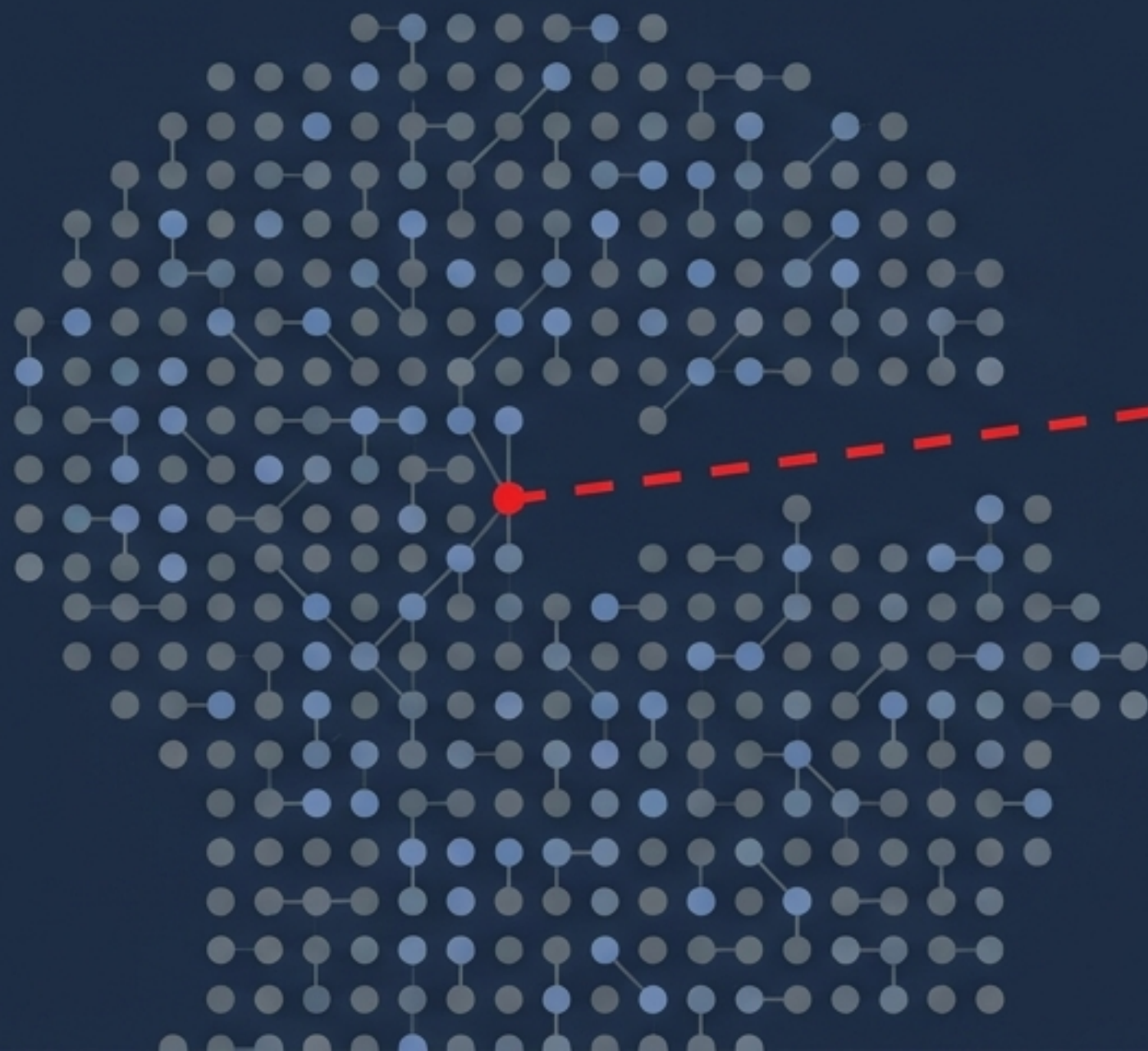


知識のギャップを埋めるAIの真価： Sony AI「Hakken」の深掘り調査

誇大広告を解体し、154万の仮説と細胞実験の現実を診断する



Target Audience Note: 本資料は、メディアの「AI創業の誇大広告」を見抜き、テクノロジーの真の価値と限界を冷静に見極めたいR&D責任者、バイオインフォマティクス研究者向けに構成されています。

Hype vs. Reality：ニュースの裏側にある正確な科学

~~AIが未知の老化遺伝子を新発見し、その老化機能を証明した~~



誤解：発見成功率66.7%（3件中2件成功）

AIが老化関連の探索空間から、従来Hakkenの知識グラフに記録されていなかった2つの『機能的遺伝子関係』を予測し、限定的な細胞実験で支持する結果を得た。



[TARGET]：既知の老化候補遺伝子



[NOVELTY]：ノード（遺伝子）ではなくエッジ（関係性）



[EVIDENCE]：細胞株（HepG2/THP-1）での薬理実験のみ。老化表現型は未検証

What is Hakken? : 未来の発見を予測する統合システム



Temporal Knowledge Graph 時系列知識グラフ

科学文献を静的文書ではなく「時間とともに変化するシステム」として処理。1990年、2020年など「知識のカットオフ」時点での歴史的バックテストを可能にする構造。



THiGERLLM & PU Learning LLM統合とPU学習

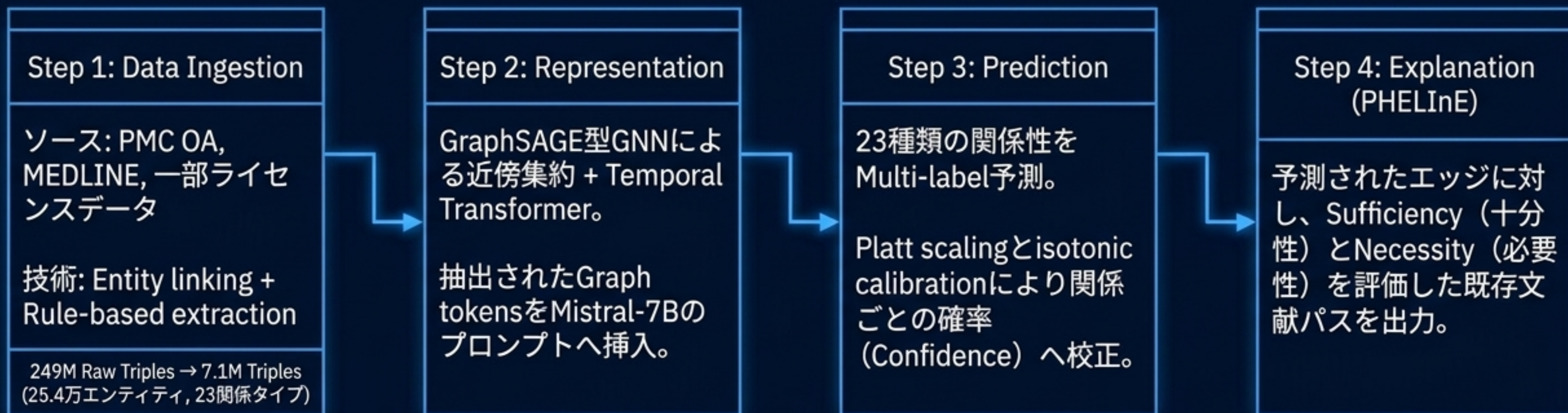
Mistral-7B-Instructの意味情報とGNNを統合。「文献に書かれていない=存在しない (Negative)」ではなく、「未発見かもしれない (Unlabeled)」として扱う Positive-Unlabeled (PU) Learningを採用。



PHELIInE Explanation 説明経路の生成

予測スコアに対し、GraphSAGEサロゲートモデルを用いて、既存の知識グラフ上に存在する「マルチホップ説明経路」を事後的に付与するメカニズム。

The AI Architecture: 予測生成のパイプライン



The Funnel of Selection Bias : 154万仮説から3件への道のり

1,543,297 仮説

専門家選定の1,386老化関連エンティティに対する予測総数。

2,804 件 (Top-k)

既存関係を除外し、各エンティティの上位3候補に機械的絞り込み。



**Human Selection Bias
(人間の選択介入)**

ランダム抽出ではなく、専門家が
Confidence > 0.8や主観的な生物学的関心
(Interestingness) を加味して選抜。

3 件

最終的な細胞実験
(Wet-lab) の検証対象

数学的警告

「2/3が検証に成功したからAIの発見率は66.7%」
という解釈は統計的に無効。
 $P(\text{Wet-lab support} \mid \text{AI generated})$ の真の確率
は、強い選択バイアスにより推定不能。

The "Unknown" Genes : 予測された対象の正体



TP53

Tumor protein p53

[NCBI: 7157]



BAMBI

BMP and activin
membrane-bound inhibitor

[NCBI: 25805]



RAF1

RAF proto-oncogene

[NCBI: 5894]



TNF

Tumor necrosis factor

[NCBI: 7124]

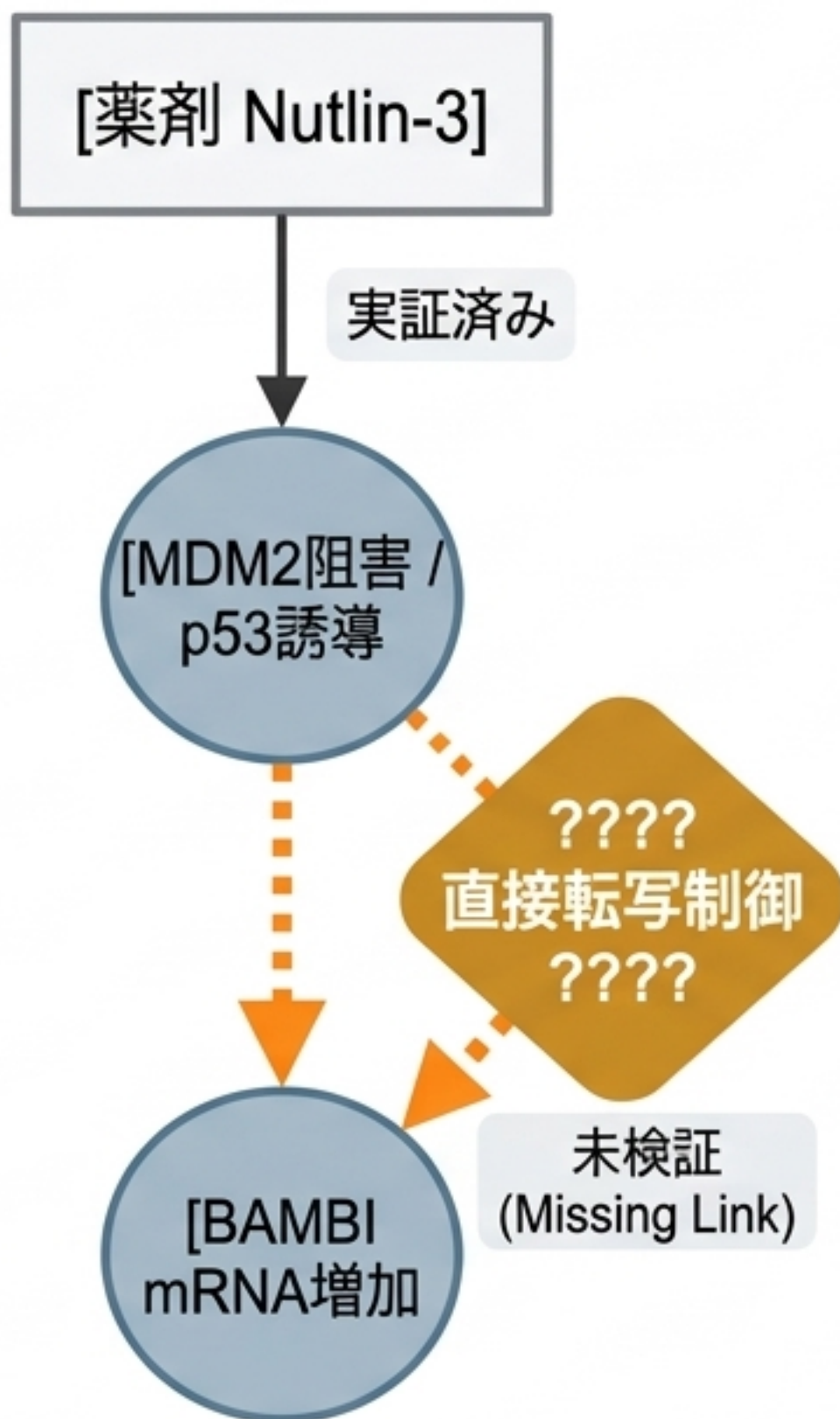
再定義されるNovelty : Hakkenが発見したのは人類にとって未知のノード（遺伝子）ではない。
既存のReference Knowledge Graph上に存在しなかった「エッジ（関係性）」である。

Evidence Matrix：3つの仮説の実証評価一覧

仮説 (Confidence)	初回 / 確認細胞実験	タンパク質 / 直接機構	老化表現型の 確認	総合評価 (Verdict)
TP53 → BAMBI (0.85391)	● mRNA +16～38%	● 直接結合未証明	● 未実施	中程度 (Supported)
RAF1 → TNF (0.83114)	● mRNA一時増、 AZ628で減	● 分泌変化なし / pERK逆説的活性化	● 未実施	中～低 (Limited)
SOAT1 → STAT3 (0.84183)	● 変化なし	● 変化なし	● 未実施	否定的 (Denied)

🔑 Key Takeaway: 論文のWet-labは「機能的遺伝子発現」を検証したものであり、細胞老化や寿命 (Aging Phenotypes) そのものへの因果的影響は未検証である。

Deep Dive 1: TP53-BAMBI (機能的関連と残る課題)



観測事実

HepG2細胞にてNutlin-3 (10 μ M) 投与。24hでBAMBI mRNA約16%、48hで約38%上昇。

決定的に欠けているエビデンス

TP53 CRISPR knockout/rescue実験、およびBAMBIプロモーターへのp53直接結合を示すChIP-seq/CUT&RUN。

臨床的意味の反転 (Implication)

BAMBIはTGF- β を抑制する。p53がBAMBIを直接「活性化」するか「抑制」するか（因果の方向性）によって、がんや線維化への臨床的意味が完全に反転する。

Deep Dive 2: RAF1-TNF (逆説的活性化と複雑な結果)

The Claimed Support (mRNA)

- RAF1阻害剤 GW5074投与で、THP-1細胞のTNF mRNAが有意に増加。
- 但し：別のpan-RAF阻害剤 AZ628では一時的に逆方向 (低下) へ変化。

The Complexity (Protein & Pathways)

- Secreted TNF：ポジコンは上昇したが、RAF阻害剤群は定量下限 (<15.6 pg/mL) 未満で無変化。
- Paradoxical ERK Activation：阻害剤投与により、かえってpERKが逆説的に活性化する現象を観測。

Conclusion: 「RAF1 kinase活性がTNF転写を直接かつ単調に抑制する」という単純な因果機構は未確立。AIの単一エッジでは、阻害剤のオフターゲットやフィードバックループを表現しきれない。

The Denied Hypothesis: SOAT1-STAT3 が示すAIの限界

Confidence
0.84183 $\neq$ 84% True

The Finding

予測: SOAT1 affects STAT3 (スコア: 0.84183 - 成功候補と同水準)。

結果: LDLによるSOAT1活性化は確認されたが、STAT3 mRNA / pSTAT3タンパク質ともに有意な変化なし。明確な「不支持 (Denied)」。

The Lesson

AIのスコアは「モデル内部の予測確信度」であり、「特定のWet-labプロトコルでの生物学的な成功確率」ではない。スコア0.8超の候補であっても実験的に完全に否定され得る。検証可能な (Experimentally falsifiable) 仮説を安価に出力することに真の価値がある。

Methodological Blind Spots : 技術と評価の限界点



文献抽出バイアス (Publication Bias)

現状のRule-based抽出は否定文 (Negation) や不確実性 (Speculation) を除外。過去の「出版されたポジティブな結果」のみを継承しやすい。



時系列リークのリスク (Temporal Leakage)

LLMの事前学習コーパスに未来の知識が含まれている可能性。過去時点でのバックテストにおいて、純粹な予測か「記憶」かの完全な切り分けは困難。



説明可能性の錯覚 (Illusion of Causality)

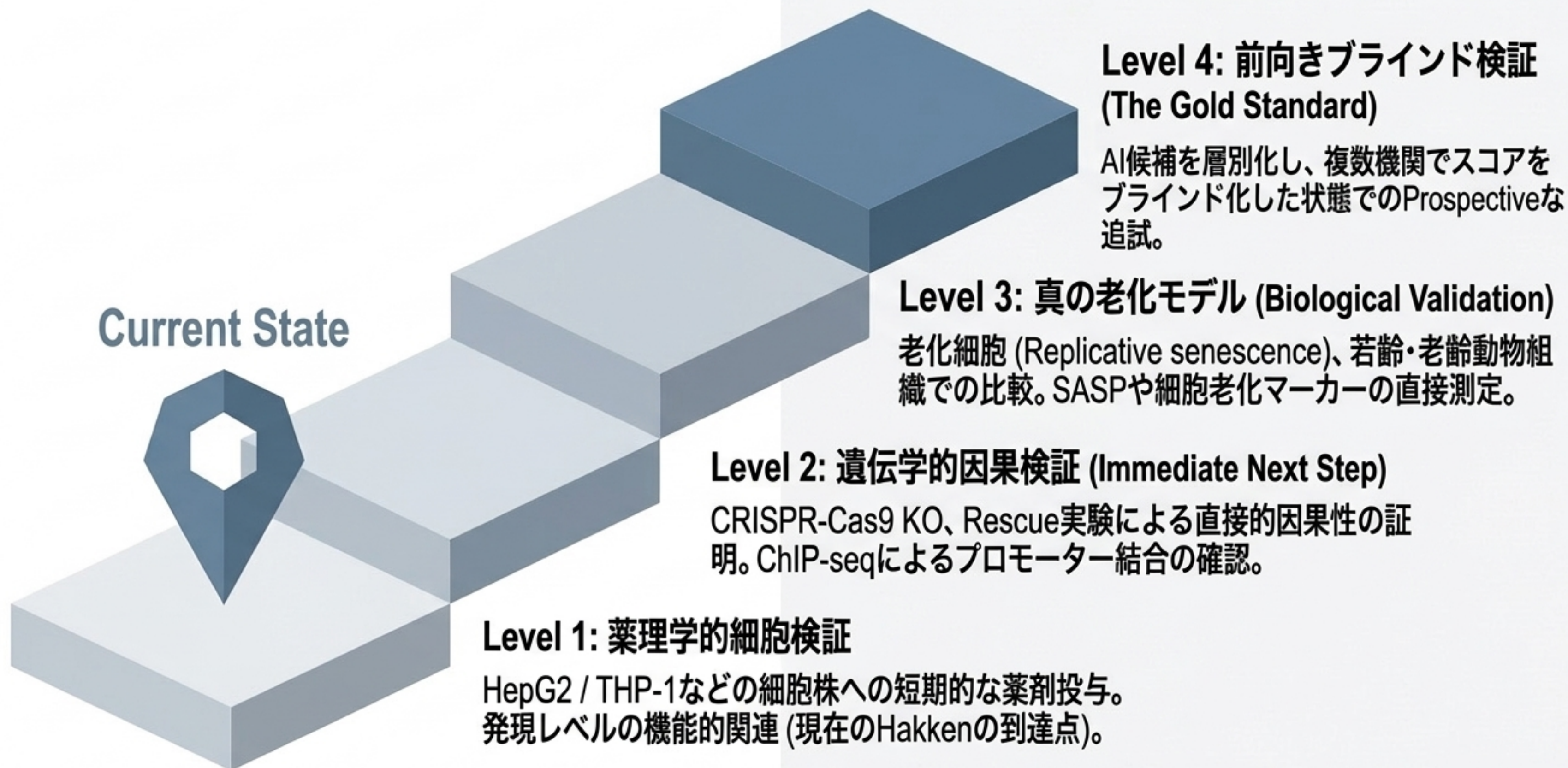
PHELIInEが示す経路は「AIがなぜそのスコアを出したか」の既存グラフ上の近似説明であり、自然界の因果メカニズムの証明ではない。



薬理的検証の限界 (Pharmacological Limits)

小分子阻害剤特有のオフターゲット効果やフィードバックループ。ノックアウト等の遺伝学的手法 (Genetic perturbation) がないため直接性が担保されない。

The Roadmap to True "Discovery" : 老化生物学への昇華に向けて



The Final Verdict：総合評価と次なるマイルストーン

確立された成果 (The Achievements)

- ・時間を考慮した時系列KG、PU Learning、LLMを統合した高度なパイプラインの構築。
- ・文献検索から「検証可能な新規候補の生成」へのPoCを細胞株レベルで実証。HakkenOSS公開による貢献。

退けるべき解釈 (The Misconceptions)

- ✗ AIが老化の分子機構を完全に解明した。
- ✗ AIの発見成功率は66.7%である（強烈な人間による選択バイアスの存在により統計的に無効）。

最優先の推奨事項 (The Mandate)

TP53-BAMBIの直接転写制御をCRISPRで確定させること。そして、154万件の全予測データを固定し、今後数年の独立研究でいくつが「自然に確認されるか」を追跡する「Prospective Benchmark」の実施が、Hakkenの真価を証明する試金石となる。