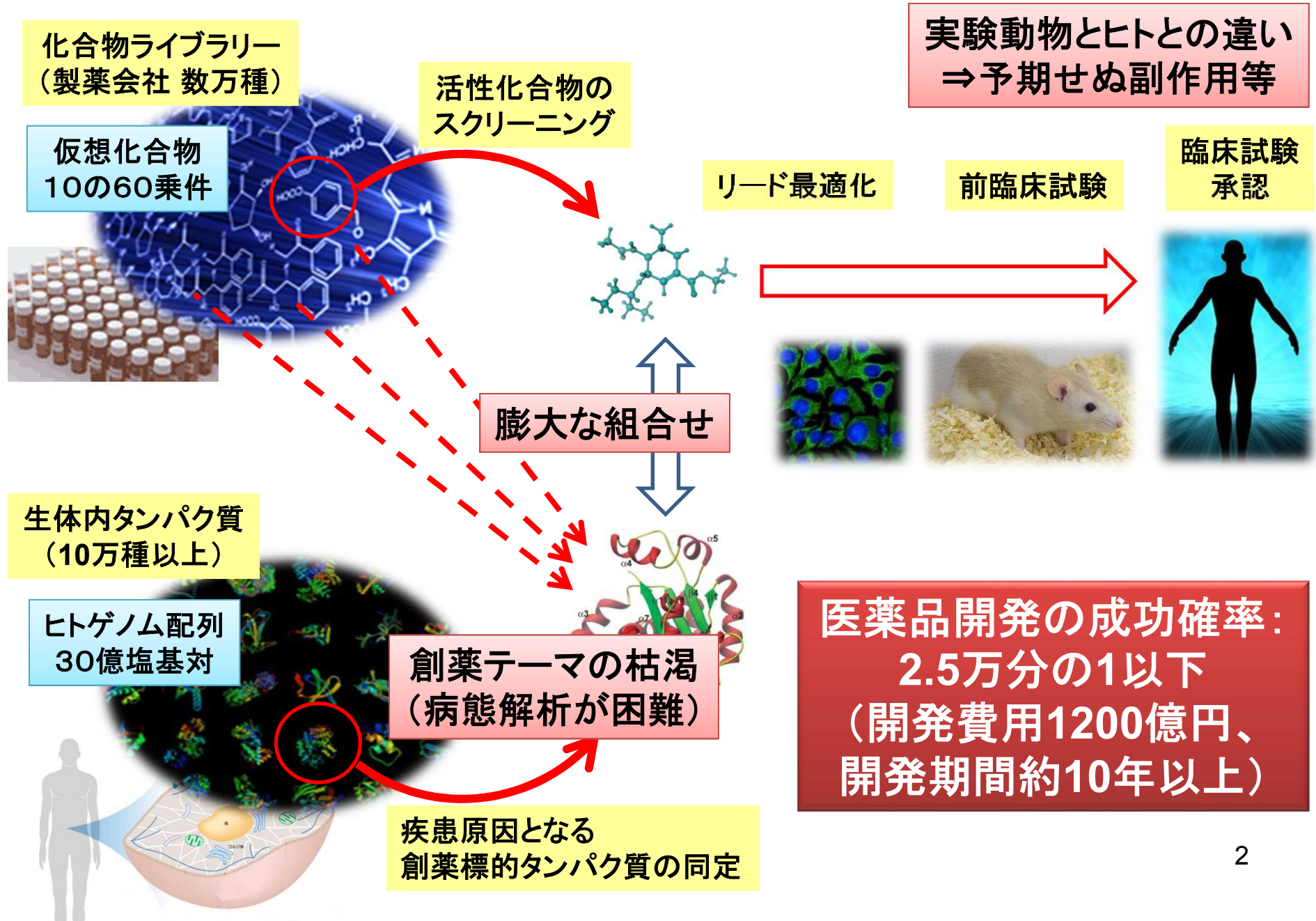


# 創薬における人工知能応用

京都大学 大学院医学研究科  
理化学研究所 QBiC/AICS/RC  
先端医療振興財団 先端医療センター研究所  
奥野 恭史

# 製薬業界の抱える課題



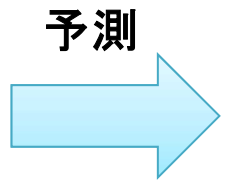
# ビッグデータ創薬：バーチャルスクリーニング法「CGBVS法」

## 人工知能技術によるタンパク質と化合物の相互作用予測

大量の人の顔画像を学習



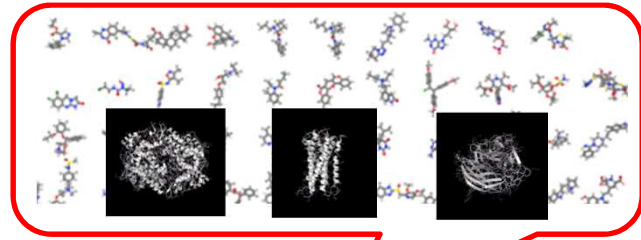
顔パターンの  
統計ルール化



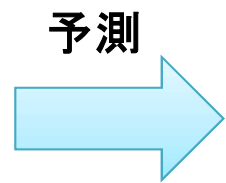
人の顔を自動認識



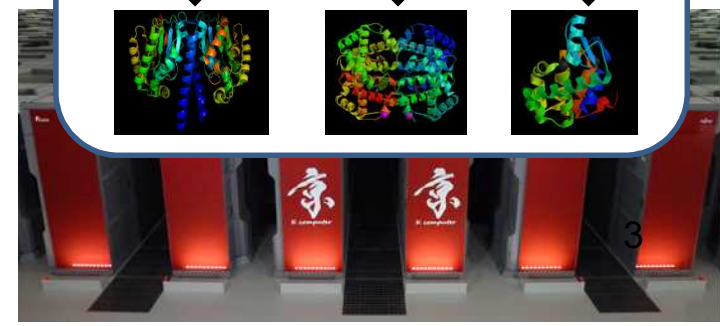
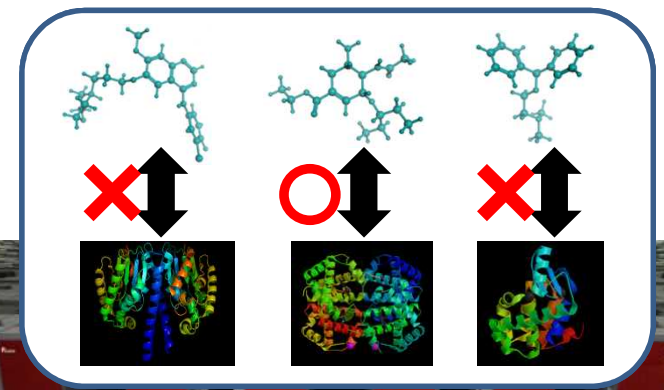
大量のタンパク質と化合物の結合データを学習



結合パターンの  
統計ルール化



病気の原因タンパク質に  
結合する化合物を「京」で認識



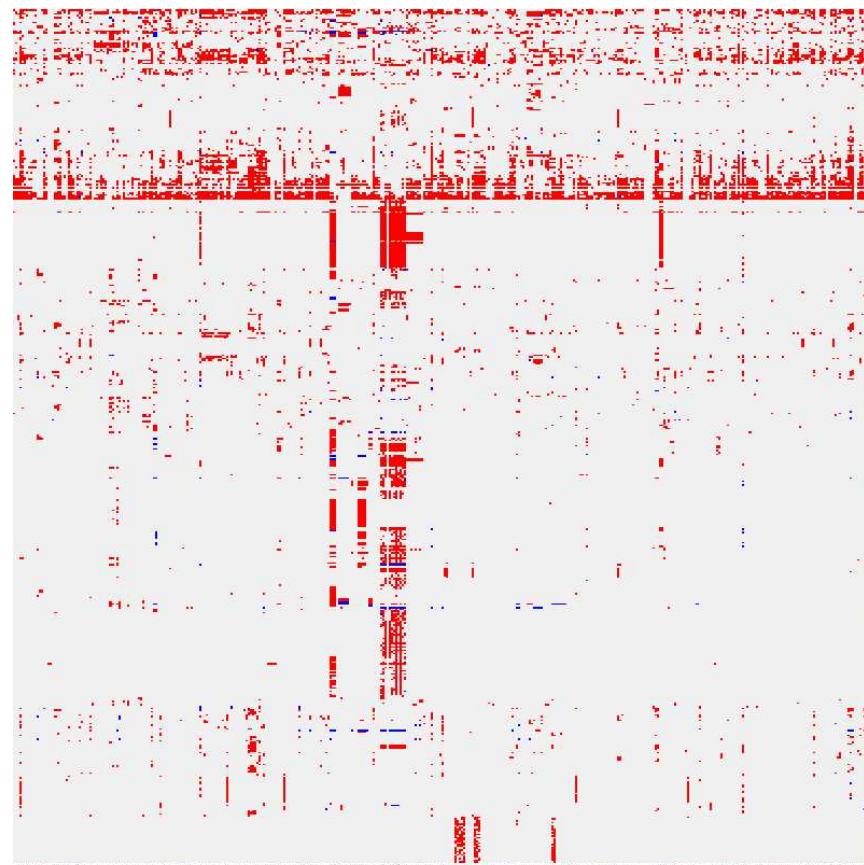
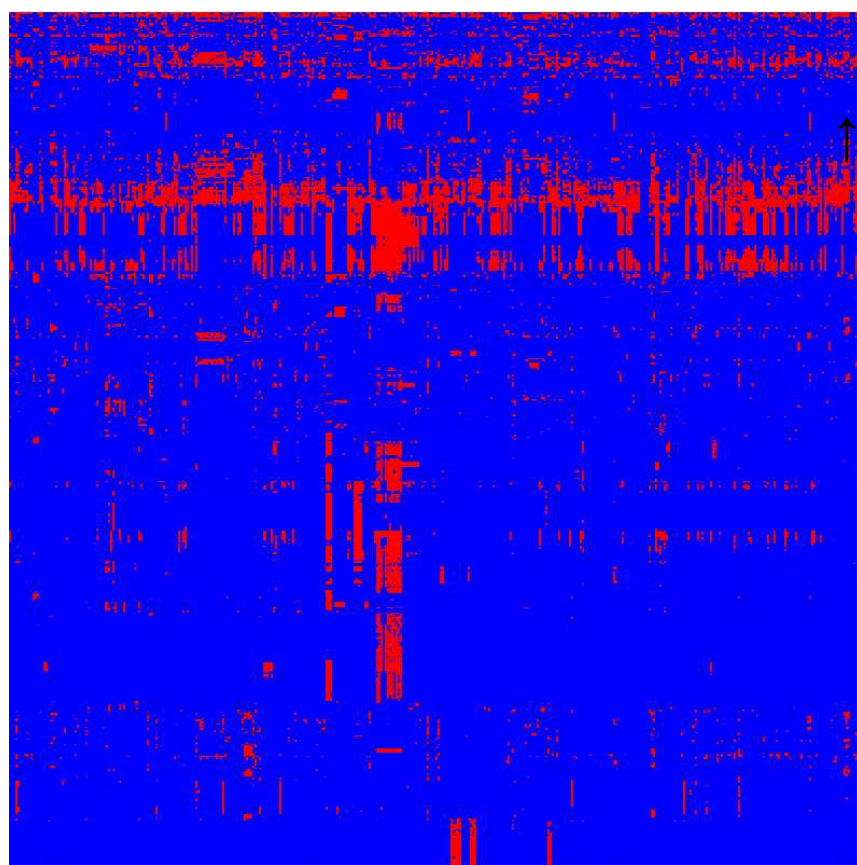


# CGBVS法による世界最大規模のタンパク質-化合物結合予測

タンパク質-化合物の全組合せ(189.3億ペア)を計算するのに、汎用計算機(16ノード使用)では約2年かかるところが、「京」をフルに利用したら5時間45分程度で計算が可能

「京」が予測した結合パターン(予測結果)

実際の結合パターン(実験結果)



Kinase (388種) → 631種 (+GPCRs)

Kinase (388種)

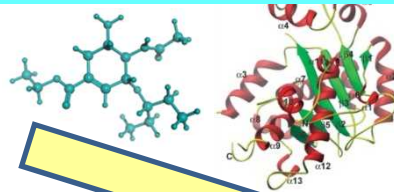
以後、予測結果を基に、各製薬会社が独自に医薬品開発を行う

# ドラッグデザインする人工知能

従来型のインシリコ創薬 (Docking計算、CGBVS)



化合物構造、タンパク質情報



活性の有／無



次世代型インシリコ創薬 (De novo Design)



標的タンパク名

新規な活性候補化合物

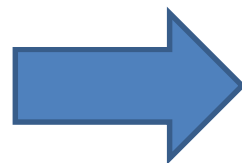
膨大な相互作用情報  
アッセイ情報の  
事前機械学習

多種多様な医薬関連ビッグデータを  
機械学習することで、薬効・安全性を  
総合的に加味した薬剤設計が可能



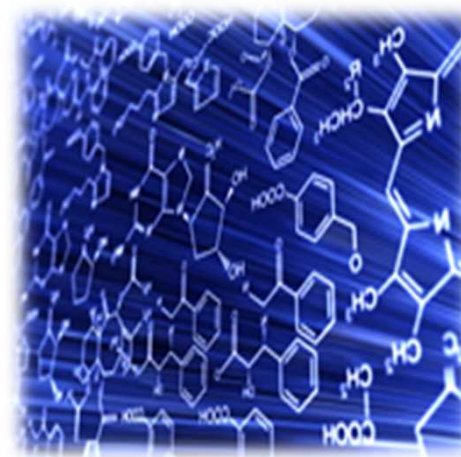
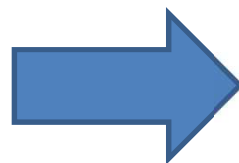
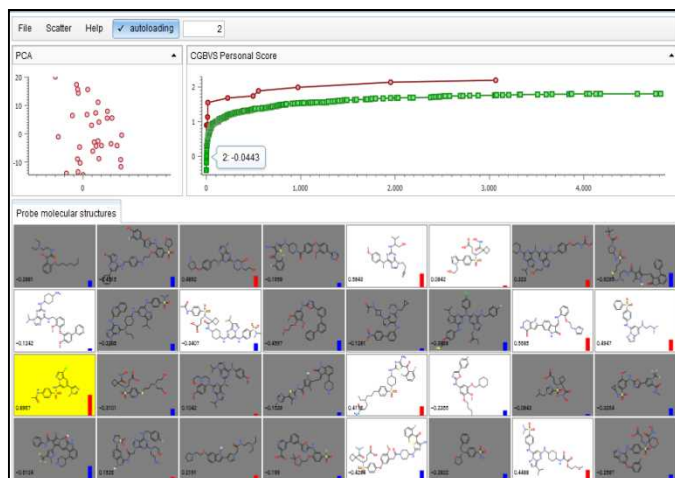
# アルファ碁と化合物合成

囲碁の人工知能が世界トップ棋士に勝利



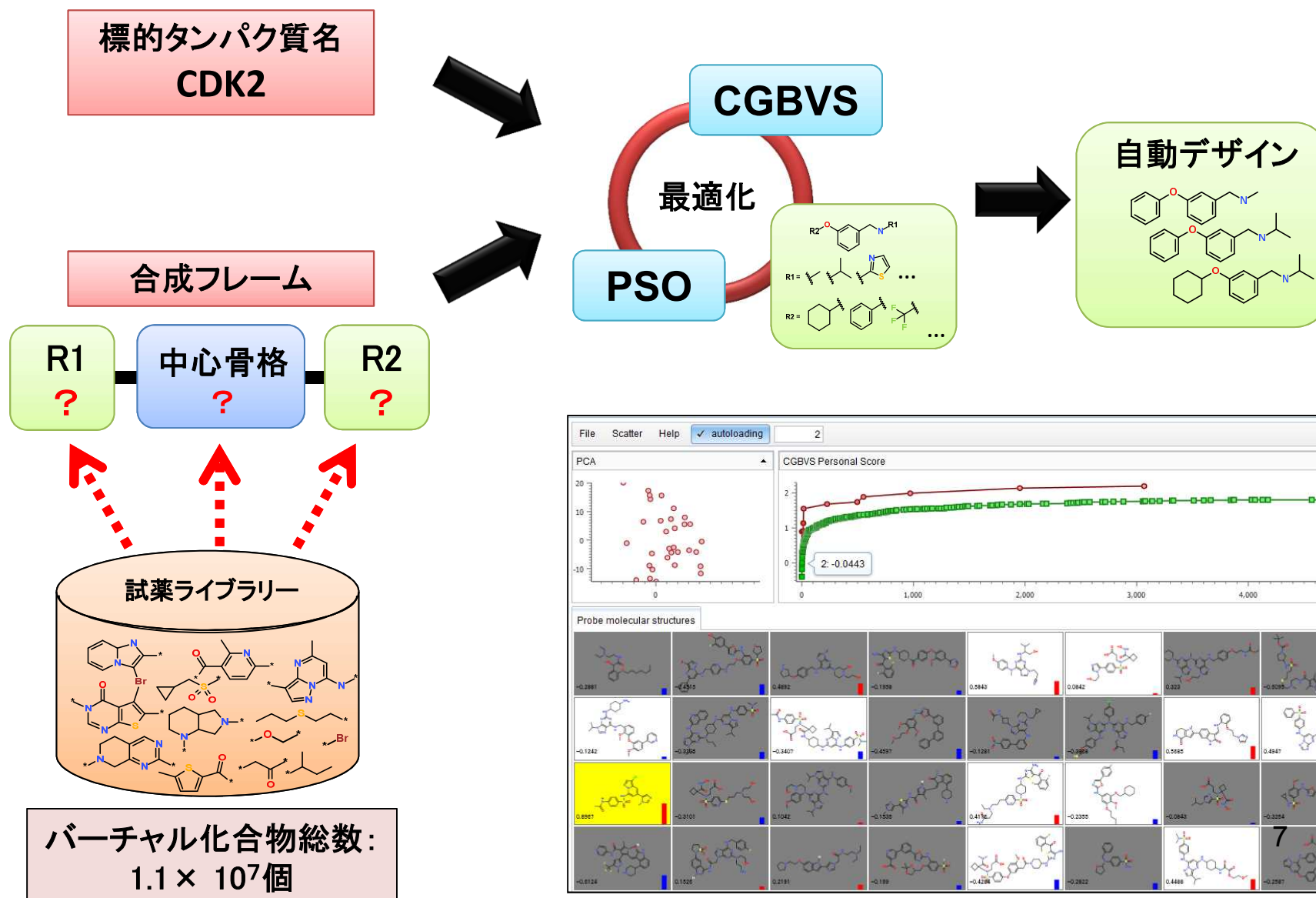
10の360乗の  
組合せから高速  
に勝ち手を推定

人工知能が医薬品の候補化合物を自動に提案



10の60乗の  
組合せから高速  
に活性化学構  
造を自動提案

# de novoドラッグデザインによる Kinase (CDK2) を標的とした活性化化合物の最適化の例



世界をリードする  
ライフAI立国を目指して：  
ライフ系とIT系企業による  
産学AIコンソーシアム設立

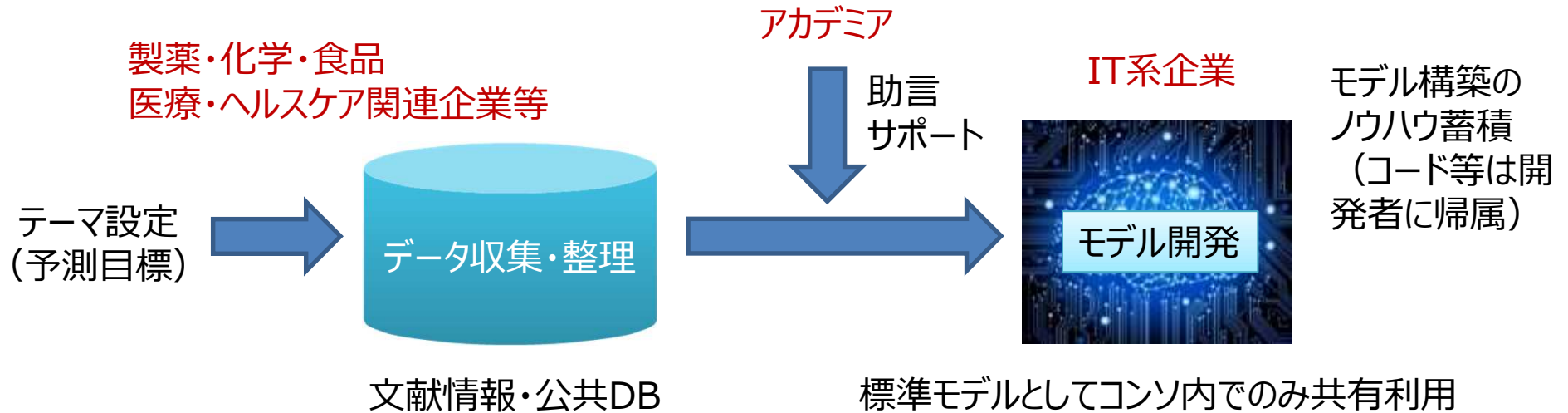


京大・理研・医薬基盤研・東大等（7機関）  
ライフ系企業（35社）  
IT系企業（27社）  
その他（3）とで産学コンソーシアムを設立



# ライフ・インテリジェンスコンソーシアム (LINC)

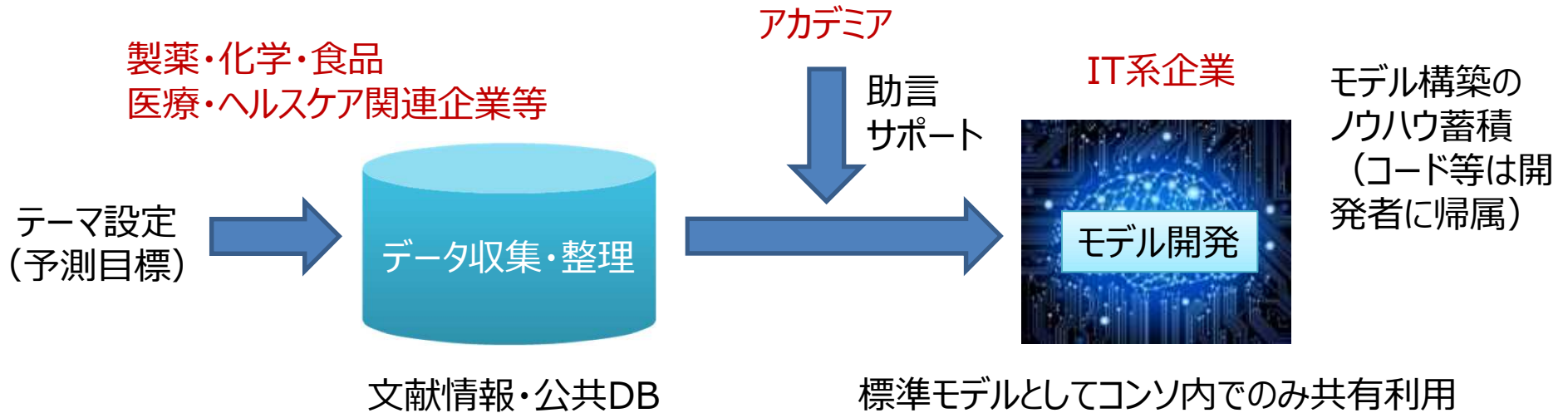
Pre-Competitive Area



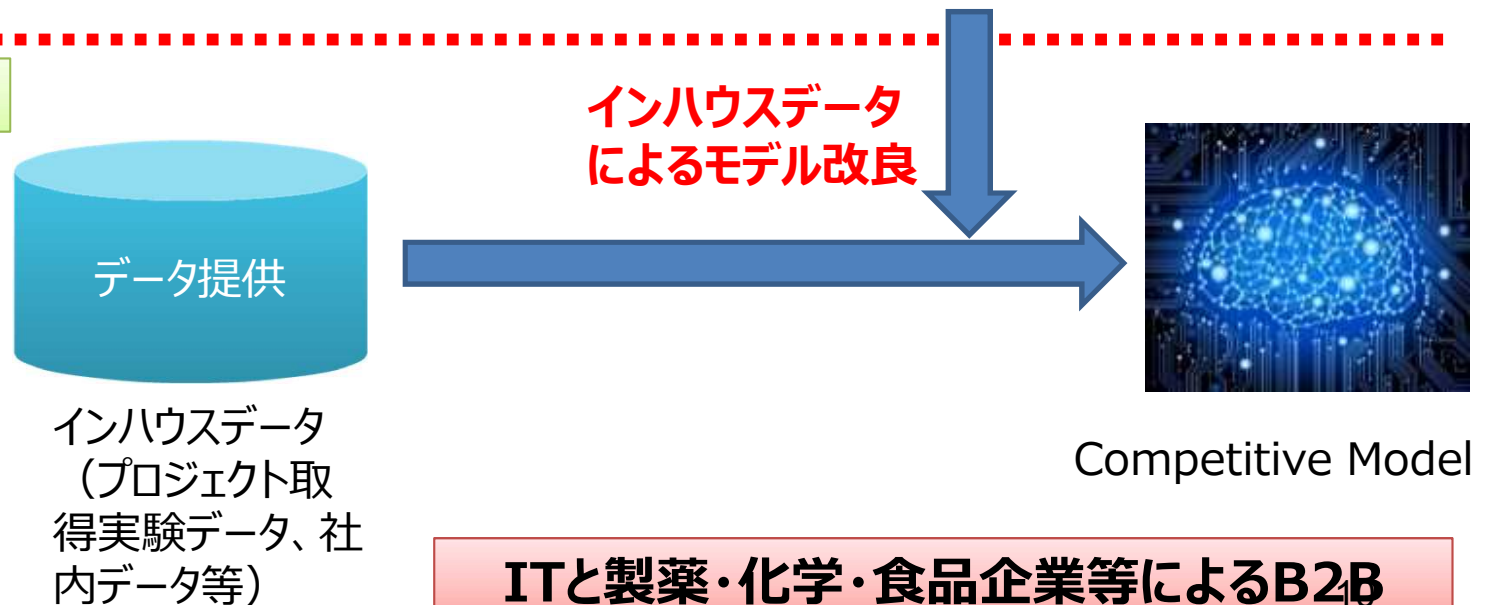
参加費は無料、ただし、参画企業はマンパワーを拠出して開発に参加することを義務

# ライフ・インテリジェンスコンソーシアム (LINC)

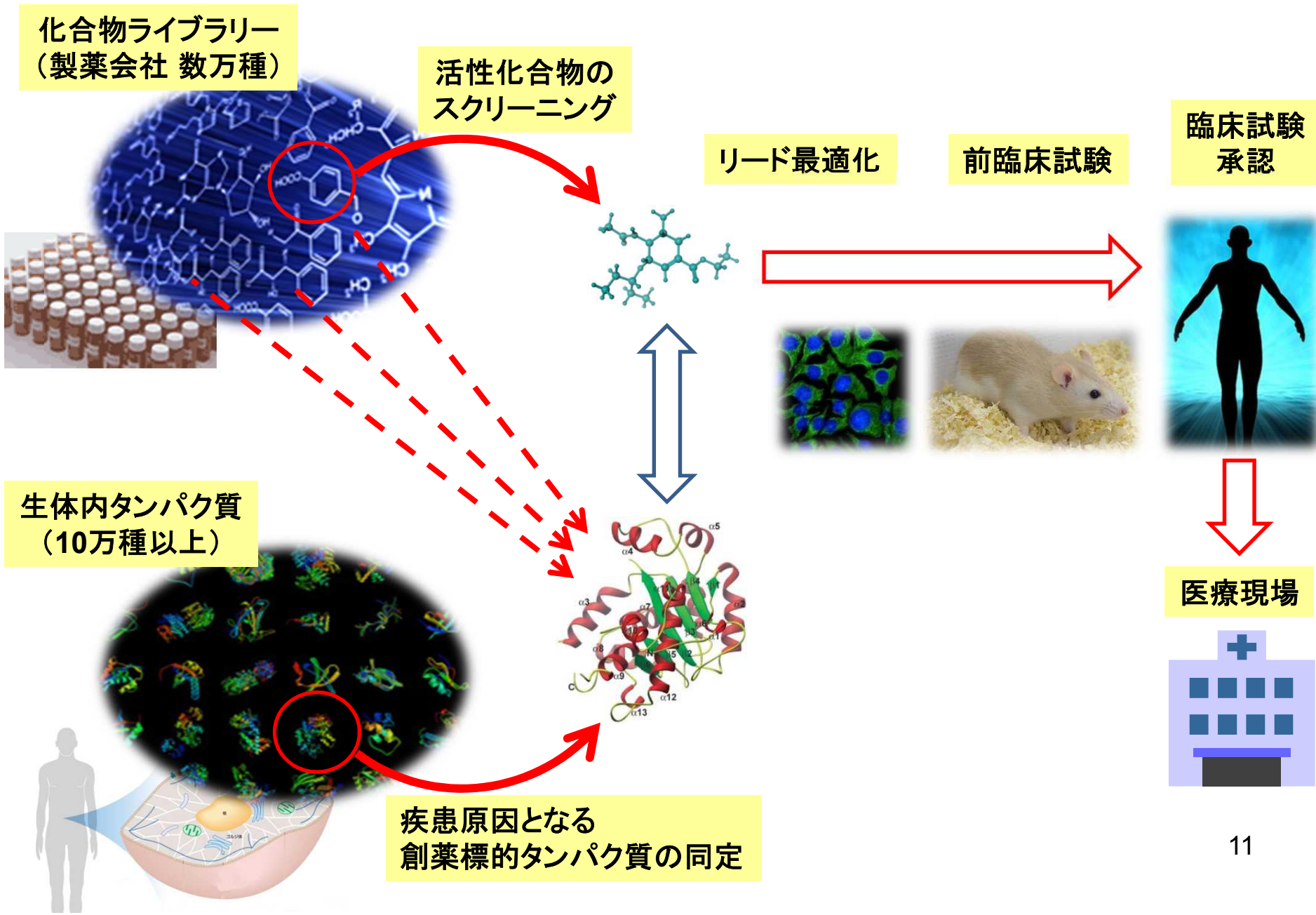
## Pre-Competitive Area



## Competitive Area

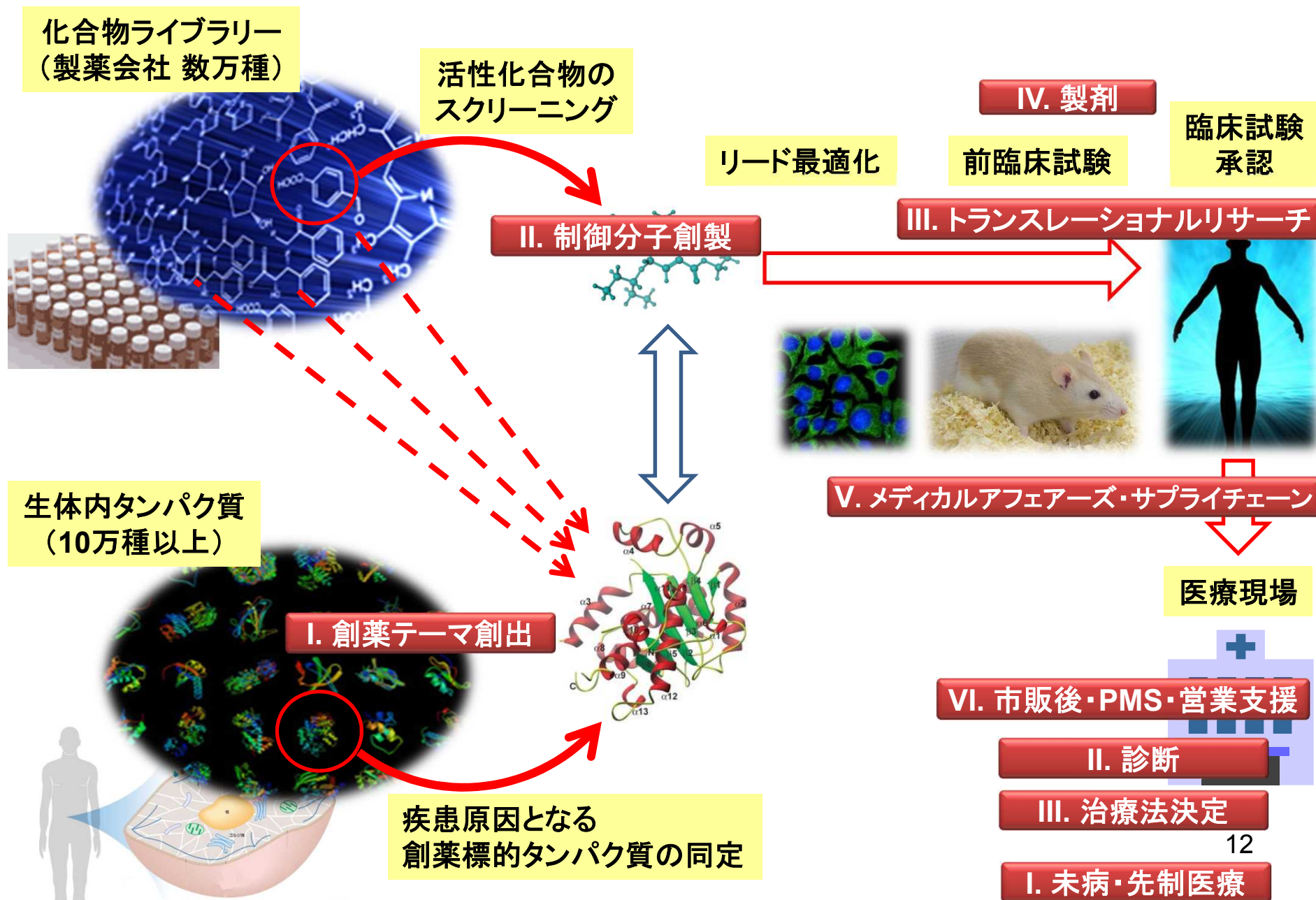


# 医薬品開発フローとAIニーズ





# 医薬品開発フローとAIニーズ

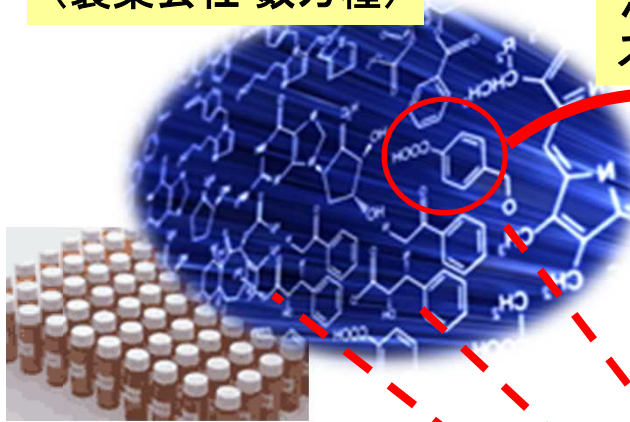


# 医薬品開発フローとAIニーズ

PMDA, FDA等の薬事承認プロセスの自動化とノウハウ蓄積のためのAI（自然言語処理等）に期待大

化合物ライブラリー  
(製薬会社 数万種)

活性化化合物の  
スクリーニング



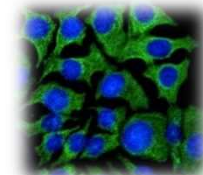
II. 制御分子創製

リード最適化

前臨床試験

IV. 臨床試験承認

III. トランスレーショナルリサーチ



生体内タンパク質  
(10万種以上)

I. 創薬テーマ創出

医療AIと創薬AIの連結による  
相乗効果が期待

V. メディカルアフェアーズ・サプライチェーン

医療現場



疾患原因となる  
創薬標的タンパク質の同定

VI. 市販後・PMS・営業支援

II. 診断

III. 治療法決定

I. 未病・先制医療

# ライフ・インテリジェンスコンソーシアム (LINC)

## Pre-Competitive Area

課題：製薬会社どうしが自社データを持ち出して共有することは困難。これに対して、国策として、PMDAが保有するデータ（特にCDISC等）を学習データとして利活用することが期待される。

文献情報DB

アカデミア

助言  
サポート

IT系企業



モデル開発

モデル構築の  
ノウハウ蓄積  
(コード等は開  
発者に帰属)

標準モデルとしてコンソ内でのみ共有利用

## Competitive Area

代理機関制  
度等を通じた  
臨床データの  
利用に期待大



データ提供

インハウスデータ  
(プロジェクト取  
得実験データ、社  
内データ等)

インハウスデータ  
によるモデル改良



Competitive Model

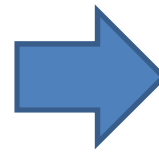
ITと製薬・化学・食品企業等によるB2B



# 世界に勝つためのAI戦略：ハイブリッド型人工知能

## 従来型人工知能

- 予測精度や探索空間が学習データの質と量に依存していた。
- 非線形予測モデルは特徴量の抽出が困難であり、因果関係の推論もほぼ不可能。

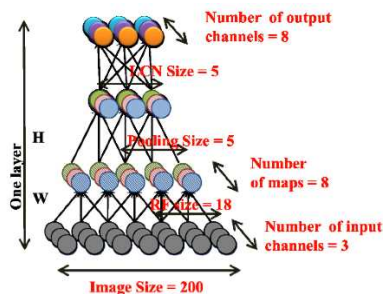


## ハイブリッド型人工知能

- シミュレーションでの学習データ生成による予測精度向上や未知空間の探索が可能
- シミュレーションとの融合による特徴量の具現化や因果推論を実現

## ハイブリッド型人工知能

### 機械学習



### シミュレーション

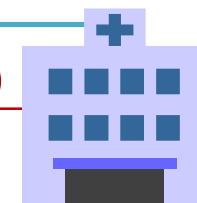


予測

データ生成

実データによる誤差補正

### ビッグデータ・リアルワールド（実験データ、医療データ等）



# 創薬AIがもたらす経済効果

開発期間：4年短縮  
開発費：業界全体で1.2兆円削減  
(1品目あたり600億円削減)



(日本製薬工業協会「DATA BOOK 2016」参考)

業界全体で1.2兆円の見積もり根拠：

1年あたり新薬承認数は平均5品目。短縮期間4年で20品目分の評価ができるはずで  
600億の削減×20品目 = 1.2兆円の削減効果