AI for Life Sciences Lab

WaveSem

Frequency-Adaptive Tokenization for Disentangling Semantics and Noise in Genomics

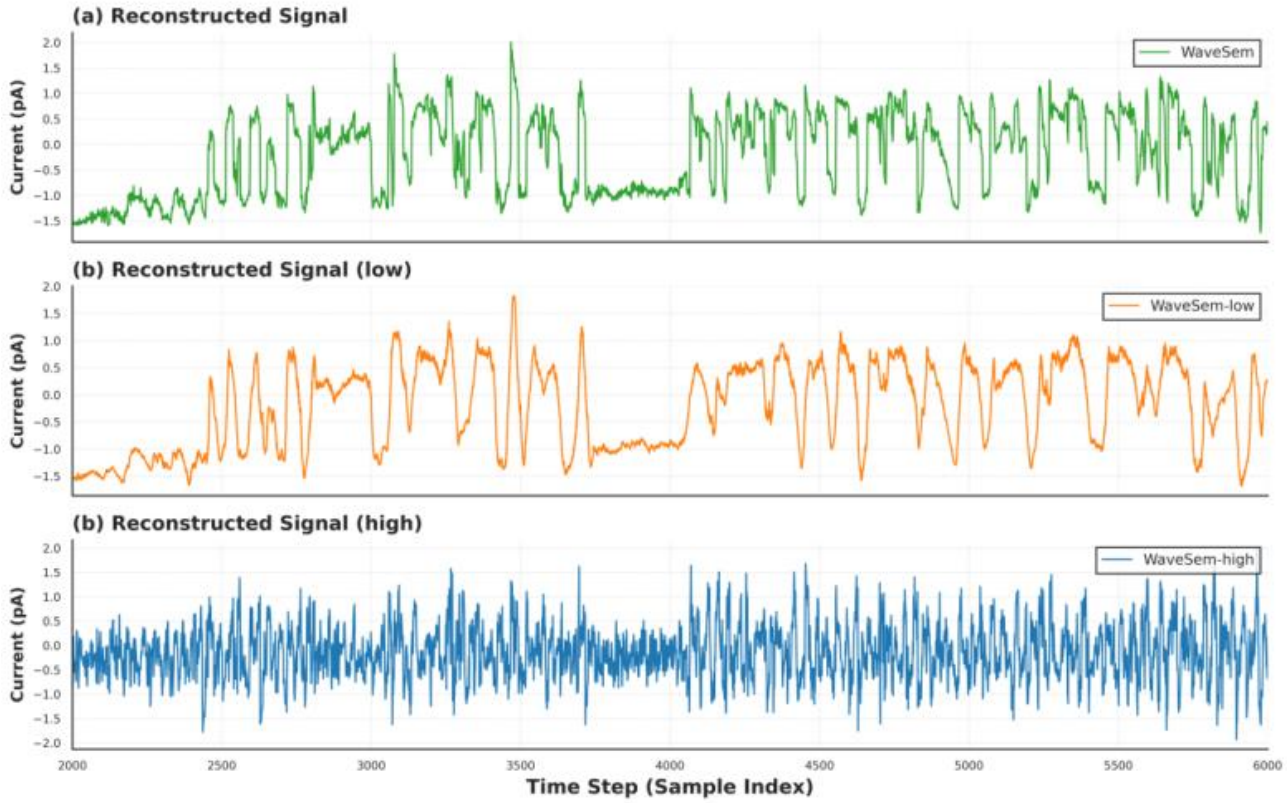
Compact tokens for useful genomic information

Shouzhi Chen[12]*, Bing He[1]*†, Zhenchao Tang[1], Jun Zhu[1], Minghao Yang[1], Tianxu Lv[1], Yu Wang[1], Jiayang Wu[1], Fang Wang[1], Yu Zhao[1], Chenchen Qin[1]†, Jianhua Yao[1]†

[1] Tencent AI for Life Sciences Lab [2] Sun Yat-sen University

* Equal contribution † Corresponding authors

Useful structure is buried in noise



Original Figure 6 (left): reconstruction, semantic component and residual.

The representation bottleneck

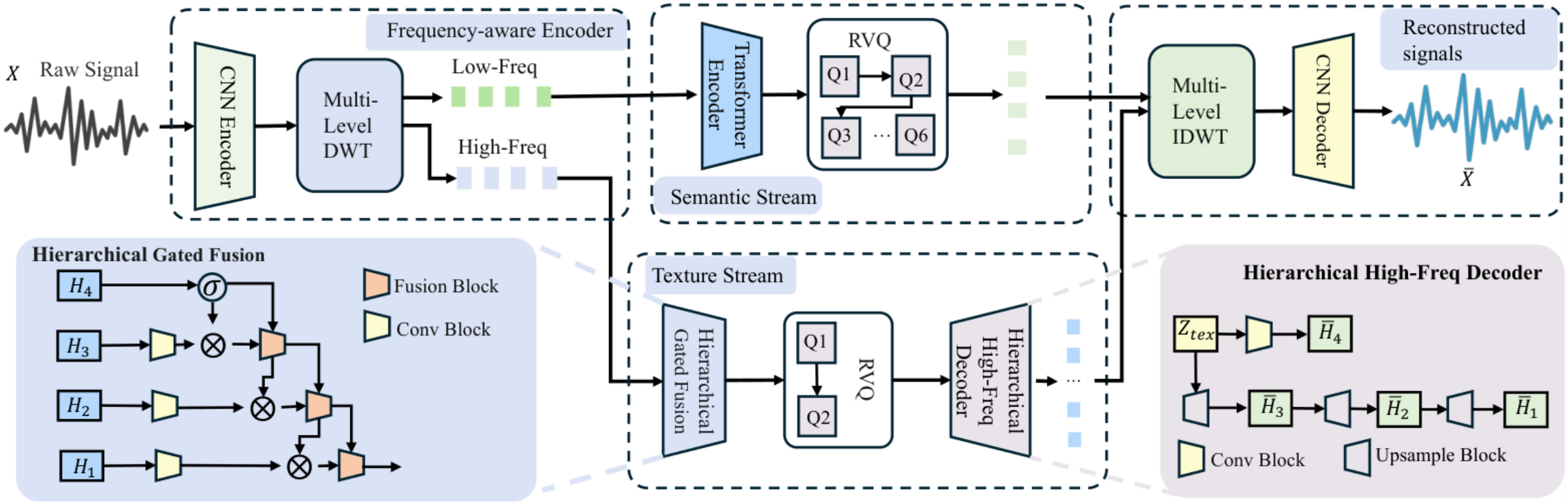
Raw waveforms are expensive.
Basecalls discard physical cues.

Our inductive bias

Give low-frequency structure and high-frequency residuals different information budgets.

Frequency-structured partial separation; not statistical independence.

Split frequencies before quantization



Semantic stream

Transformer + 6-stage RVQ
More capacity for the low-frequency scaffold.

Texture stream

Gated fusion + 2-stage RVQ
Structured residuals complete reconstruction.

Original Figure 1 · 4× CNN encoder + four-level db4 DWT · IDWT + CNN reconstruction

Compact tokens support two downstream tasks

Basecalling

Accuracy · 64× temporal downsampling

Representation	Input Type	Downsampling	ACC (%) ↑
Bonito	Continuous Waveform	1×	98.12 ± 0.05
SoundStream	Discrete Tokens	64×	92.64 ± 0.41
Standard VQ	Discrete Tokens	32×	73.92 ± 0.54
	Discrete Tokens	64×	65.13 ± 0.91
WAVESEM	Discrete Tokens	32×	**96.88** ± 0.09
	Discrete Tokens	64×	**95.61** ± 0.09

+2.97 pp over retrained SoundStream

Token-adapted decoders · original Table 1

5mC methylation

Segment-level 5mC · AUROC

Frozen Encoder	ACC ↑	F1 ↑	AUROC ↑	AUPRC ↑
Random Frozen CNN (Control)	0.811	0.812	0.889	0.897
SoundStream	0.733	0.743	0.808	0.813
RVQ-VAE	0.812	0.814	0.886	0.901
Single-Codebook VQ-VAE	0.836	0.835	0.914	0.924
WAVESEM (No DWT)	0.779	0.790	0.859	0.870
WAVESEM (Texture Only)	0.791	0.795	0.864	0.877
WAVESEM	**0.852**	**0.851**	**0.928**	**0.936**

+0.039 over random frozen CNN

Frozen-encoder probe · original Table 2

5mC: balanced 20K-segment test from distinct runs; relative information retention.
Basecalling: task-adapted utility, not drop-in compatibility with raw-signal callers.

Each component serves a distinct role

Model Variant	Spectral Bias (DWT)	Dual Stream	Contextual Gating	Bit Split	MSE ↓	STFT ↓	ACC (%) ↑
Single-Stream	×	×	×	×	0.5125	1.4265	65.13
Dual (No DWT)	×	✓	✓	✓	0.2406	1.1602	95.20
Dual (No Gate)	✓	✓	×	✓	0.1250	1.0161	94.80
Dual (Swap Bits)	✓	✓	✓	×	0.1845	1.1032	92.10
WAVESEM (All)	✓	✓	✓	✓	**0.0757**	**0.8697**	**95.80**
WAVESEM (Texture Only)	✓	✓	✓	✓	0.6871	5.8670	81.22
WAVESEM (Semantic Only)	✓	✓	✓	✓	0.1523	1.0847	95.61

DWT improves reconstruction

MSE: 0.2406 without DWT → 0.0757 full
Spectral bias protects waveform fidelity.

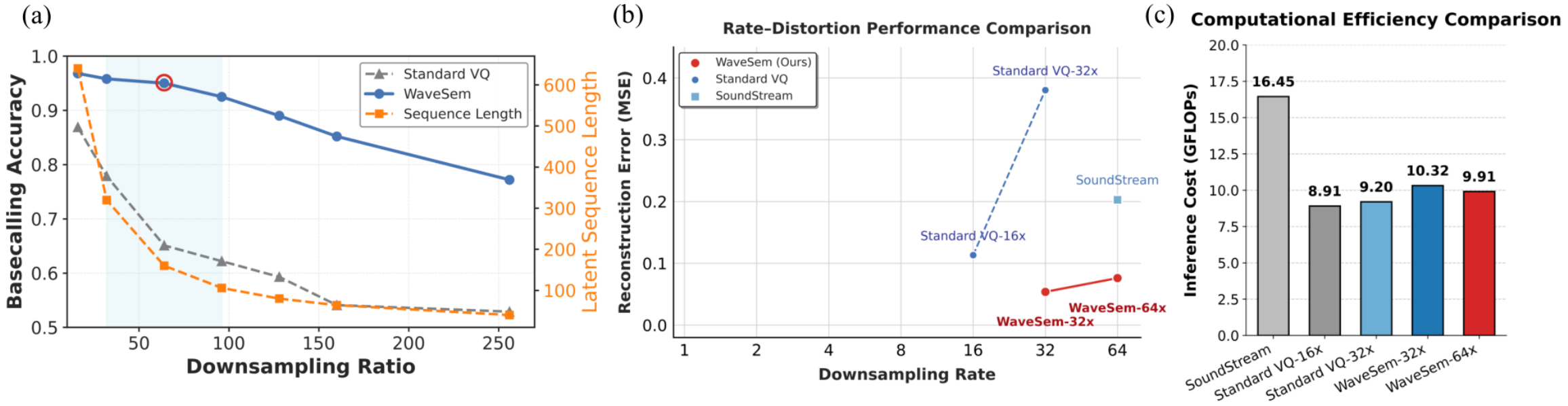
The streams are complementary

Semantic only: 95.61% ACC, 2.0× full MSE
Texture adds detail to the semantic scaffold.

Original Table 3 · Swapping bit allocation reduces ACC from 95.80% to 92.10% in this ablation.
Compare within Table 3: its full-model 95.80% is a separate result from the Table 1 main experiment.

Compression, fidelity and computational cost

Original Figure 4 · Basecalling, rate–distortion and inference cost



Utility under compression

Accuracy remains high at 64×.

Low reconstruction error

MSE 0.0757 at 64×.

Moderate inference cost

9.91 vs. 16.45 GFLOPs.

64× is temporal downsampling; nominal storage reduction is 14.2× versus Int16 (1.125 bits/sample, excluding overheads).
Panel (c): WaveSem-64× versus SoundStream; GFLOPs measure computational cost, not runtime.

Preserve useful structure.
Spend fewer bits.

A spectral inductive bias for compact genomic tokens.

Frequency-aware allocation · Asymmetric RVQ · Downstream utility

For lossy token archival and task-adapted analysis.

Code: github.com/GEHAH/WaveSem