Adversarial Attack Mechanisms in Genomic Models

PGD Attack Mechanism

Gradient-based Iterative Process

Legend:
- Loss Gradient
- Model Prediction
- Perturbation

FIMBA Attack Mechanism

Feature Importance Ranking

Target Features

- Top-5 Threshold
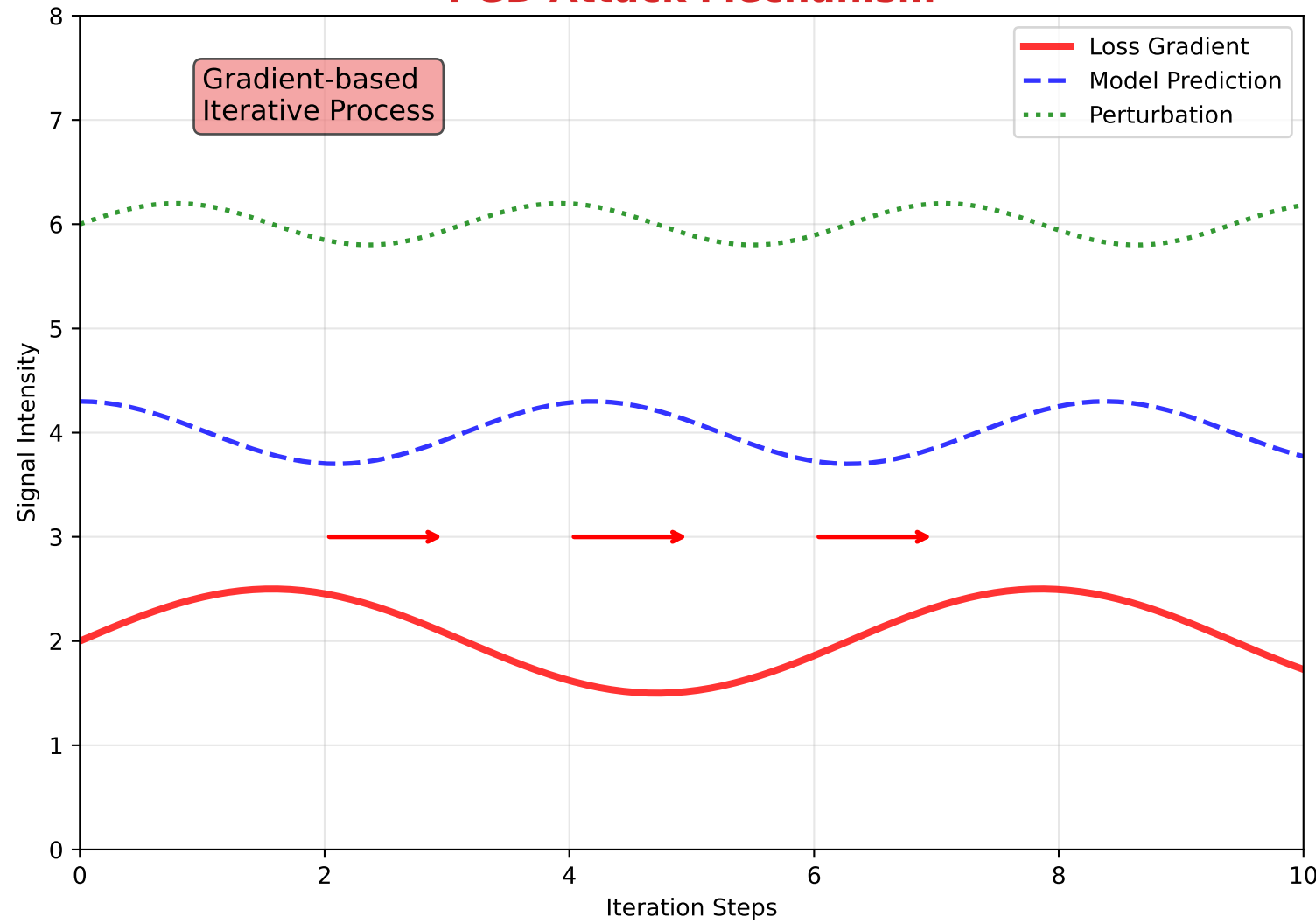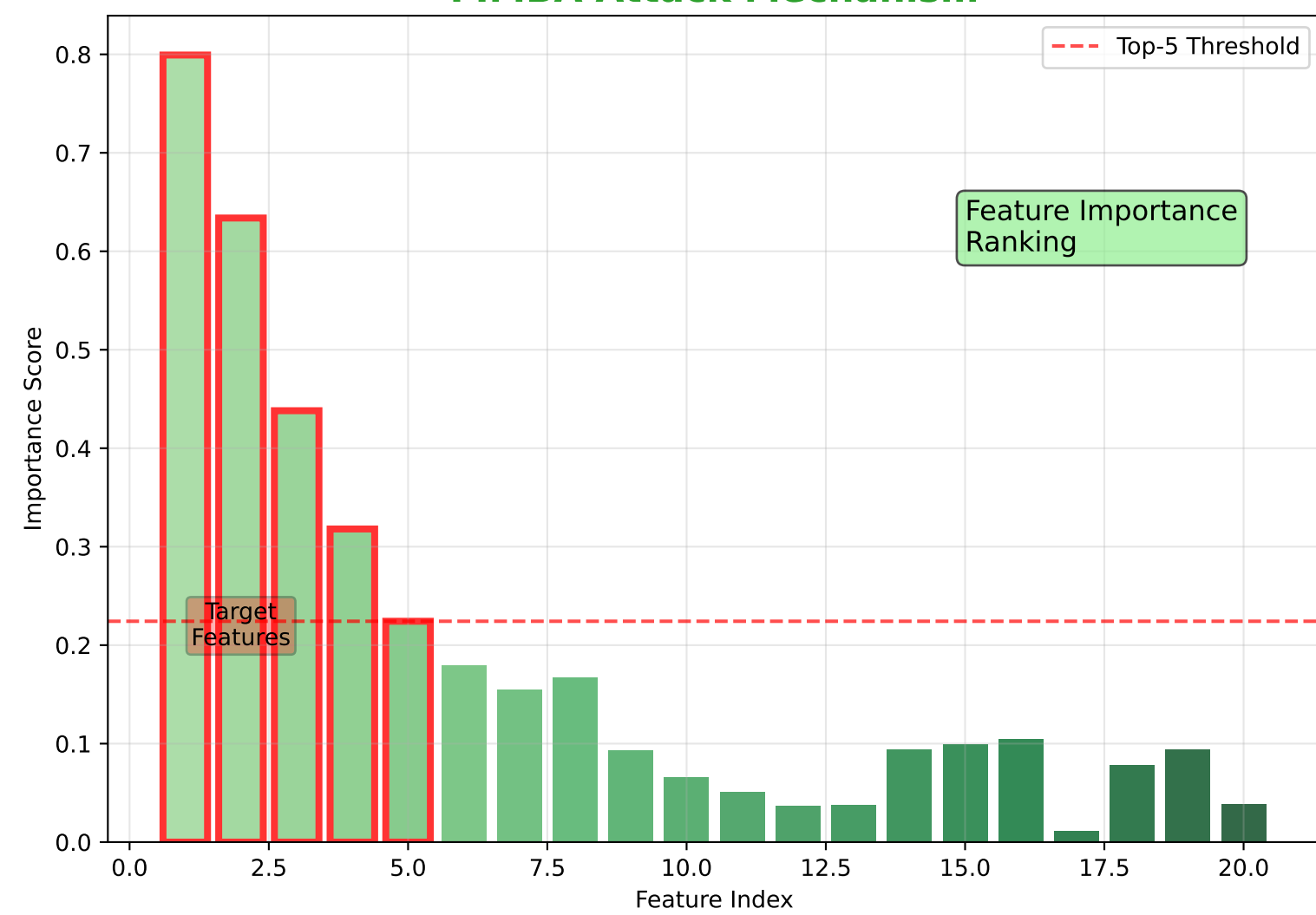
DNA Sequence Perturbation Example

Original: A T C G A T C G A T C G A T C G

Perturbed: A T C G A T C A A T C G A T C G

Mutation

Attack Effectiveness Comparison

- Low Impact
- Medium Impact
- High Impact

PGD L2: 3.8%
PGD L∞: 58.9%
FIMBA: 27.1%