



## Drug Discovery AI Factory

# メディア・IRセミナー FRONTEO AI 「KIBIT」イノベーション事業 July 2024

# 10年を超えるAI運用実績

## 事業領域

AIソリューション事業と  
リーガルテックAI事業を展開

## 人工知能 (AI)

進化する特化型AI「KIBIT (キビット)」

有効特許数 **82件**

2003年 (創業期) ~

### リーガルテックAI事業

- ディスカバリ (証拠開示)
- フォレンジック (不正調査)



2012年~

自社開発特化型AI「KIBIT」誕生  
少量データで判断の機微を捉えるAI

**KIBIT**

米国司法省調査 / 米国民事訴訟 / 国内不正調査

実績 **10,600**件以上<sup>※</sup>

2014年~

### AIソリューション事業

- ビジネスインテリジェンス分野
- ライフサイエンスAI分野



2018年~

独自の分散表現による  
ベクトル型自然言語処理AI

2020年~

### AIソリューション事業

- 経済安全保障分野
- サプライチェーン解析
- 株主支配ネットワーク解析
- 最先端技術・研究者ネットワーク解析

2021年~

ネットワーク解析AI

ビジネスインテリジェンス分野  
ライフサイエンスAI分野 / 経済安全保障分野

**348社** 導入<sup>※, ※※</sup>  
(累積)

# FRONTEOの事業コンセプト

➤ FRONTEOは特化型AI「KIBIT」の提供を通じて、日夜社会課題と向き合う各分野の専門家を科学的に支援

## 社会課題

訴訟 不正 コンプライアンス 経済安全保障 技能伝承 病気 事故

### 解決の努力

## 専門家

弁護士

犯罪捜査官

特許分析員

創薬研究者

医師

看護師・介護士



自然言語による判断  
(会話、文書、論文)

専門家自らが  
理解する仮説生成

### 判断支援

NLP AI 特許取得済み



KIBIT+



FRONTEO

豊富な社会実装経験

マップ化する技術 特許取得済み



## 専門領域におけるAI実用化の要点



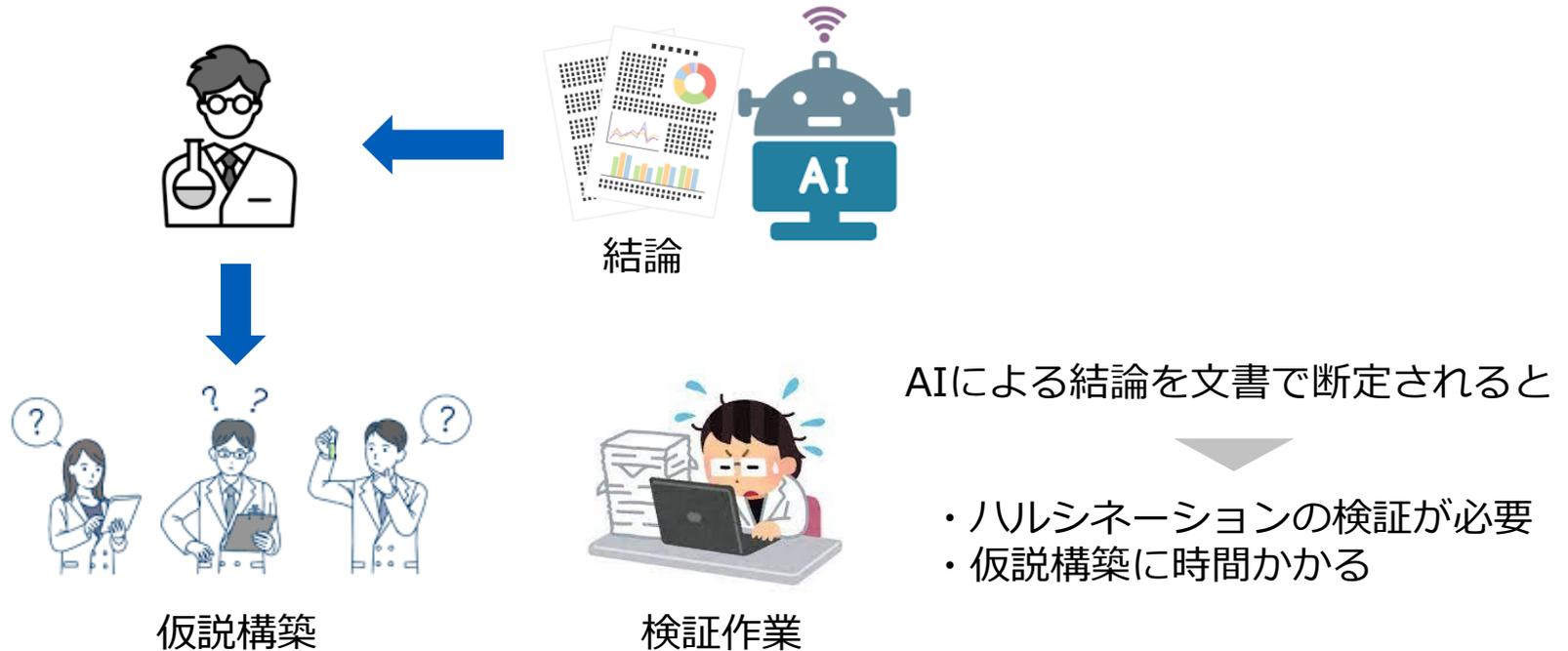
訴訟・犯罪捜査領域



医療・医薬品・ヘルスケア領域

人が理解できるような説明（仮説）が必要

## 専門家の判断支援：結論だけではなぜダメなのか？



**AIによる結論は簡単に受け入れられない**  
**AIに全て任せると、実用化できる新しい発見は困難**

# FRONTEO AI KIBIT : 特化型AI



## 特化型AI

***Serendipity*** 意味のある偶然の一致：直感

**KIBITは高度に訓練した者（専門家）専用AI**



KiBi+ はなぜ  
イノベーションをもたらすのか？

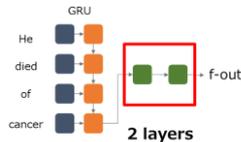


# FRONTEO 自社開発 特化型AI「KIBIT」の特徴

数学的アプローチで開発された、軽くて高性能なAI

## Green microAI

シンプルな構造



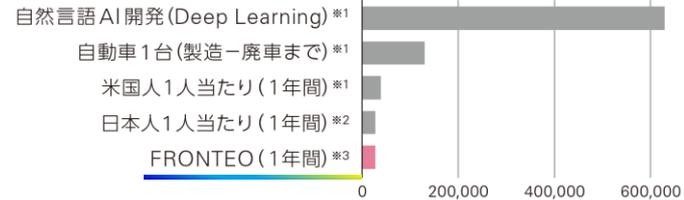
解析が速く、精度が高い



CPUレベルで解析



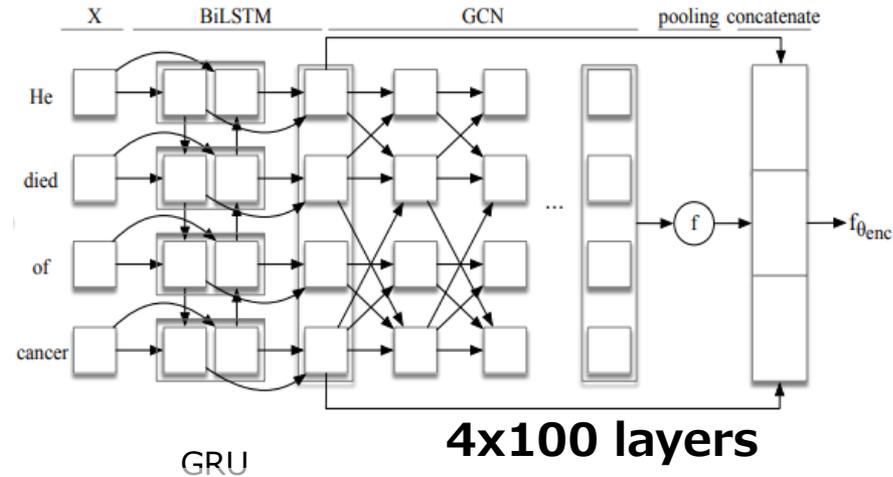
CO<sub>2</sub> 排出量の比較 (単位: lbs)



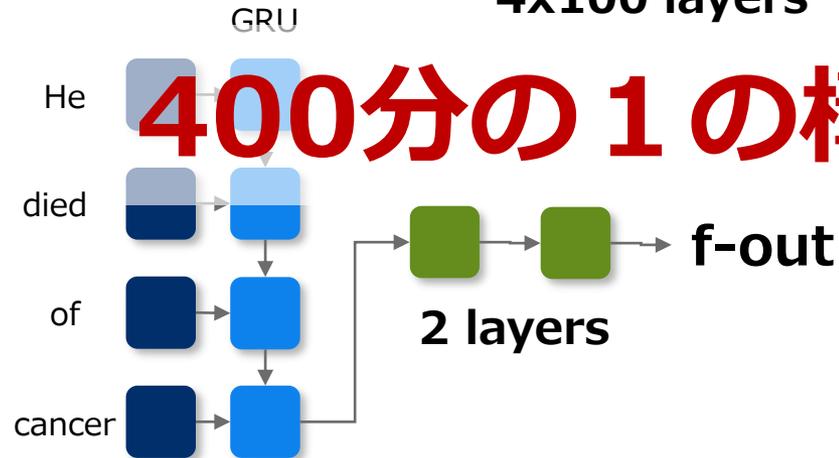
- ※1 Energy and Policy Considerations for Deep Learning in NLP, College of Information and Computer Sciences University of Massachusetts Amherst (Jun 2019) から抜粋
- ※2 日本のCO<sub>2</sub>排出量及び日本の人口からFRONTEO作成
- ※3 ※1の論文と同様の計算方法により、FRONTEO作成

# KIBITのアーキテクチャ

## GCE モデル



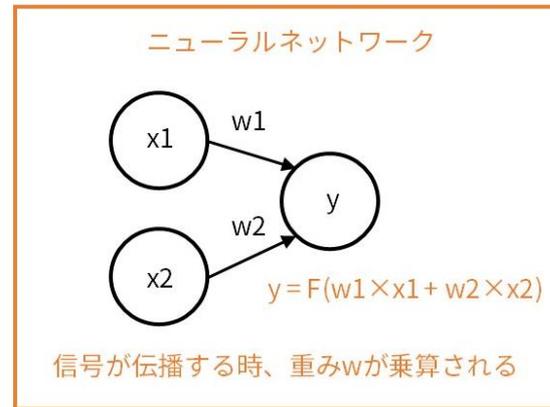
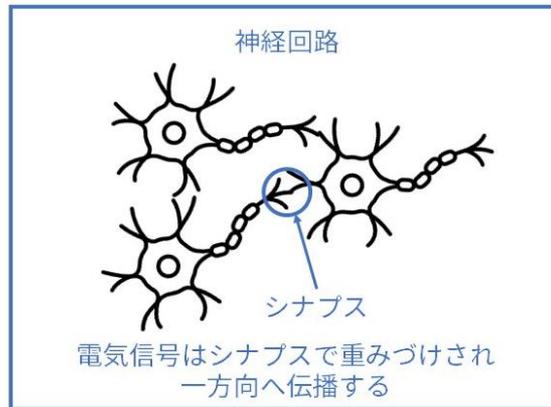
## FRONTEO モデル



# FRONTEO 特化型AI「KIBIT」設計思想① 世界の趨勢

～ニューラルネットワーク～

人間の脳内の神経細胞同士の**つながり**を再現した数理モデル



- ◆ 学習が神経細胞(ニューラルネットワーク)を育てる工程からスタート
- ◆ 計算量が膨大

## FRONTEO 特化型AI「KIBIT」設計思想② FRONTEOモデル

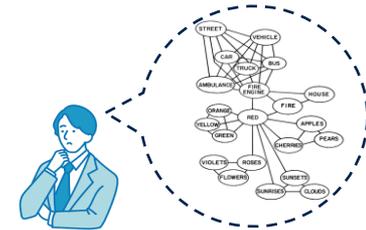
KIBIT：以下の二つの過程を再現した数理モデル

- 1 \* 人間の子供が新しい言葉を学ぶ過程
- 2 \* 人が思考を発展させる過程



### 確かな理論

- 1 \* 分布仮説[Harris, 1954]
- 2 \* 拡散活性化[Collins & Loftus 1975]



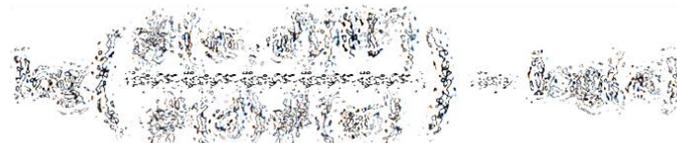
- ◆ 学習過程の再現の為、すぐに学習が始まる
- ◆ 計算量が極めて少ない
- ◆ 専門家に新たな発見（アイデア）を誘起

# FRONTEO 特化型AI「KIBIT」設計思想③ WHY KIBIT?

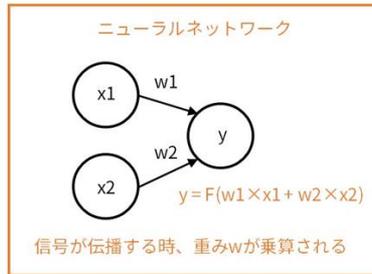
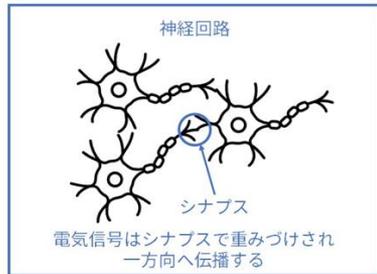
専門家の**脳力をブースト**し、**新しい発見を促す**  
日米特許技術

1	<b>分布仮説</b> <b>Distributional hypothesis</b>	<b>独自の</b> <b>分散表現技術</b>
2	<b>拡散活性化</b> <b>Spreading Activation</b>	

$$DW = D * W^t = \begin{pmatrix} dw_{11} & dw_{12} & \dots & dw_{1n} \\ dw_{21} & dw_{22} & \dots & dw_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ dw_{m1} & dw_{m2} & \dots & dw_{mn} \end{pmatrix}$$

$$\vec{w} \cdot \vec{d} =$$


## WHY KIBIT?: 自社開発特化型AIエンジン「KIBIT」の設計思想



脳(神経ネットワーク)を  
創造する



人の学び、思考過程を  
創造する

KiBi+

## WHY KIBIT?: FAQ よくある質問

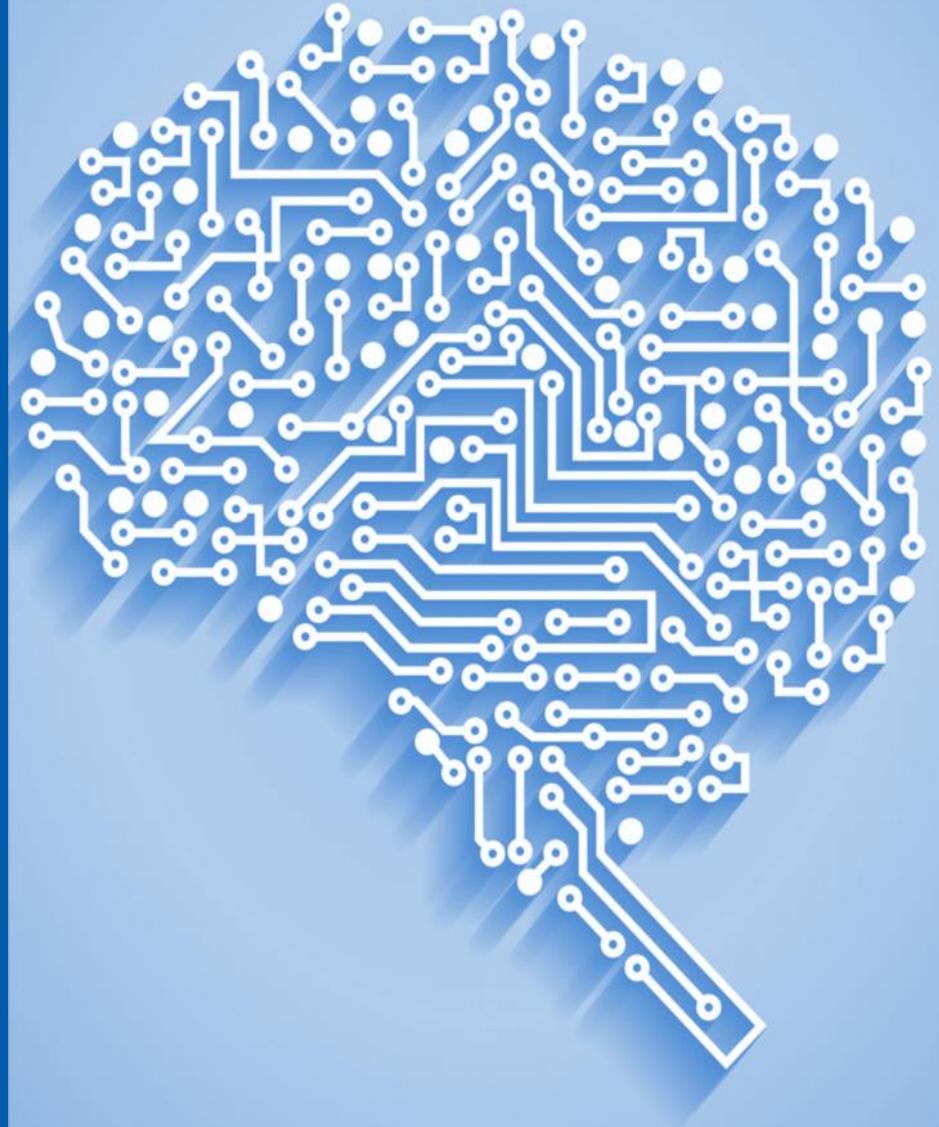
Q : あやしい論文が含まれていれば、精度が悪くなるのでは？

Q : 論文解析では、新しい発見は無理では？

Q : 定量的分析手法のほうが優れているのでは？  
(例 QSP(定量的システム薬理学) 等)

**答え : NO !**

# 非連続成長事業



## 経済安全保障事業

### KiBi+ Seizu Analysis

#### サプライチェーン ネットワーク解析

chokepointや懸念  
組織との直接・間接の  
取引関係を検知



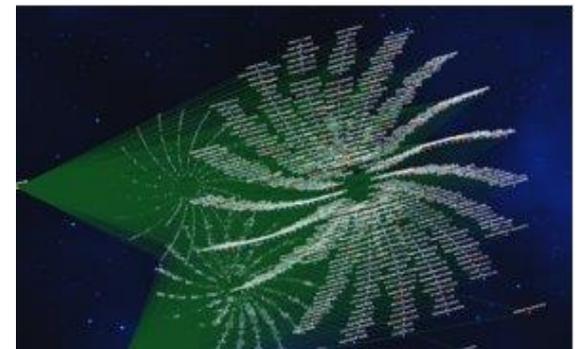
#### 株主支配 ネットワーク解析

持ち株ネットワークを  
通じた間接的な  
支配を検知



#### 研究者 ネットワーク解析

共著関係に基づき  
研究者間の  
ネットワークを解析



# AI医療機器領域：非医療機器・産業横断アライアンス、海外市場展開

## 日本市場

FRONTEO  
独自技術

標的疾患

医療機器

非医療機器

例) 産業横断アライアンス

インパクト

・特許  
・独自技術

認知症

うつ病

ADHD

統合失調症

etc....

FRONTEO  
×  
SHIONOGI

FRONTEO  
×  
SHIONOGI

一定の研究成果が  
確認され次第展開

一定の研究成果が  
確認され次第展開

銀行

地銀ネットワークを  
通じた網羅性の確保

保険  
(生命保険)

健康促進支援を  
通じた新たな保険設計

通信

キャリアの新たなサー  
ビス開発 (スマホ)

自動車

スマートシティを視野  
に入れた新たなMaaS  
サービス

産業X

-

社会実装

海外市場 A

海外市場 B

海外市場 C

## AI創薬領域： 医薬品開発産業イノベーション



# 医薬品開発プロセスの課題

## 1. 非効率的な活動実態（すでに破綻...）

製薬企業、バイオベンチャー、アカデミアが個々に活動

優先順位づけが困難。抜け漏れ、重複が発生  
他社、他国の着目するターゲット、個々のカバー範囲が不明瞭

## 2. 優先的に改善すべき工程が手つかず

最も重要な最上流工程である仮説生成のAI活用ができていない



論文解析のハードルが高い  
従来AIアプローチは撤退

## 財政と安全保障に大きなダメージ

- ✓ 新薬開発の成功確率が低く、失敗分のコストが転嫁され価格高騰
- ✓ 新薬を海外からの輸入に依存
- ✓ 希少疾患に対する新薬開発への投資抑制
- ✓ 効果的な新薬が国民に行き渡らない

## 解決案① Drug Development Direction Center構想

1. 国家戦略に基づいた、新薬・既存薬再開発・ワクチン・薬価改善方針
2. 方針に基づいた、標的分子、適応症、化合物等の洗い出し、及び企業・アカデミア・研究者等の特定
3. プロジェクト開始～完了までDD-AIFのOSINT仮説検証システムによる伴走

事象やそれに対する活動の全体像を把握  
 優先順位を付けた戦略的な対策  
 得意分野を生かした分業や協業

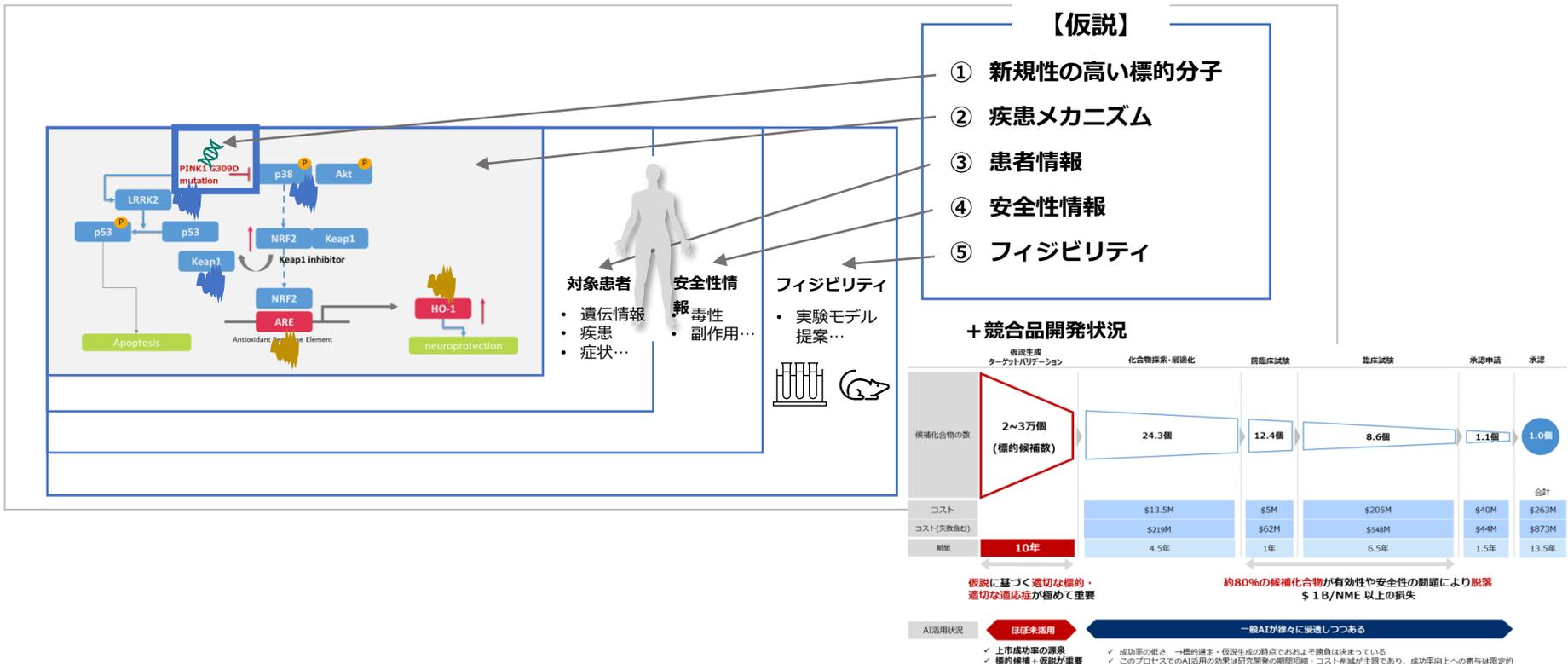


製薬企業・大学研究機関・バイオベンチャー…

## 解決案② 仮説生成能力の向上

### 医薬品開発の設計図ともいえる仮説生成

- 疾患関連性が未報告の、新規性の高い標的分子
- 網羅的かつアンバイアスに関連性の高いつながりで構成される疾患メカニズム
- 一般的な統計・解析技術と補完的に組み合わせることが可能

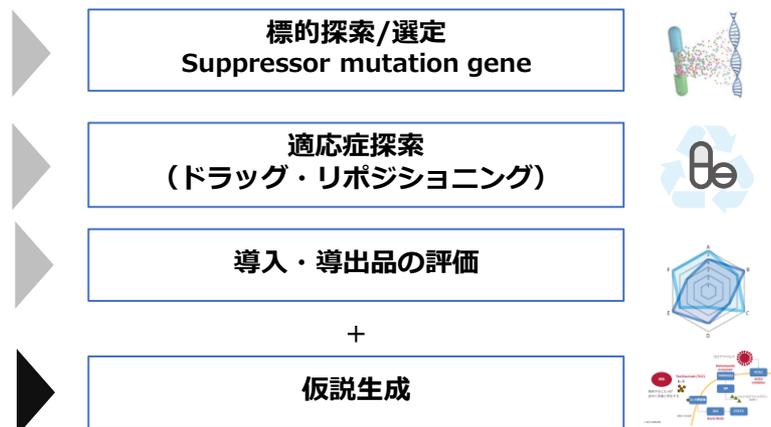


# 医薬品開発産業におけるイノベーション ～Drug Discovery AI Factory～

- 高度化された仮説生成により医薬品開発成功確率の大幅な改善
- インテリジェンス機能により中央情報指揮所として医薬品開発のプロセスを変革



活用領域





# ステージ4

## FY23 主な経営指標の増減

売上高  
7,375百万円  
(前期比 : +2%)

EBITDA  
695百万円  
(前期比増減 : +1,105百万円)

営業利益  
▲185百万円  
(前期比増減 : +1,176百万円)

当期純利益  
▲2,843百万円  
(前期比増減 : ▲1,142百万円)

FCF  
1,528百万円  
(前期比増減 : +3,067百万円)

# ステージ4

引き続きAIソリューション事業をステージ4の成長事業と位置づけ、分野別の成長種別（リニア/非連続）とその成長因子を明確にし、経営資源の選択と集中を徹底



# KIBITの独自性と 新たな文献検索の幕開け

FRONTEO CTO  
Hiroyoshi Toyoshiba, Ph.D.

文章情報探索で、記述されていない  
関連性を探すことは可能か？

もし確立できれば、これまでの常識  
を根底から覆すことになる。

既知情報から未知情報を推定

= 「創造性」

# 考えられる応用先

## 創薬研究



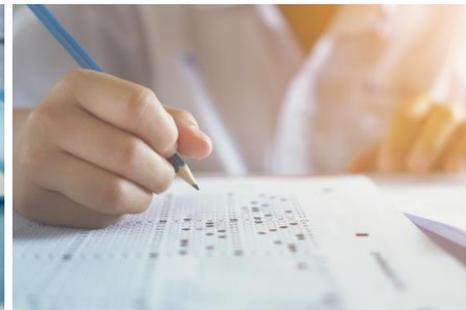
新規標的/バイオマーカー探索

## ヘルスケア



食事と健康の関係

## 教育



学習内容

■ ■ ■ ■ ■ ■

# 製薬企業の企業戦略

## 企業戦略

- **First in class**を狙う場合
- 出口戦略として**導出**を考える場合

他社との**差別化**が必須

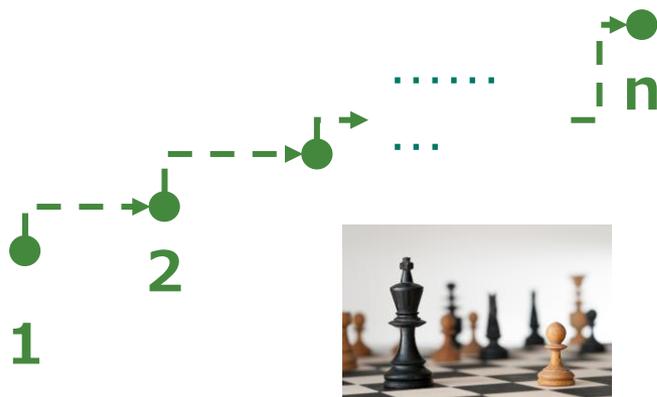
**新規性**の高い標的が必要



# 文章情報からの新たな発見とは？

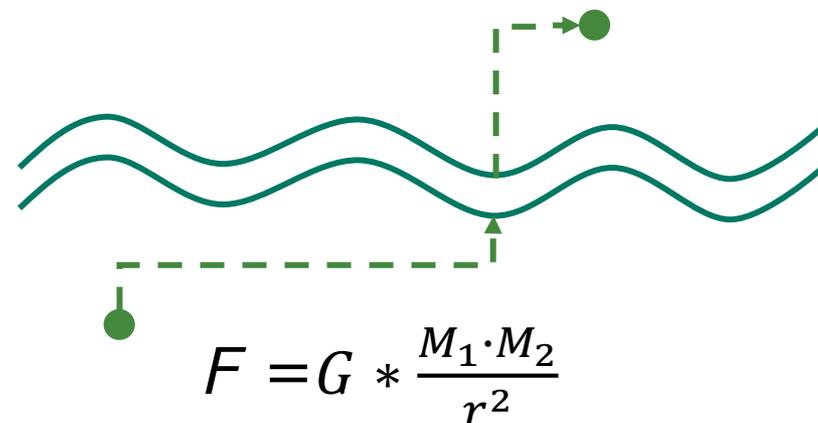
## 連続的

連続的なつながりであるが、nが非常に大きいと、1からnの想起は難しい



## 非連続的

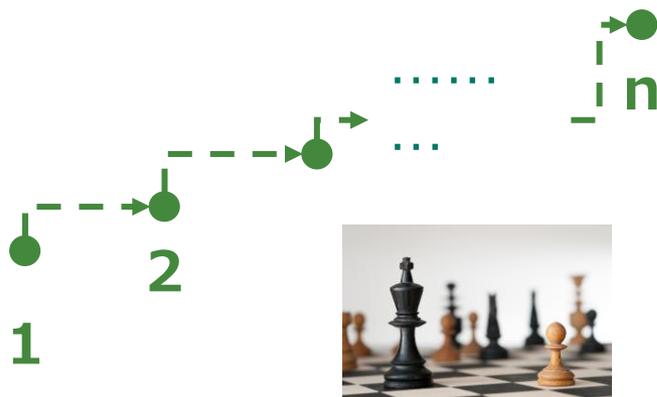
1からnはそもそも連続的なつながりで想起できない



# 連続的な発見

## 連続的

連続的なつながりであるが、nが非常に大きいと、1からnの想起は難しい



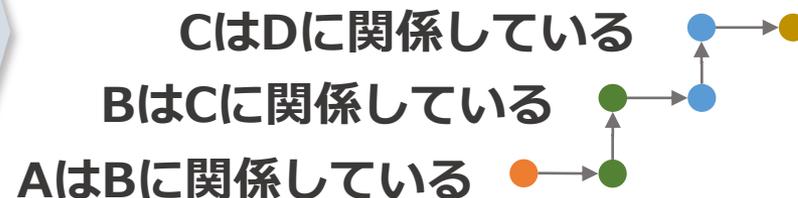
AはDに関係している



CはDに関係している

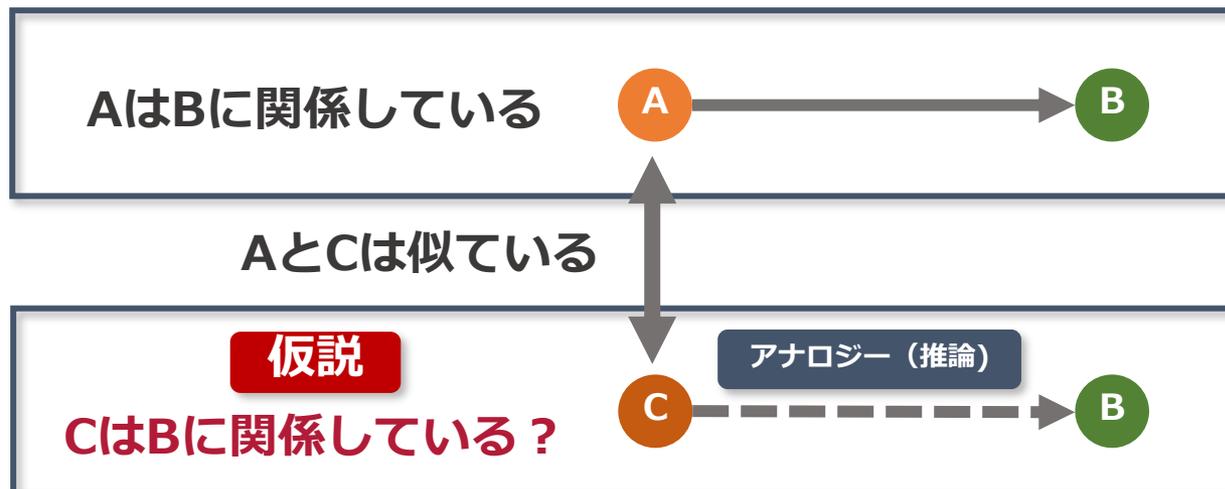
BはCに関係している

AはBに関係している



世の中のNLPの100%が、連続的な発見にとどまっている

## 非連続的な発見の例



“CはBと関係している”ということは、  
どこにも記述されていない

非連続な発見 = 記述されていない関連性を見つけ出す



# 記載のない関連を見つけ出す 文献検索アルゴリズム



# SPRINGER NATURE

## Advancing discovery

発見を促進するためにシュプリンガーネイチャーができること

シュプリンガーネイチャーは、発見の進展に貢献するために、信頼性が高く洞察に富む研究の出版を通して新領域の知識の成長を支えるとともに、アイデアや情報への世界からのアクセスを可能にし、またオープンアクセスの推進を牽引します。

その実現に向けて、シュプリンガーネイチャーは研究コミュニティ全体に最良のサービスを提供することを目指します。

- ✓ 著者が発見した知見を共有できるようにすること
- ✓ 他研究者の研究成果を見つけ、アクセスし、理解できるようにすること
- ✓ 図書館員や研究機関をテクノロジーやデータのイノベーションを通して支援すること
- ✓ 高品質な出版サポートにより学会を支援すること
- ✓ 政府や資金配分機関のにとっての重要な課題に真摯に取り組むこと



### Value of Publishing

シュプリンガーネイチャーと出版する価値



# SPRINGER NATURE

FRONTEOだけが活用できる独自ソース

発見の進展に貢献するために、  
信頼性が高く洞察に富む研究の出版

SPRINGER NATURE

600雑誌、25年分、フルテキスト

センテンス数：約3億5000万

2回以上出現単語：約1260万単語



FRONTEOだけの独自エンジン **Kibi+**

忠実に分散表現を再現し、  
Spreading Activationで新たな関連を抽出

**Kibi+**

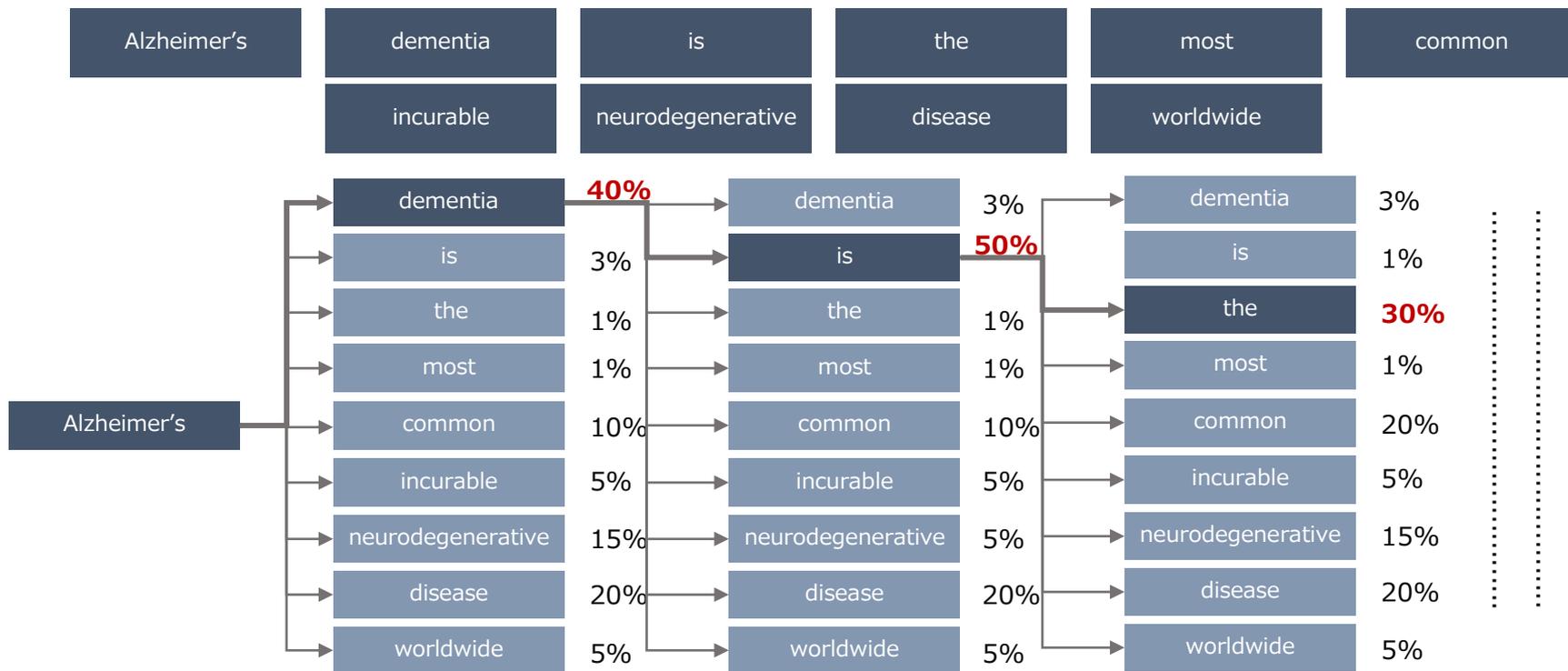
独自の分散表現エンジン

精度の高い概念ベクトル

**Kibi+** Cascade Eye

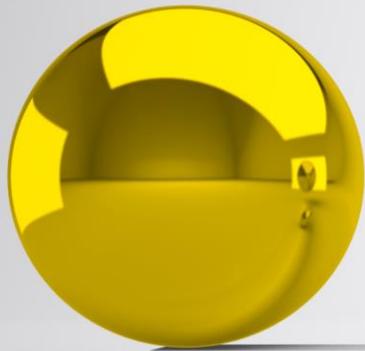
# 生成AIでも不可能なのか？

Alzheimer's dementia (AD) is the most common incurable neurodegenerative disease worldwide



**記述内容が間違っているとしてもそれを学習すると生成してしまう、  
記述が間違っているか否かよりも、表現が自然か否か**

人の意識を超えていく = 法則に従って決める



## 2つの重要な定理

1

**分布仮説**  
**Distributional Hypothesis**

2

**拡散活性化**  
**Spreading Activation**

# Distributional Hypothesis

"a word is characterized by the company it keeps"

Harris, Z. (1954). Distributional structure. *Word*, 10(23): 146-162.

Firth, J.R. (1957). A synopsis of linguistic theory 1930-1955. In *Studies in Linguistic Analysis*, pp. 1-32.

In recent years, **the distributional hypothesis** has provided the basis for the theory of similarity-based generalization in language learning: the idea that **children can figure out how to use words they've rarely encountered before by generalizing about their use from distributions of similar words.**

**分布仮説**は、言語学習における類似性に基づく汎化理論の基礎を提供した。つまり、子どもは類似した単語の分布からその使用法を一般化することで、ほとんど出会ったことのない単語の使用法を見つけ出すことができるという考え方である。

Gleitman, Lila R. (2002). **Verbs of a feather flock together II**. The child's discovery of words and their meanings. Published online: 22 November 2002. DOI logo.

Yarlett, D. (2008). *Language Learning Through Similarity-Based Generalization* (PDF) (PhD thesis). Stanford University. Archived from the original (PDF) on 2014-04-19. Retrieved 2012-07-12.

# Distributional Hypothesis

1

単語は周りの単語で特徴付けられる  
単語の意味は、周りの単語の種類と頻度  
のみ決まる

文章の正否を学んでいない  
単語の意味を規定しているだけで、それは  
事実である

2

子供が新しい単語の意味を学ぶという考え方

滅多に出てこない単語の使い方を学習する  
発見を行うことを可能にする

3

まったく異なる論文でも周りの単語の種類と  
頻度が近ければ、関係性を見出せる

人間の認知レベルを超えた関連性の発見

KIBITは分布仮説をベースにしている

Distributional Hypothesis  Kibit+

こどもが、ほとんど出会ったことのない単語の使用法を見つけ出す（学習する）ことができるという考え方である

# 分布仮説の下でAとCが似ているとは？



Aの周りが出てくる単語の種類と頻度

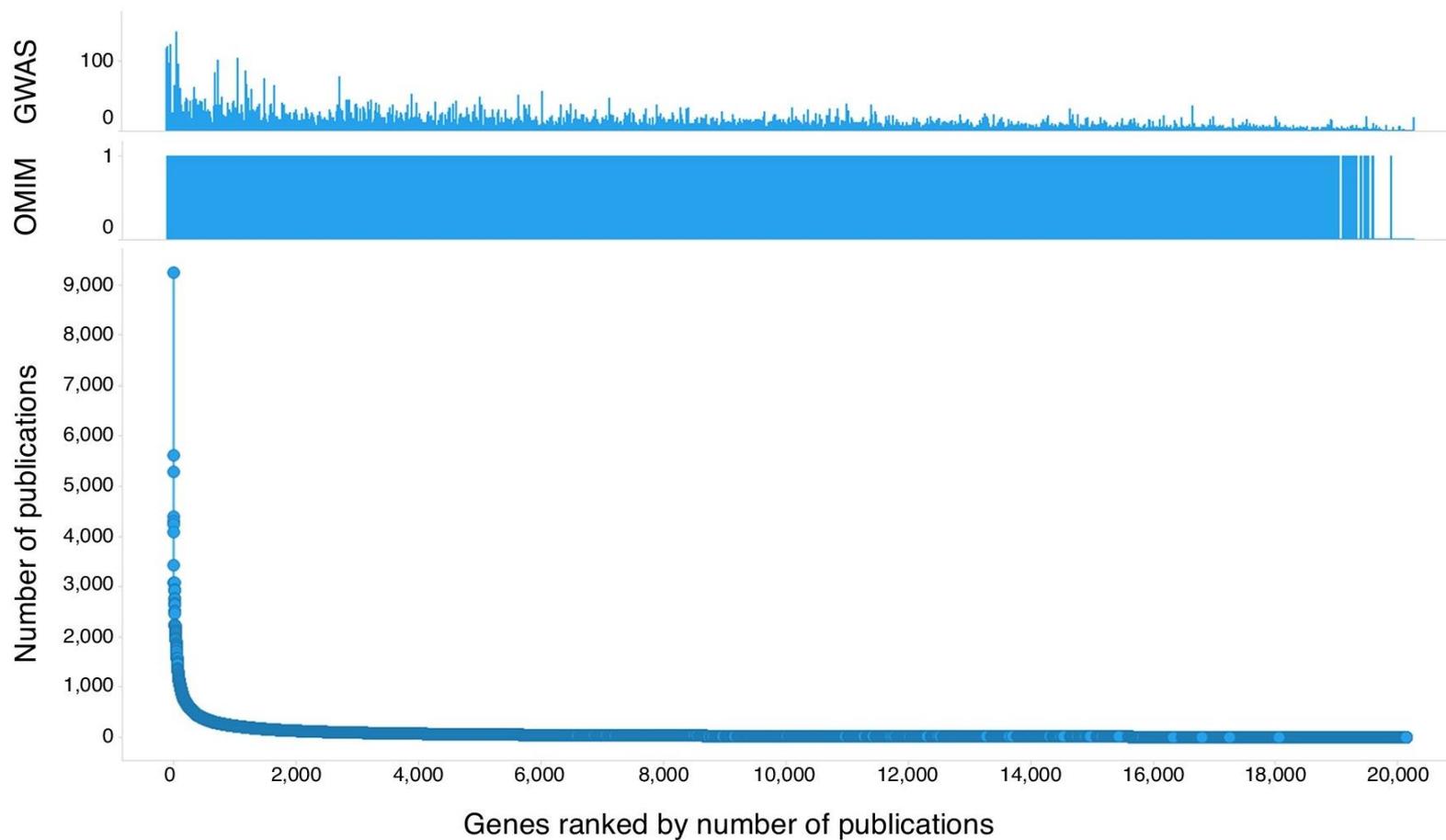
≒

Cの周りが出てくる単語の種類と頻度



プリンペラン	類似度*
プリンペラン	1.000
むかむか	0.768
ムカムカ	0.748
エリー	0.739
テン注	0.733
吐き	0.721
エリーテン	0.713
吐剤	0.703
制	0.701
吐き気	0.693
吐く	0.690
<b>まぐまぐ</b>	<b>0.687</b>
吐い	0.684
むかつき	0.663
時制	0.661
...	...

# バイオメディカル領域におけるバイアス



# Confirmation Bias

Confirmation Bias is the tendency to search for, interpret, favor, and recall information in a way that confirms or supports one's prior beliefs or values[1].

**確証バイアスとは、認知心理学や社会心理学で取り上げられるバイアスの一つ。自分の思い込みや願望を強化する情報ばかりに目が行き、そうではない情報は軽視してしまう傾向のことを指す**

[1] Nickerson 1998,  
pp. 175-220

研究者に尋ねてみて欲しい

「疾患との関連性が未報告\*な標的を提案してください」

\* PubMedでno hits



# Kibi+

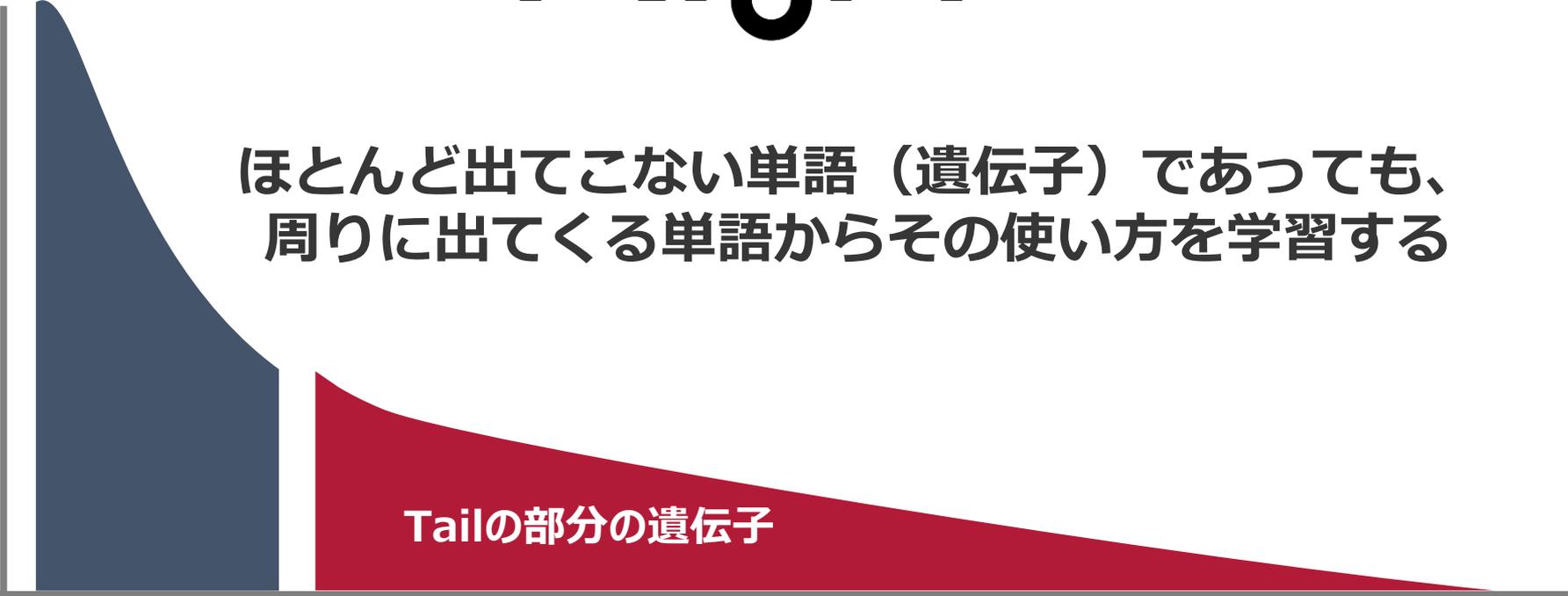
分布仮説は、言語学習における類似性に基づく汎化理論の基礎を提供した。つまり、子どもは類似した単語の分布からその使用法を一般化することで、ほとんど出会ったことのない単語の使用法を見つけ出すことができるという考え方である。



ほとんど出てこない単語であっても、周りに出てくる  
単語からその使い方を学習する

# Kibi+

ほとんど出てこない単語（遺伝子）であっても、  
周りに出てくる単語からその使い方を学習する

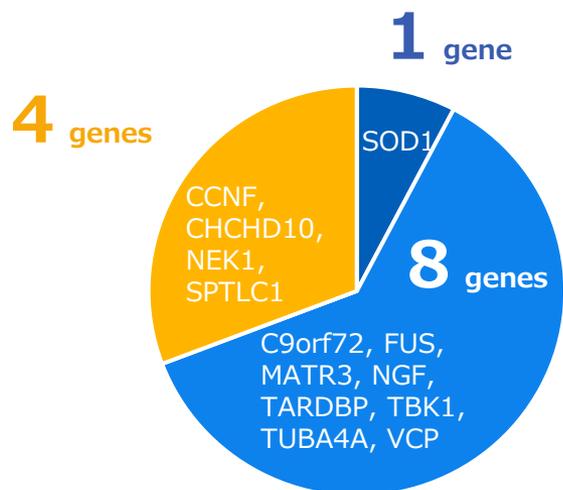


Tailの部分の遺伝子



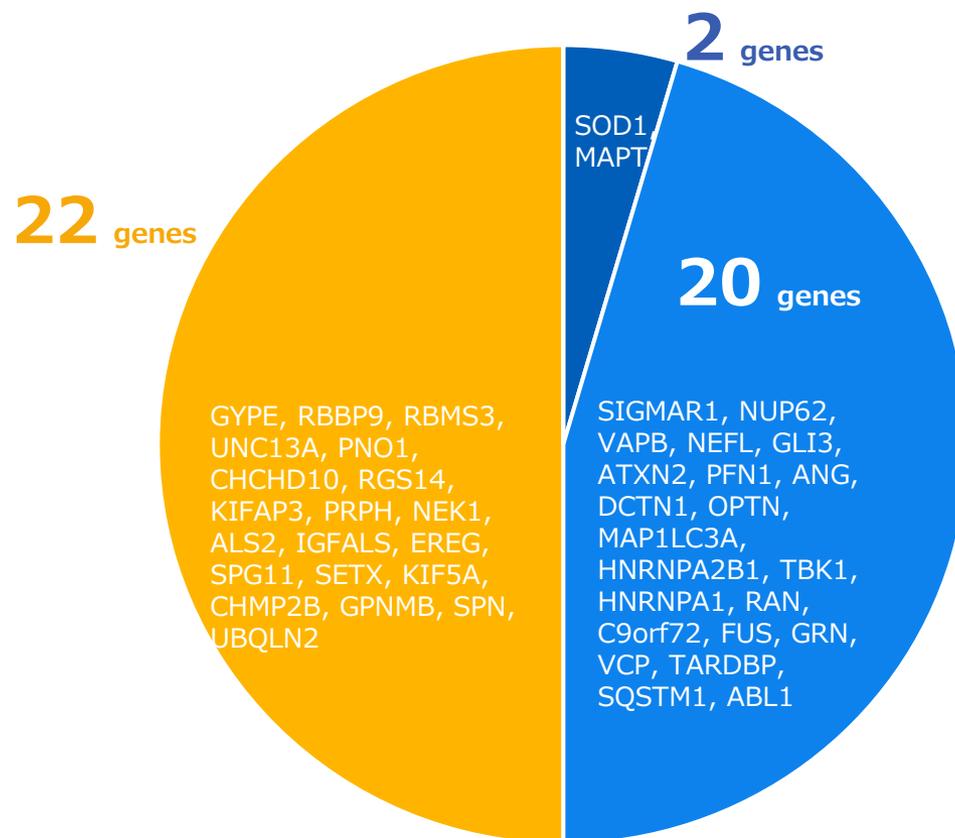
# 関連性の高い分子に出会う機会が増える

## PubMed Best match



- ultra major genes (1000- PMID counts)
- major genes (100-999 PMID counts)
- minor genes (10-99 PMID counts)

## KIBIT Amanogawa Discovery



There were no genes categorized into ultra minor genes (1-9 PMID counts).

## 2つの重要な定理

1

**分布仮説**  
**Distributional Hypothesis**

2

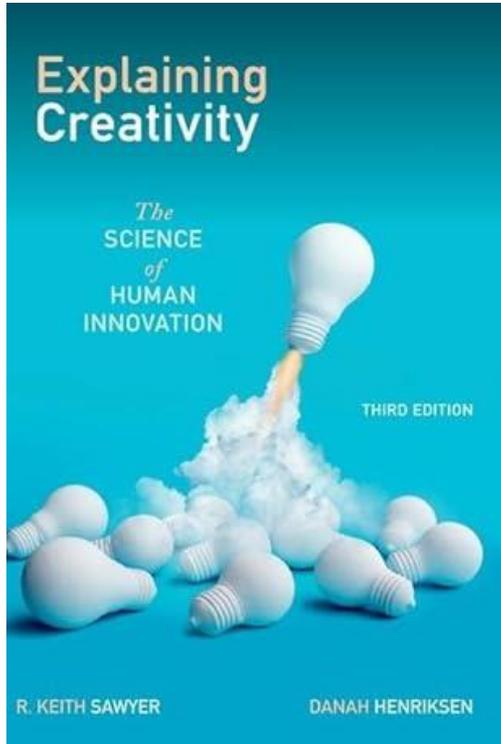
**拡散活性化**  
**Spreading Activation**

# Spreading Activation Theory

Spreading activation is a term used in psychology, psycholinguistics and cognitive linguistics. It refers to a model describing processes thought to be taking place during language production and comprehension.

拡散活性化とは、心理学、心理言語学、認知言語学で使われる用語で、**言語生成**や**言語理解**の過程で起こっていると考えられているプロセスを説明するモデルを指す

# Explaining Creativity : 創造性に必要な5つの要素



1. You continue consciously working on the problem.
2. Rest.
3. Selective forgetting.
4. Random subconscious recombination.
5. Spreading Activation.

# 2,3,5は実験で確かめられている

# 創造性を誘発する

Spreading activation model accounts for associative priming by assuming that **words which are semantically associated with each other are represented in the form of a network** and activation spreads through this network from presented words to their associated words in memory (Anderson, 1995).

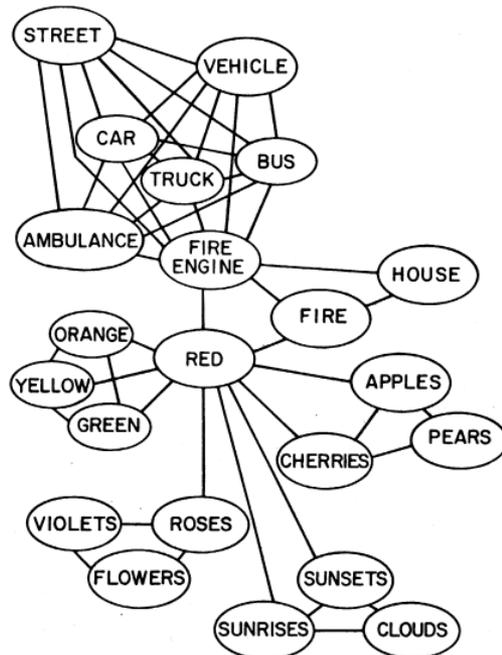


FIGURE 1. A schematic representation of concept relatedness in a stereotypical fragment of human memory (where a shorter line represents greater relatedness).

概念によって結びつけられた 概念空間に分布する  
複雑な関連ネットワーク上に構築されている



分散表現でベクトル化し（概念を埋め込んだ）  
コサイン距離の近さで結び付けた



KiBi+ Cascade Eye

Anderson, J. R. (1995). Cognitive psychology and its implications (4th ed.). New York: W. H. Freeman.

# 連続性と非連続性の関連

Gene	PubMed	人工知能のスコア
CYP2E1	586	0.75748
CYP3A4	171	0.80779
ABCB11	89	0.95925
UGT1A1	69	0.75419
CYP7A1	63	0.72498
DDC	47	0.95564
CYP2B6	45	0.88155
HNF4A	26	0.89395
ABCB4	23	0.81159
CYP8B1	22	0.87051
.....	.....	.....
.....	.....	.....
.....	.....	.....
.....	.....	.....
<b>MAT2A</b>	<b>1</b>	<b>0.70161</b>
<b>ADH4</b>	<b>0</b>	<b>0.91</b>
<b>ZFYVE19</b>	<b>0</b>	<b>0.74481</b>

## PubMedのカラム

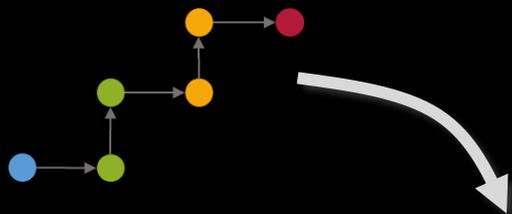
“drug-induced liver injury” と “遺伝子名”  
をPubMedで検索した時の論文のヒット数

ヒット数が0：非連続的な関連性

ヒット数が多いものは、連続的な関連性

# 概念によって結び付けられた分子間ネットワーク

連続的な遺伝子のつながり



Gene	PubMed	cos_sim	causality score
CYP2E1	586	0.404093	0.75748
CYP3A4	171	0.255596	0.80779
ABCB11	89	0.409593	0.95925
UGT1A1	69	0.314609	0.75419
CYP7A1	63	0.295381	0.72498
DDC	47	0.399839	0.95564
CYP2B6	45	0.256335	0.88155
HNF4A	26	0.152759	0.89395
ABCB4	23	0.381347	0.81159
CYP8B1	22	0.358319	0.87051
.....	.....	.....	.....
MAT2A	1	0.209926	0.70161
ADH4	0	0.152856	0.91
ZFYVE19	0	0.291886	0.74481

非連続的な遺伝子のつながり

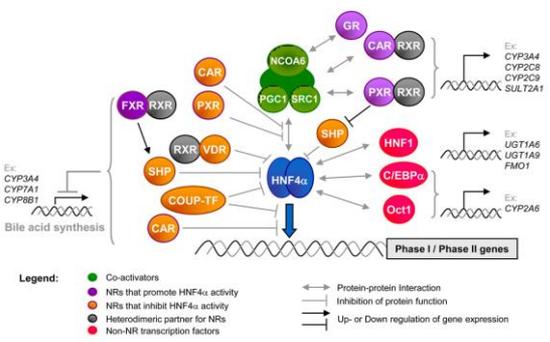
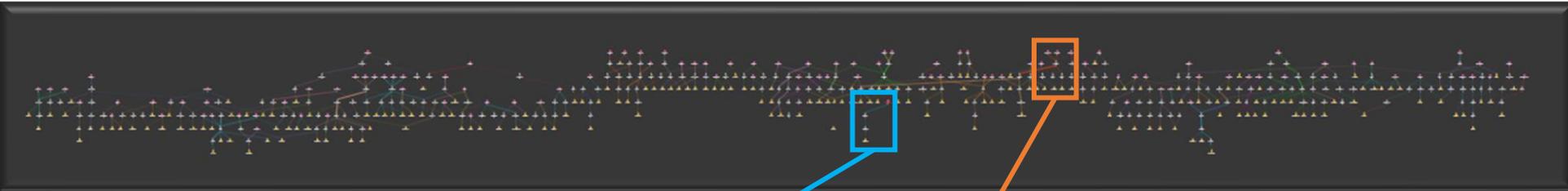


KiBi+  
Cascade Eye

Spreading Activation!



## Network for drug-induced liver injury



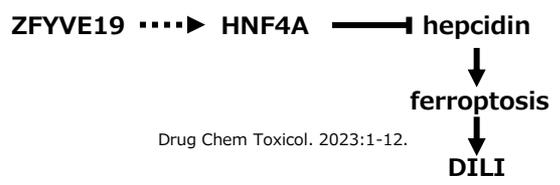
**Legend:**

- Co-activators
- NRs that promote HNF4 $\alpha$  activity
- NRs that inhibit HNF4 $\alpha$  activity
- Heterodimeric partner for NRs
- Non-NR transcription factors

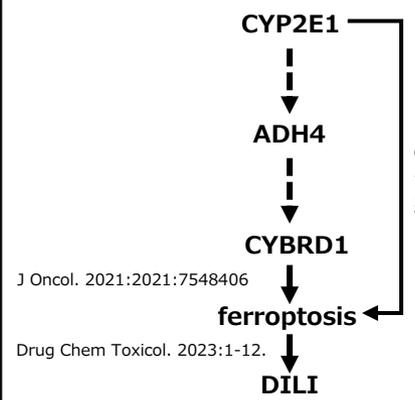
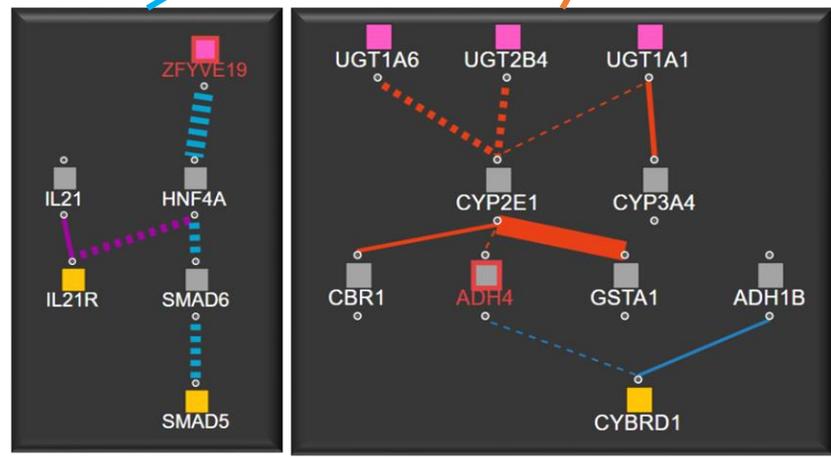
← Protein-protein interaction  
 ⊣ Inhibition of protein function  
 ⊢ Up- or Down regulation of gene expression

Curr Opin Pharmacol. 2010;10:698-705.

Biol Trace Elem Res. 2017;176:294-304.



Drug Chem Toxicol. 2023;1-12.



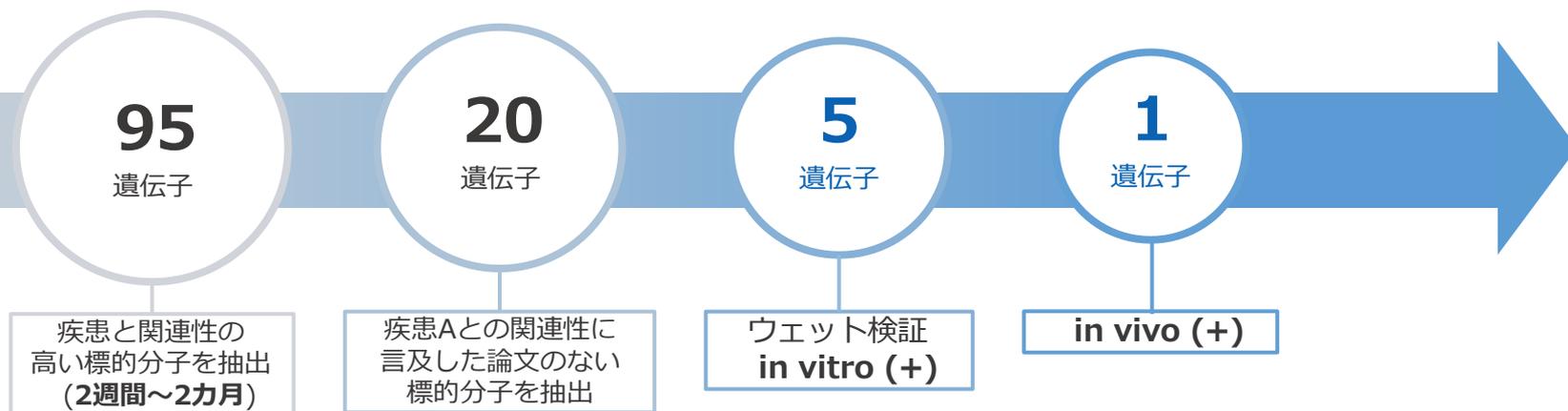
J Oncol. 2021;2021:7548406

Drug Chem Toxicol. 2023;1-12.

Cancer Med. 2021;10(22):8100-8113.

# 論文に報告のない関連性を発見し、実験で検証

## 新規標的探索で 効果実証済



非常に **新規性の高い標的候補** を見出すことのできる技術

## FRONTEOだけが見つけ出せる標的分子



非構造データから非連続的な発見を  
体系的に実現する





この資料は下記により作成されました。

株式会社FRONTEO

108-0075 東京都港区港南2-12-23 明産高浜ビル

TEL : 03-5463-7577

FAX : 03-5463-7578

東証グロース上場

本資料は、株式会社FRONTEOの承諾なしに貴社以外の第三者の閲覧に供することは出来ません。  
本資料のいかなる部分も、一切の権利は株式会社FRONTEOに帰属しており、電子的または機械的な方法を問わず、  
いかなる目的であれ、無断で複製・転送されることは固くお断りいたします。

Confidential