

Homework 4
BENG 183

1) How many clusters per lane do you need on an Illumina flow cell to qualify for the X Prize? This is a high-level question for which you will need to research a number of facts and make a number of assumptions. Make sure to state all of them. You will find the following reference useful: Lander ES, Waterman MS. "Genomic mapping by fingerprinting random clones: a mathematical analysis." Genomics. (1988)

2) Pyrosequencing. One difficult problem in sequencing is handling heterozygous loci. Typically, PCR amplification of an exon from a heterozygous individual will produce both alleles in roughly equal amounts. You use pyrosequencing on a short piece of DNA that has a known SNP and get the following result:

A	T	G	C	A	T	G	C	A	T
0	1	3	2	.5	0	.5	1	2	1

a. What are the 2 alleles?

b. For a DNA fragment of length N, what is the smallest peak intensity possible? What is the largest possible?

c. What would a SNP look like in a four-dye chromatogram?

3) Solexa Sequencing. Many people use a Solexa machine in "paired end" mode to capture short reads from both ends of a longer cloned sequence. Using your own reasoning, draw a diagram showing how this works along with accompanying text to explain the procedure. As long as you explain yourself and your approach makes sense, you will not be penalized for not guessing the "True" method used by Illumina.

4) Base Calling. Writing a real base caller is difficult, we will write a toy peak caller on simple data.

Download trace.txt off the assignments page.

a) Visualize the sequence traces. Show a graph of each trace and which base it is identifying. What do you think the sequence should be?

b) Write a base calling algorithm to identify the base at each position. Your output should be the sequence of bases your base caller calls i.e "GATTACA".

You make work with a partner for this problem, if you do you may turn in the same code, but make sure you clearly label in your code who you worked with. The rest of the assignment should still be done independently.