

## **Clinical and Research Genomics Assignment #4**

### **From Lecture\_10-12 (April 11th): Complex Genome Re-arrangements, Transposons, and Genetic Variant Calling**

---

**Assignment: Answer questions on genetic variation and clinical genomics.**

**Due Date: 5:00PM on April 21<sup>st</sup>**

1. How is the positive strand of a chromosome defined?
2. What are the differences between indels, structural variants (SVs), and copy number variants (CNVs)? Why is this a useful distinction to make for both detection purposes and general biology?
3. Where in terms of chromosome and strand would each of a pair of DNA WGS short reads align if they flanked (but did not span) the breakpoint of a fusion joining gene A on the negative strand of chr 1 to gene B on the positive strand of chr 2?
4. We use long-read nanopore DNA sequencing to confirm the gene fusion in (3). These data comprise thousands of single end ~10 Kbp reads, each a 5' to 3' sequence. We align these reads to the reference genome in a piecewise fashion, so a gene fusion will yield a multi-part alignment. What are the two expected alignment patterns (successive chromosomes and strands) for reads that span the fusion in (3)? Answers should be written in 5' to 3' order along the read.
5. Describe a strategy to improve identification of species from short reads in metagenomic experiments (from class or another source).
6. What assumption underlies the Perfect Phylogeny Model? Briefly comment on the relevance of the model to cancer research.

---

**Please hand the assignment on the day of the lecture, or email if you cannot attend.**

**For any questions, please contact Chandrima Bhattacharya ([chb4004@med.cornell.edu](mailto:chb4004@med.cornell.edu)), Ebrahim Afshinnkoo ([eba2001@med.cornell.edu](mailto:eba2001@med.cornell.edu)), or Professor Mason ([chm2042@med.cornell.edu](mailto:chm2042@med.cornell.edu))**