90-Minute Discussion Session Plan
Lecture 6: Genome Annotation

Course: BINF301 – Computational Biology
Instructor: Tom Michoel
Date: 2/2/2026
Created with Copilot

0–10 min — Warm-Up

Prompt:

> Which part of genome annotation seems most challenging: repeat masking, gene prediction, or functional annotation?

10–25 min — Think–Pair–Share

Main Prompt:

> Why is repeat masking the first step in most genome annotation pipelines?

Follow-up Questions:

- "What specific problems do unmasked repeats cause for ab initio gene predictors?"
- "How does softmasking vs. hardmasking influence downstream tools?"
- "Why can repeat detection be slow, and why might RED be used instead of RepeatMasker?"

25–45 min — Structured Group Discussion

Students form groups of three with rotating roles.

Starter Question

> Why is eukaryotic gene prediction dramatically harder than prokaryotic annotation?

Roles

- **Summarizer:** Explain differences between prokaryotic and eukaryotic gene structure (Slides 10–14 vs. 20–24).
- **Questioner:** Raise a question about ab initio gene prediction (HMM states, GC content, training).
- **Connector:** Link RNA-seq/protein homology evidence to gene prediction quality (Slides 25–27).

Break (15 min)

45–60 min — Deep Dive: Integrating Evidence Sources

Discussion Prompts:

- "When is ab initio prediction alone insufficient?"
- "How do BRAKER2 and BRAKER3 integrate intrinsic HMMs with extrinsic RNA-seq and protein hints?"
- "What are limitations of homology-based predictions in non-model organisms?"
- "How would you resolve conflicts between RNA-seq evidence and ab initio predictions?"

60–70 min — Deep Dive: Annotation Quality Assessment

Prompts:

- "What exactly does BUSCO completeness mean?"
- "How does OMArk detect contamination and taxonomic inconsistency?"
- "Why might duplicated BUSCOs indicate fragmentation rather than real gene duplication?"

70–90 min — Assignment

Students work on the Portfolio Assignment Genomics.