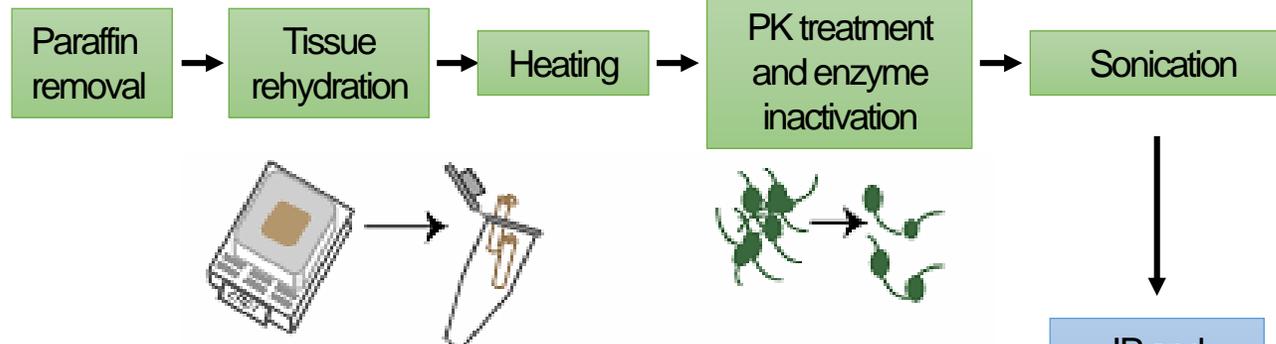
### Cancer cells/tumor samples

Culture cells

Fresh-frozen (FF) tumors

Formalin-fixed paraffin-embedded (FFPE) tissues

### I. Tissue preparation



バイオメディカル研究支援  
LabDroidを活用し、再現性の高い高精度のデータを取得する。

現在のがん研究分野において、エピジェネティクス解析（特にヒストン修飾）が遅れている。

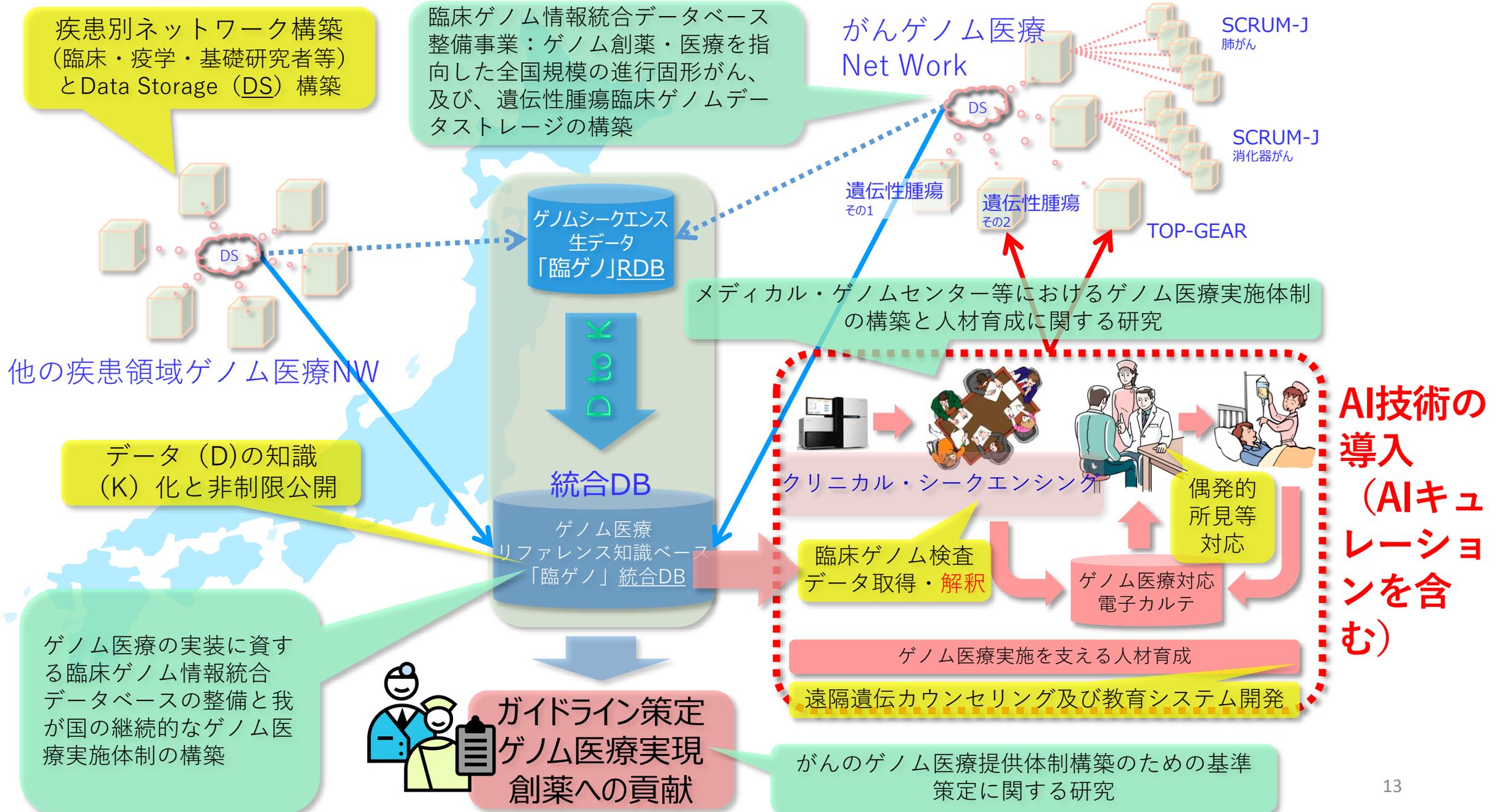
[理由]

高速かつ高精度に測定する技術が確立されていない。

AI技術を導入し、高速かつ高精度にヒストン修飾を測定する手法の開発

データベースの構築及び創薬・診断への応用

# 現在推進しているPrecision Medicine実現化プロジェクトへのAI導入



# 将来展望

---

- CREST事業が求める最初の2年4カ月でProof of Concept (POC:概念実証) の取得を目指す。
- 5年後を目処に実用化を目指していく。