10X GEM-X

-Vivian interested in percent duplication (sequence same read more than once—there’s a goldilocks range

--computer % of reads/cell

-look at the percent loading

-CellRanger has an output .html with quality metrics

Vivian had some concern about how the sequencing was done, and asked if the genomics core might have underloaded the flowcell.  From what I can tell, Illumina doesn’t think that should cause any problems, just lower quantity of data.  To me, the surprising things are the high numbers of cells and high numbers of UMIs per cell.  I’d pay special attention to the possibility of cell multiplets in the data.

--more than 2 alpha or beta chains

--more than 1 hashtag per cell.

the UMIs per cell are really high and the sequencing saturation is really low.

If you set threshold for upper level of features and umis, you can lose activated cells

Don’t have thrsholds on # features and UMIs

Is relationship between hashtag multiplets and high counts

Could be multiplets or could just be really activated

r$> table(htDF$manualHT)

CerosalettiLab1059994-CD3CD28 CerosalettiLab1059994-CEFX CerosalettiLab1059994-Islet CerosalettiLab1464776-CD3CD28 CerosalettiLab1464776-CEFX CerosalettiLab1464776-Islet CerosalettiLab942655-CD3CD28

470 358 4667 858 558 3257 620

CerosalettiLab942655-CEFX CerosalettiLab942655-Islet Multiplet Negative

517 21240 1722 120