Metrics on RNA secondary structure ensembles

Bioinformatics Module Project

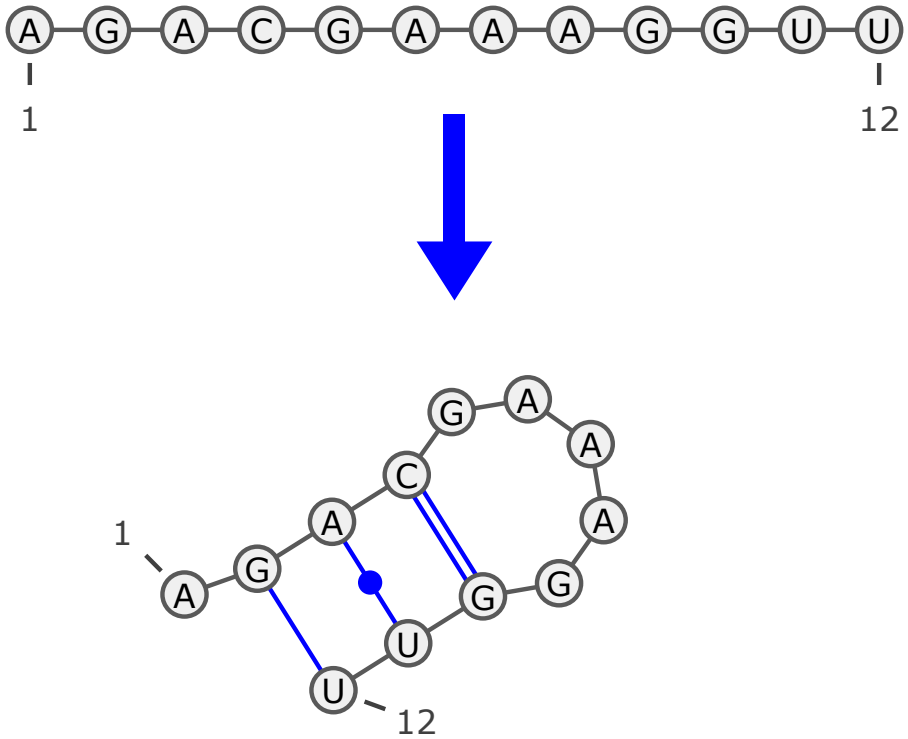
Michael Golden[1], Alain Laederach[2], Jotun Hein[1]

Students

- David Kell
- Hannah Patel
- Tammo Rukat

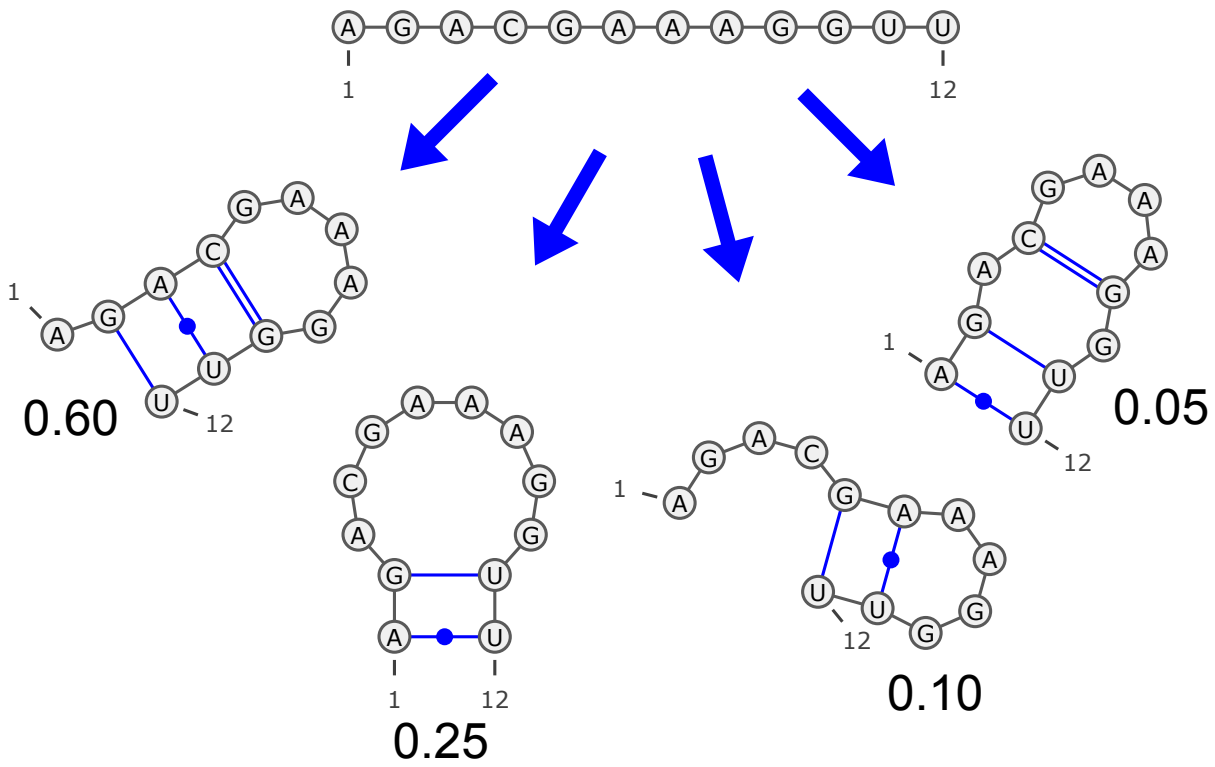
Introduction

- RNA secondary structure is defined as the set of base-pairing interactions between the constituent bases of a RNA sequence.



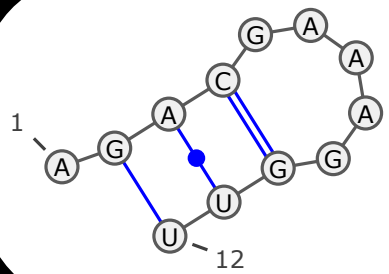
Introduction

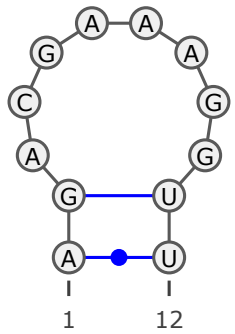
- It is more accurate to think of a RNA sequence as able to adopt a range of possible conformations with varying degrees of probability.



Introduction

- Various metrics exist for measuring distance between two RNA secondary structures.

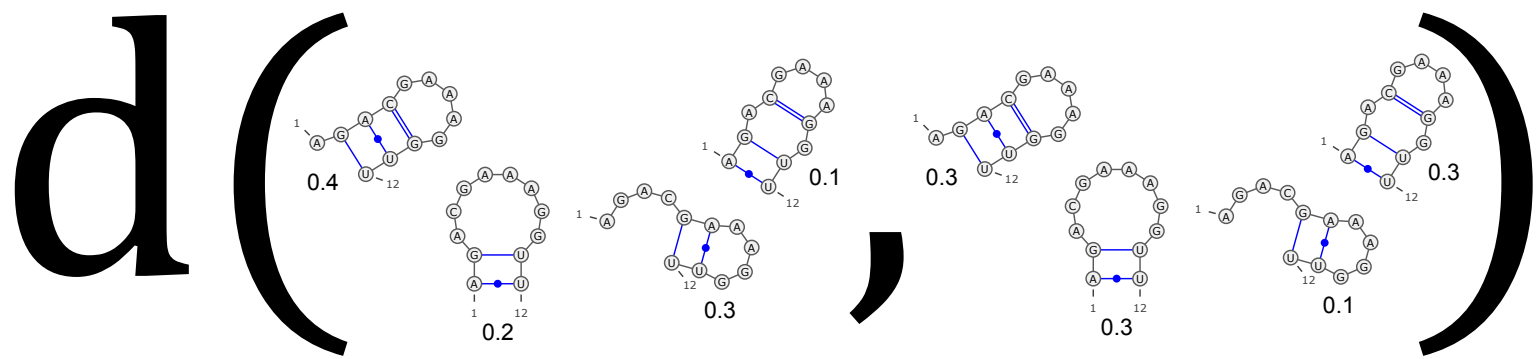
$$d\left(\right.$$ 

$$,$$ 

$$\left.\right)$$

Objective

- But no metric exists for measuring distance between two RNA secondary structure ensembles.



Objective

- Design two metrics:
  1. Within-ensemble variance
  2. Ensemble distance metric

Applications

- Test whether associations exist between the measured properties of RNA ensembles and different functional annotations.
  - For example: are flexible sequences (higher ensemble variance) associated with certain functional annotations?
  - Can distances between wild-type and mutant ensembles be used to identify disease-associated mutations?