

NVIDIA JEPA-DNA

生成型事前学習と JEPA を組み合わせた新しいゲノム基盤モデル

調査レポート | 2026 年 7 月 24 日

Claude Opus 4.8

Opus 4.8 に切り替えました

Fable 5 のセーフガードがこのメッセージを検知しました。セーフガードは現在意図的に広範囲に設定されており、安全で日常的なコーディング、サイバーセキュリティ、または生物学に関する作業にもフラグが立つ場合があります。これらの対策により、Mythos レベルの機能をより早くお届けできるようになっており、現在も改善を進めています。

概要

JEPA-DNA は、NVIDIA が開発し Hugging Face 上で公開したゲノム基盤モデル（genomic foundation model）である。従来のゲノムモデルが主に用いてきたトークン（塩基）予測型の学習に加え、Joint-Embedding Predictive Architecture（JEPA、ジョイント埋め込み予測アーキテクチャ）による潜在空間での予測学習を組み合わせた点が特徴で、DNA 配列の機能的な理解を深めることを目的としている。公開されたモデルは DNABERT-2 を基盤とする「JEPA-DNA-DNABERT2」で、2026 年 7 月 23 日にリリースされた。

リリースの背景

これまでのゲノム基盤モデルは、自然言語処理（NLP）モデルと同様に、マスク言語モデリング（欠損した塩基／トークンを復元する学習）に依存してきた。しかし DNA には、単なる配列（トークン）の予測だけでは捉えきれないパターンや関係性が含まれているという課題があった。NVIDIA はこの限界を超えるアプローチとして、機能的な表現を潜在空間で学習させる JEPA をゲノム領域に応用した。

アーキテクチャと手法

JEPA-DNA は、ひとつのモデルに 2 種類の学習目標を組み合わせる（デュアルオブジェクティブ）構成をとる。

- トークンレベル予測：従来型の生成的事前学習。配列中の塩基（トークン）を予測することで局所的なパターンを学習する。
- JEPA（潜在空間予測）：欠損トークンを直接再構成させるのではなく、配列全体のグローバルな埋め込み表現を潜在空間で監督（予測）する。これにより機能的・意味的な表現の獲得を促す。

公開モデルの基盤：DNABERT-2（パラメータ数 約 1 億 1,700 万）を土台とし、上記 2 目標を組み合わせて学習させた「JEPA-DNA-DNABERT2」として提供される。

目的と想定される用途

JEPA-DNA は研究ワークフローでの利用を想定して公開されており、主に次のような用途をサポートする。

- 特徴抽出（feature extraction）
- 線形プローブ（linear probing）による表現の評価
- 継続的事前学習（continual pretraining）実験
- DNA 配列の変化に対するゼロショット評価

なお、公開は非商用の研究目的に限定されている。

関連論文

本モデルは、論文「JEPA-DNA: Grounding Genomic Foundation Models through Joint-Embedding Predictive Architectures」に基づく。JEPA（ジョイント埋め込み予測アーキテクチャ）をゲノム基盤モデルに接地（grounding）させる手法を提案している。

留意事項

本レポートは、公開時点で参照可能なニュース記事およびモデル情報に基づいて作成した。具体的なベンチマーク結果や定量的な性能数値については、一次資料（論文本文）の一部が取得できず本レポートに反映できていない。厳密な性能評価が必要な場合は、Hugging Face 上のモデルカードおよび原論文を直接参照することを推奨する。

主な参照元

- The New Stack — “Nvidia’s new DNA model learns what token prediction misses”
(thenewstack.io/nvidia-jepa-dna-genomics/)
- Hugging Face — NVIDIA 組織ページ / JEPA-DNA モデル (huggingface.co/nvidia)
- 論文: JEPA-DNA: Grounding Genomic Foundation Models through Joint-Embedding Predictive Architectures