Multiple Choice Questions Day 1 (IFAT, form #E028)

1. Which of the below sequencing techniques require DNA amplification during the library preparation step (is considered a 2nd generation sequencing technique)?

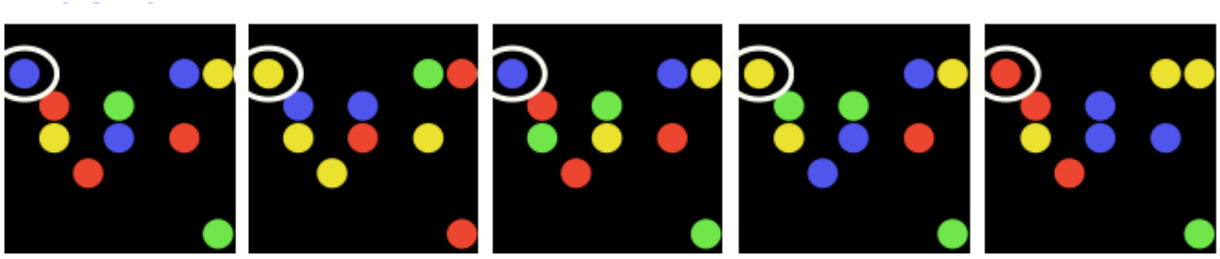
A: PacBio AND Oxford Nanopore
B: Only Illumina
C: Illumina AND Oxford Nanopore
D: Only PacBio
E: PacBio, Illumina, AND Oxford Nanopore

2. Which of the below sequencing techniques use(s) fluorescently labelled nucleotides for identifying the nucleotide sequence of the template DNA strand?

A: Illumina AND PacBio
B: Illumina AND Oxford Nanopore
C: Only PacBio
D: Only Illumina
E: All sequencing methods use fluorescently labelled nucleotides for identifying the nucleotide sequence of the template DNA strand

3. The below figure illustrates five cycles of Illumina sequencing, where each coloured spot represents a flow cell. What is the sequence of the DNA template in the top, left corner according to the figure?

A: Yellow
C: Red
G: Blue
T: Green



A: GAGAC
B: CTCTG
C: ATGCG
D: GGACT
E: GTCAC

4. Which of the below statements are correct when it comes to 3rd generation sequencing techniques?

A: In general, they produce relatively short reads with a low error rate
B: In general, the reads have a high error rate in stretches with the same nucleotide, e.g., AAAAA (homopolymers)

C: Illumina is considered a 3rd generation sequencing methods
D: In general, they produce very long reads with a high, unspecific error rate
E: All 3rd generation sequencing machines are huge

5. Which file format is seen below?

```
@M10_0139:1:2:18915:1321#ATCACG/1
TATCAAGAAAGATTTTAACAGCATTGACTCTGTTATCGAGTTTCATTTTAAACATAGTTTCCAGTGGT
+M10_0139:1:2:18915:1321#ATCACG/1
_bbeeeccgfgecgiiiihfhchiiiiiiiiihhfhhh^dghhhhf_fffghhhhhhacgeeghgbb]
@M10_0139:1:2:18915:1321#ATCACG/2
AGTTCATAGTGACAAGGTAATATTTGTCAAATTATATCGACCTAAAACGGTAGGATATATAACAAAAT
+M10_0139:1:2:18915:1321#ATCACG/2
a__eceeeeggffhihe^bhfiifh_edeg_agbgd]dd`g`fgdhedffaedadhhchhfhiicfhX
@M10_0139:1:2:12256:1321#ATCACG/1
ACGGGTGAACTGTACGGCATCGAAGCCCTTGCGCGCTGGCACGATCCCCAGCATGGTCATGCCCCCTC
+M10_0139:1:2:12256:1321#ATCACG/1
___`c_c`egge[bfghdeghfhhhhhfiii_ffhhN`ghhfddbcddadcddbccb_bbbcbc^aac
@M10_0139:1:2:12256:1321#ATCACG/2
AATCCGGAAAAGCCCGTACCAAAATCATCTACCGATAAGCCCACGCCCATATCACGCAGGATGAATCG
+M10_0139:1:2:12256:1321#ATCACG/2
a_ZcccWHO_bgadgc_WbaceZefda^f`egd`HO[ega\G\b`F_dggeca_cad`Y]^b__bKYZ
```

A: Sanger format
B: FASTA format
C: Illumina format
D: ASCII format
E: FASTQ format

6. In a file containing raw sequence reads, one of the bases is associated with the quality score “N”. What is the probability that this base is incorrect?

A: 0.0000016 %
B: 0.015 %
C: 45 %
D: 78 %
E: 0.00316 %

7. Which file format is used for storing the sequence data of assembled genomes (draft and finished)?

A: Sanger format
B: FASTQ format
C: Illumina format
D: FASTA format
E: ASCII format

8. You have sequenced a genome. Following assembly the draft genome has a size of 5,000,000 bp. The raw output from the assembly is 2,500,000 reads. The average read length is 100 bp. What is the depth of coverage?

A: 200x
B: 10x
C: 125x
D: 150x
E: 50x

9. You have sequenced a genome. Following assembly the draft genome has a size of 5,000,000 bp. The known size of this genome is 5,500,000 bp. What is the breadth of coverage?

A: 0.5
B: 1,1
C: 0.91
D: 0.80
E: 5.5

10. A draft genome consists of seven contigs. Their lengths are listed below. What is the N50 value of the draft genome?

```
Contig1:    2500
Contig2: 150.000
Contig3: 100.000
Contig4:  90.000
Contig5:  25.000
Contig6:     500
Contig7: 130.000
```

A: 498.000
B: 249.000
C: 130.000
D: 100.000
E: 150.000