Rethinking Genomic Modeling Through Optical Character Recognition

Hongxin Xiang[1,2], Pengsen Ma[1,2], Yunkang Cao[2,3], Di Yu[1,2], Haowen Chen[1,2], Xinyu Yang[1,2], Xiangxiang Zeng[1,2]

[1]College of Computer Science and Electronic Engineering, Hunan University [2]Yuelushan Laboratory
[3]School of Artificial Intelligence and Robotics, Hunan University

ICML International Conference On Machine Learning

Background & Motivation

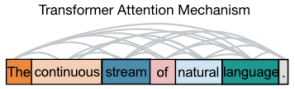
Fundamental Mismatch in LLM-style Genomic Modeling

- **Genomes are sparse, not dense:** Functional signals are discontinuous and embedded in vast low-information background.
- **Sequential reading is wasteful:** Genomic LLMs exhaustively process base/k-mer tokens, spending computation uniformly across background and functional regions.

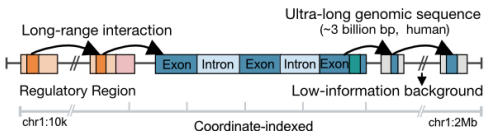
Scalable modeling requires understanding-driven compression:

- **Compression as first-class operation:** Low-information-density genomes make compression essential for scalable genome-scale modeling.
- **Understanding-driven compression:** High-fidelity compression requires semantic understanding — preserving sparse, task-relevant structures while suppressing redundant background.

Natural Language: dense, continuous & compositional

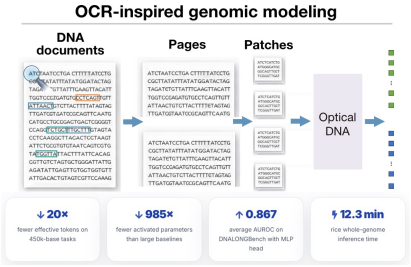


Genomic Reality: sparse, discontinuous & jump-like semantics



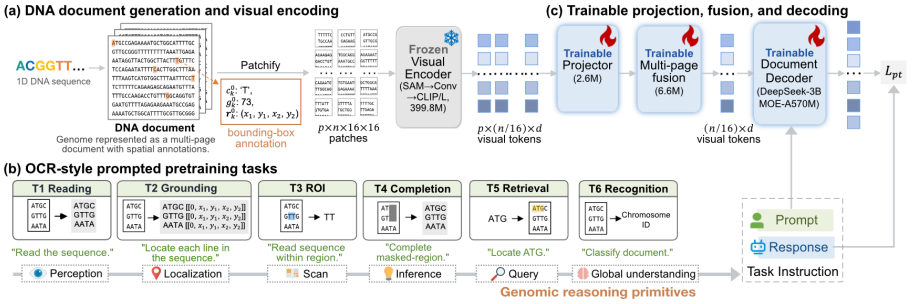
Overview of OpticalDNA

- We introduce **OpticalDNA**, the first **vision-based DNA foundation model** that learns genomic representations through structured visual layouts and region-centric reasoning.
- We design **six OCR-inspired pretraining tasks aligned with practical genomic analysis workflows,** covering core genomic operations.
- Extensive long-range experiments show that OpticalDNA consistently outperforms state-of-the-art baselines, including models up to **985x larger**, while using nearly **20x fewer effective tokens**.



Our Framework

- **Visual DNA representation:** Render DNA into structured multi-page images with nucleotide-level bounding boxes, mapping 1D genome coordinates to 2D visual regions.
- **Understanding-driven compression:** A visual encoder produces compact, reconstructible tokens that preserve fine-grained sequence information while reducing the effective token budget for long inputs.
- **Region-aware reasoning:** OCR-style supervision supports explicit localization, subsequence retrieval, ROI transcription, and masked-region completion through prompt-conditioned genomic tasks.



Long-range Experiments

(1) DNALONGBench summary: accuracy with lightweight adaptation

OpticalDNA achieves the best average AUROC on eQTL tasks while using a lightweight probing setup under **0.45M-base inputs**.

Models #Trainable param	Expert Model (252M)	HyenaDNA (1.6M)	Caduceus-Ph (7.7M)	NT-v2-500M* (1.03K)	GENERator-1.2B* (10.24K)	JanusDNA w/o mid-Attn (7.662M)	JanusDNA mlp w/o mid-Attn (7.745M)	OpticalDNA Linear Probing (256K)	OpticalDNA MLP (1.3M–2.3M)
AT	0.741	0.479	0.690	0.764	0.795	0.803	**0.852**	**0.852**	**0.852**
AS	0.736	0.513	0.759	0.730	0.730	0.741	0.769	**0.813**	0.788
CF	0.639	0.584	0.690	0.744	0.703	0.771	0.802	0.812	**0.819**
MS	0.621	0.487	0.789	0.828	0.819	0.803	0.864	**0.880**	0.877
NT	0.683	0.511	0.842	0.824	0.828	0.877	**0.914**	0.884	0.900
SNSES	0.710	0.471	0.812	0.833	0.854	0.875	0.903	0.904	**0.931**
SSELL	0.700	0.544	0.692	0.749	0.726	0.706	**0.846**	0.835	0.832
Thyroid	0.612	0.529	0.703	0.706	0.749	0.752	0.793	0.791	**0.876**
WB	0.689	0.512	0.769	0.773	0.834	0.794	0.821	0.895	**0.927**
Average	0.681	0.514	0.750	0.772	0.782	0.791	0.840	0.852	**0.867**

256K ↓ 30x fewer
Parameter-efficient adaptation
Linear probing reaches 0.852 average AUROC using only 256K trainable parameters, outperforming JanusDNA-MLP while requiring far fewer task-specific parameters.

(2) RiceSubBench: in-domain to far-OOD generalization

OpticalDNA generalizes from in-domain japonica to near-, mid-, and far-OOD rice subspecies under **0.45M-base inputs**, with especially clear accuracy gains on rufipogon, barthii, and glaberrima.

Model	In-Domain japonica	Near-OOD aus	Mid-OOD rufipogon	Far-OOD barthii	Far-OOD glaberrima
Evo-2 (7B)	0.486 / 0.700	0.509 / 0.714	0.500 / 0.714	0.532 / 0.725	0.489 / 0.705
LucaOne (1.8B)	0.510 / 0.703	0.551 / 0.723	0.589 / 0.760	0.556 / 0.745	0.526 / **0.736**
OpticalDNA (409M)	**0.590 / 0.739**	**0.556 / 0.725**	**0.639 / 0.762**	**0.608 / 0.747**	**0.599** / 0.731

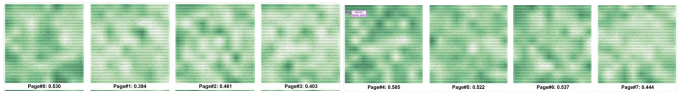
(3) RiceWGPB: whole-genome phenotype prediction

RMSE and inference time under **~400M-base** rice genome inputs.

Model	TGW (g) ↓	LRI-15SZ (%) ↓	Time ↓
Evo-2 (7B)	3.056	9.617	5h40m
LucaOne (1.8B)	8.817	9.740	32.5m
OpticalDNA (409M)	**2.952**	**9.531**	**12.3m**

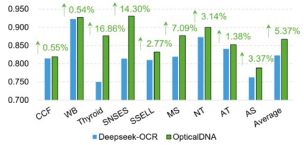
Interpretability

Grad-CAM highlights region-level attributions and biologically meaningful splice signals, showing that OpticalDNA can localize relevant genomic evidence.



Efficiency Analysis & Ablation Study

- **Ablation on DNALONGBENCH eQTLs task**



OpticalDNA improves DeepSeek-OCR on all tissues, yielding a +5.37% relative gain on the overall average AUROC.

- **Visual compression remains stable across rendering resolutions**

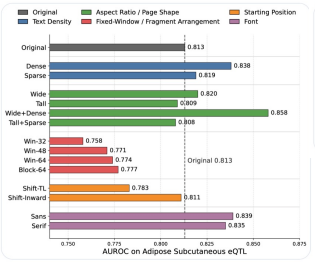
Average AUROC on nine eQTL tasks changes minimally while the visual-token compression ratio increases.

Resolution	#Page	#Token	Compression ratio	AUROC
512	370	64	19.0	0.849
640	226	100	19.9	**0.852**
1024	84	256	20.9	0.851
1280	53	400	21.2	0.849

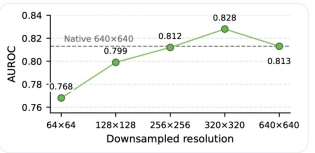
- **Rendering Robustness:** OpticalDNA is not tied to a single visual layout

OpticalDNA remains robust under realistic rendering perturbations, including text density, aspect ratio, starting position, and font changes, suggesting that the OCR-style formulation captures genomic content rather than overfitting to a brittle page style.

Rendering perturbations



Low-resolution robustness

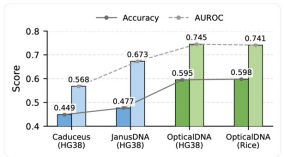


0.858 ↑ robust
Layout robustness
Rendering perturbations do not break the visual formulation: wide+dense rendering reaches 0.858 AUROC on AS eQTL, above the original 0.813 layout.

✓ 0.812–0.828 stable under moderate downsampling
✓ 0.839 / 0.835 robust under sans / serif font changes
↓ Fixed windows main degradation source, due to artificial fragment boundaries

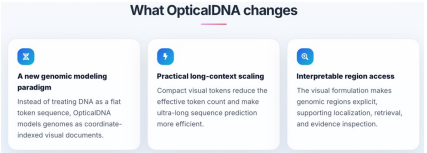
- **Cross-Species Transfer:** visual genomic representations generalize beyond pretraining species

HG38 → rice transfer



0.595/0.745 ↑ transfer
Cross-species transfer
The HG38-pretrained OpticalDNA transfers strongly to rice, approaching the rice-pretrained model and clearly surpassing HG38-pretrained sequence baselines.

Takeaways



Scan for paper, code & updates
We welcome questions, feedback, and collaborations.