Fall 2015

Discussion Questions: Genome Mining

Sarah O'Leary-Driscoll
Illinois Mathematics and Science Academy, soleary@imsa.edu

Discussion Questions: Genome Mining

Think about how the elements listed help to determine whether or not a sequence is random.

1. Proportions of elements
2. Patterns
3. Rules
4. Predictability
5. Frequency

- How does this relate to letters in the alphabet and writing? What about binary code in computers?

- How do you think that this relates to how we mine for useful information in a DNA sequence? What challenges would there be in applying these sorts of criteria to DNA that were not found in the other examples (written words and computer language)?

Go to: DNA Analysis and open up "Gene Boy"

Click on one of the sequences on the left, and then on the right side, select "Analyze Composition" to see the kind of information that can be collected with DNA sequences.

Move on to "Gene Features"

- What are reading frames? Why does it matters for analysis which reading frame that you choose?
- What are the specific characteristics of an open reading frame?
- How do matrices help determine most likely positioning of the promoter? Why are imperfect matches still accepted if they have high scores? How does this relate to your understanding of the promoter, biologically speaking?
- How is this process used to look for splice sites? What criteria did they use and what conclusions were drawn?
- The poly A tail is post transcriptionally added, but the poly-adenylation sequence is still an important part of gene structure. What was determined about that particular sequence in terms of consensus? Why might it be important to know what that sequence is and where it is if you're examining or manipulating a gene?

Move on to "Gene Finding"

- What are the two different ways in which genes are identified within a DNA sequence? What do you think the pros or cons of each might be?