Broken String Biosciences Test Answers:

1. Based on the scatter plot generated in question 2)d) one would conclude that the samples that are most likely to be controls are samples 1, 2, 4, 5, 6, 7 and 8 while the ones that are most likely to be treated are samples 3, 9, 10, 11, 12, 13, 14, 15 and 16
2. Samples 3 is one of the samples which I am unsure whether it was a control or a treated sample.
3. Even though sample 3 did not have a normalized sum AsiSI breaks of 0 (which is why I placed it in the treated group to begin with), I believe that sample 3 might be a control sample. Sample 3 has a normalized sum AsiSI breaks <0.5 while most of the treated samples have a normalized sum AsiSI breaks >=2. I believe that the break site reported in sample 3 that was intersected with the predicted AsiSI site was not actually caused by the AsiSI enzyme itself but by another phenomenon which created a double stranded break at the position that coincided with the predicted AsiSI site. Following that same reasoning it is also possible to have doubts about sample 13 which has a normalized sum AsiSI breaks <1.25 in comparison to the rest of the treated samples that have a normalized sum AsiSI breaks >=2.
4. The maximum percentage of possible AsiSI cut sites on chromosome 21 that is observed in a single sample is 5.63%. Both samples 9 and 15 have 4 different predicted AsiSI break sites intersecting with their break sites, the total number of predicted AsiSI break sites on chromosome 21 is 71 -> (4/71)\*100=5.63%.

Commands used on the terminal to answer the question:

wc -l ./results/final/Sample\*\_AsiSI\_breaks.bed | sed '$d' | sort -r

wc -l ./data/chr21\_AsiSI\_sites.t2t.bed