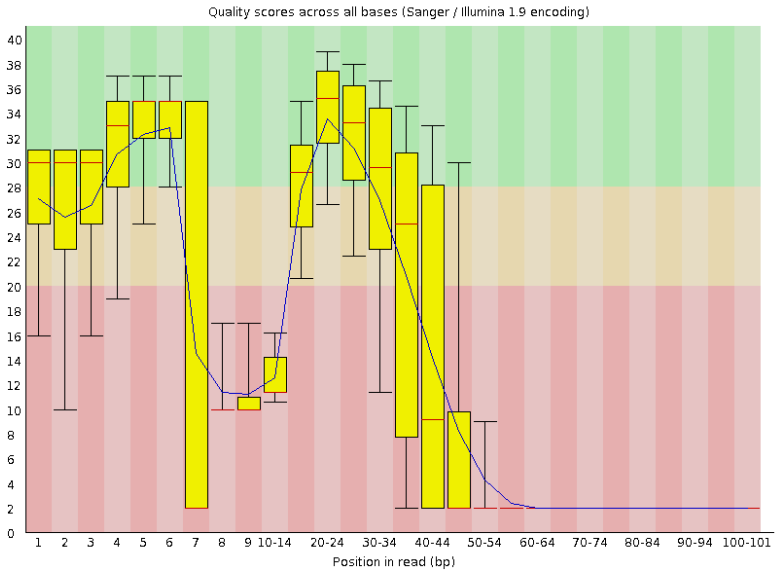
Library GQI19 Sequencing Failure Analysis

Library and sequencing information:

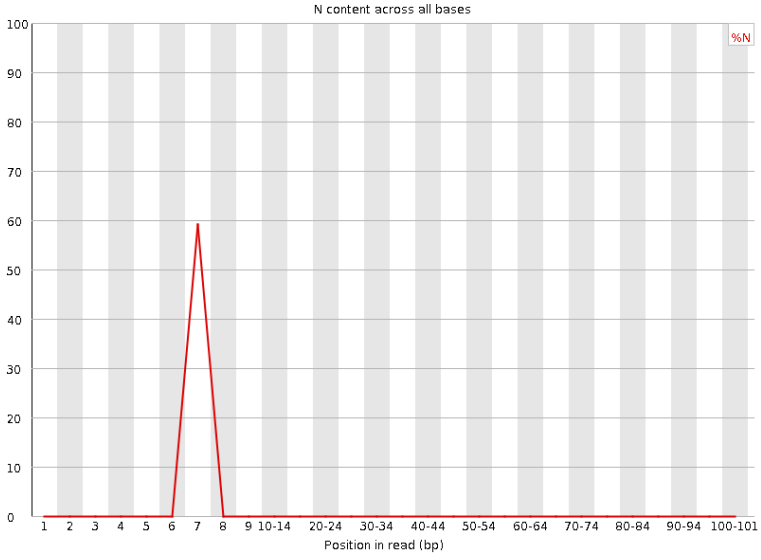
- The library is a RADtag library with 45 different 6 bp long barcodes followed by a restriction site sequence (6 bp, identical for almost all fragments)
- The fragments are 350-600 bp long
- The library was spiked with 20% PhiX to increase the complexity
- It was sequenced single end on an Illumina HiSeq 2500 lane
- The other 7 lanes also contained RADtag libraries with equal characteristics (same barcodes, identical restriction site, fragment length distribution, 20% PhiX…) that all resulted in high numbers (160-200 million) of good quality reads

Results:

- 47 million reads with very low quality
- 60% of the reads have an “N” at the 7th position
- Cluster density: ~750 k/mm2 (comparable to the other libraries)
- Cluster distribution and base intensities look normal

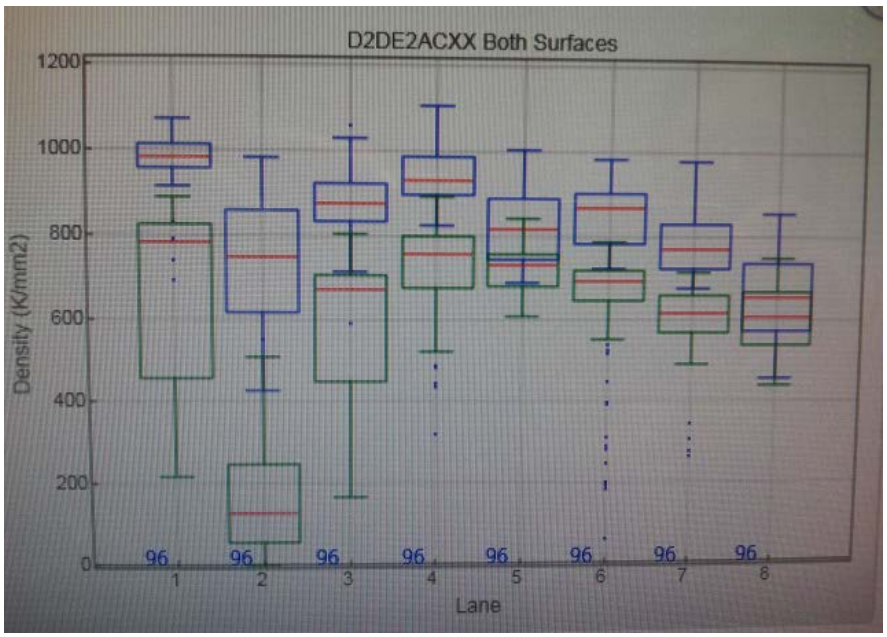


FastQC plot showing Phred Quality scores per base position



FastQC plot showing percentage of uncalled bases along the reads

-> 60% of the reads have an uncalled base (N) at position 7



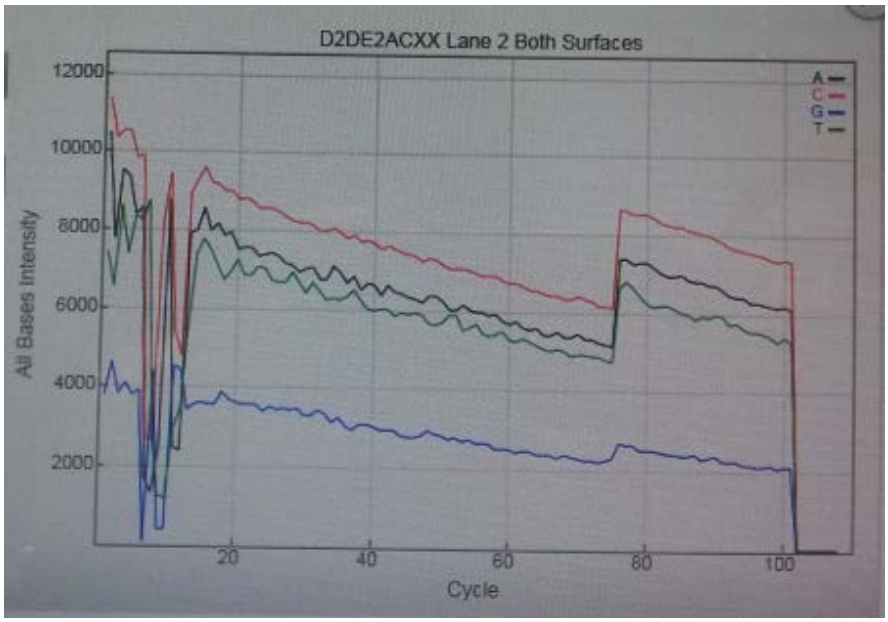
Cluster density of all lanes

Blue: total cluster density

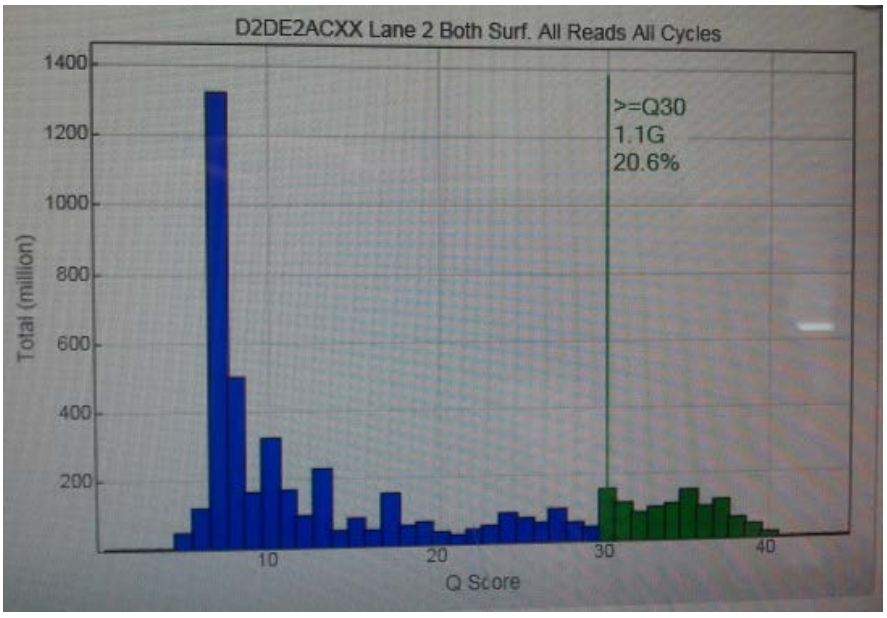
Green: Density of clusters that passed the quality filter

(recommended cluster density: 850k/mm2)

-> Library GQI19, on lane 2, has a good original cluster density but most of them fail the quality filter

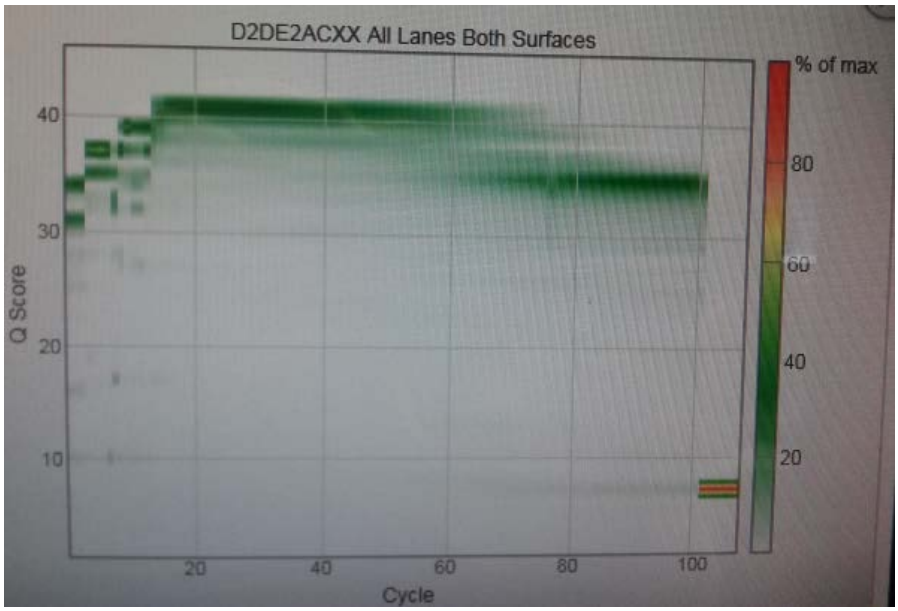


Base intensity per cycle of lane 2 (very similar for the other libraries)

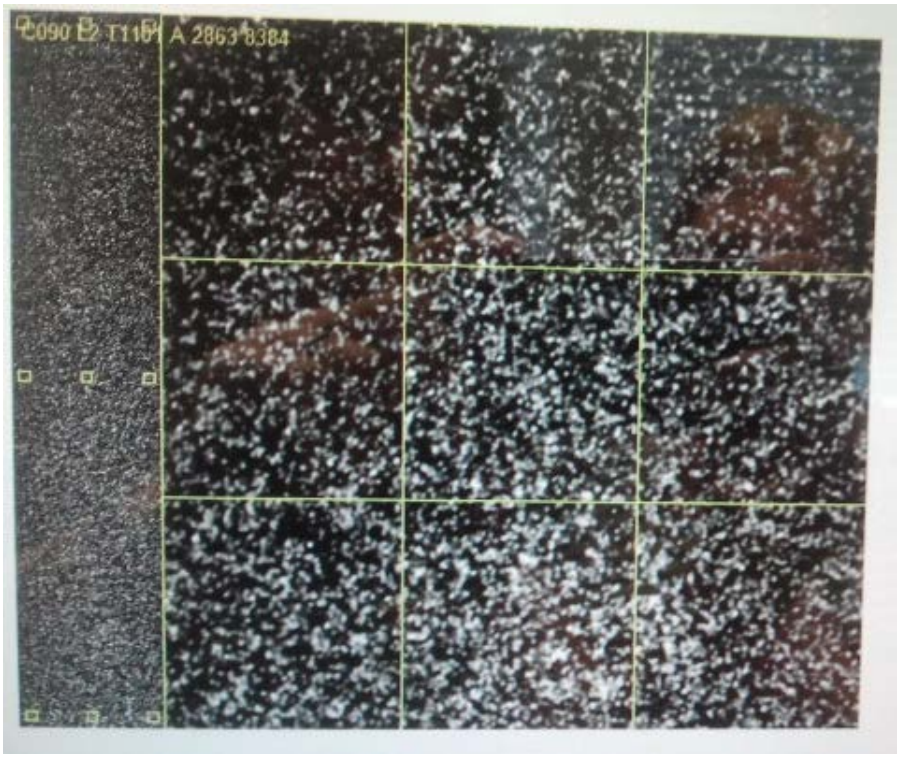


Quality distribution of lane 2

(Very low, should be mostly above Phred quality score 30)



Average quality score across all lanes



Cluster distribution