Questionnaire for Prospective Candidates

Questions Provided by the NGS Team

1. Which of the data structures is fastest for searching / locating exact matches
   1. Burrow Wheeler Transformation
   2. **Associative Arrays**
   3. Burrow Wheeler Transformation + FM Indexing
   4. Myer’s sublinear matching method
2. What is the difference between Smith Waterman algorithm and BLAST?
   1. They are the same
   2. First has a run time that scales as (1-1/d), where d is dependent on size of database
   3. Second has a worst case run time of O(mxn)
   4. **Second is a Heuristic method**
3. Which of the following is not a Next Generation Sequencing platform?
   1. Illumina
   2. Roche 454
   3. **ABI Prism**
   4. ABI SOLiD
4. Which of the following depicts an advantage of NGS over Microarrays, when interrogating transcriptomes?
   1. RNA sequence is portable, while array intensities are not
   2. RNA sequence data is an absolute representation of the transcriptome
   3. **RNA sequence data allows for discovery**
   4. RNA sequence data is cheaper per GB than Microarrays
5. Read Mapping is a memory intensive process. True or False?
   1. True
   2. **False**
6. Which of the following tools perform mapping?
   1. **Bowtie**
   2. **SOAP-denovo**
   3. Inchworm
   4. Newbler
7. What do you do when you want to sequence mRNA
   1. Extract RNA from cells and sequence them
   2. Extract RNA, convert to DNA, and sequence them
   3. Extract RNA, do PCR, convert to DNA and sequence
   4. **Extract RNA, select for poly-A and then proceed to sequencing after conversion to DNA**
8. If you want to BLAST the non-redundant database using a new protein sequence as query, which is the BEST search program to use?
   1. **BLASTP**
   2. BLASTN
   3. TBLASTX
   4. BLASTX
   5. PRSS
   6. BLASTQ
9. If you want to align two sequences that are about 90% identical, which of the following scoring matrices would be most appropriate?
   1. Blosum 35
   2. Blosum 80
   3. **Blosum 90**
   4. Blosum 65
   5. Blosum 10
10. The first mRNA codon to specify an amino acid is likely to be
    1. TAC
    2. UAA
    3. UAG
    4. **AUG**
11. Which mode of information transfer usually does not occur?
    1. DNA to DNA
    2. DNA to RNA
    3. **DNA to protein**
    4. all occur in a working cell
12. A DNA strand with the sequence AACGTAACG is transcribed. What is the sequence of the mRNA molecule synthesized?
    1. AACGTAACG
    2. UUGCAUUGC
    3. **AACGUAACG**
    4. TTGCATTGC
13. Which one of the following is a protein sequence database
    1. DDBJ
    2. EMBL
    3. Genebank
    4. **PIR**
14. Which of the following is not a sequence alignment tool
    1. bowtie
    2. blast
    3. **prosite**
    4. BLAT
15. BLAST X program is used for
    1. translate protein sequence
    2. translate DNA database
    3. **translate input sequence**
    4. none of these
16. PDB is
    1. Primary database for macromolecules
    2. can be determined by gel electrophoresis
    3. composite database
    4. **database for three dimensional struture of biological macromolecules**
17. Phred score of 30 is
    1. Probability of Error = 0.05
    2. Probability of Error = 0.03
    3. Probability of Error = 0.3
    4. **Probability of Error = 0.001**
18. BLAST E-values are calculated based on
    1. Gaussian background distribution
    2. Poisson background distribution
    3. Hypergeometric distribution
    4. **Gumbel’s Distribution**
19. PAM 30 can be calculated as
    1. PAM 1 + 30
    2. PAM 1 – 30
    3. **PAM 1 x 30**
    4. PAM 1 / 30
20. The natural mutation (error) rates of DNA Polymerase III system is
    1. **10-6  Errors / Base**
    2. 10-4 Errors / Base
    3. 10-3 Errors / Base
    4. 10-9 Errors / Base
21. Which of the following genes can be used as the marker genes for metagenomic analysis
    1. **16s rRNA**
    2. mRNA
    3. lnRNA
    4. DNA
22. Which of the following can be used as the functional annotation systems
    1. **GeneOntology**
    2. chromosome position
    3. number of genes
    4. metabolic pathways
23. Which of the following are a short-read aligners
    1. BLAST
    2. Bedtools
    3. **Bowtie**
    4. Samtools
24. Which of the following is a database linking medical condition and variants based on publication
    1. **OMIM**
    2. ExAC
    3. dbSNP
    4. 1000G
25. Which of the following are variant callers
    1. GATK - UnifiedGenotyper
    2. samtools
    3. platypus
    4. **All of the above**
26. Whole exome sequencing covers which part of the genomes
    1. **All exons**
    2. All genes
    3. All genomic region
    4. All non-coding regions
27. Polyphen and SIFT score indicate
    1. Quality of the variant called
    2. Amino acid change in the transcript due to the variant
    3. **Prediction that the variant is deleterious and probability of impact of variant on protein function**
    4. Probability that the variant is a true positive
28. For given sequence below, how many DNA variants are possible without changing the amino acid sequence? (SEQUENCE: MFCW)
    1. 6
    2. **8**
    3. 16
    4. 4
29. Human genome contains about
    1. 2 billion base pairs
    2. **3 billion base pairs**
    3. 4 billion base pairs
    4. 5 billion base pairs
30. Gene duplication has been found to be one of the major reasons for genome expansion in eukaryotes. In general, what would be the selective advantage of gene duplication?
    1. **If one gene copy is nonfunctional, a backup is available**
    2. Larger genomes are more resistant to spontaneous mutations
    3. Duplicated genes will make more of the protein product
    4. Gene duplication will lead to new species evolution
31. What would be a likely explanation for the existence of pseudogenes?
    1. gene duplication
    2. **gene duplication and mutation events**
    3. mutation events
    4. unequal crossing over
32. Which of the following genes can be used as the marker genes for metagenomic analysis
    1. **16s rRNA**
    2. mRNA
    3. lnRNA
    4. DNA
33. Which of the following can be used as the functional annotation systems
    1. **GeneOntology**
    2. chromosome position
    3. number of genes
    4. metabolic pathways
34. Which of the follwing are a short-read aligners
    1. BLAST
    2. Bedtools
    3. **Bowtie**
    4. Samtools
35. In Bisulphite sequencing, which bases are converted in R1 and R2
    1. **C>T in R1 and G>A in R2**
    2. G>A in R1 and C>T in R2
    3. C>T in R1 and C>T in R2
    4. None of the above
36. Structural Variation effects sequence length of about
    1. 1-2 bp
    2. **500bp to 3Mb**
    3. Less than 10 bp
    4. None of these
37. Percentage of CpGs methylated in mammals
    1. 10 %
    2. **60% - 90%**
    3. 4 %
    4. 0.003%
38. Expand BLAT
    1. Basic Local Alignment Tool
    2. **BLAST-like Alignment Tool**
    3. Basic Large Annotation Tool
    4. Basic Local Annotation Tool
39. Which of the following is not *de novo* assembly tool?
    1. Trinity
    2. SOAPdenovo
    3. **Bowtie2**
    4. MASURCA
40. \_\_\_\_\_\_\_\_\_ genes are required for the maintenance of basic cellular function
    1. Pseudogenes
    2. Regulatory genes
    3. **Housekeeping genes**
    4. Protein coding genes
41. Which one of the following is not an annotation tool
    1. BLAST
    2. DAVID
    3. BLAT
    4. **SAMTOOLS**
42. \_\_\_\_\_\_\_\_\_\_\_ is a high-performance visualization tool for Next generation data, micro array and genome annotation
    1. Samtools
    2. Bamtools
    3. **IGV**
    4. MAFFT
43. Literature database include
    1. **MEDLINE and PubMed**
    2. MEDLINE and PDB
    3. PubMed and PDB
    4. MEDLINE and PDS
44. Linux command to search string in a text file is
    1. sudo
    2. awk
    3. **grep**
    4. sed
45. Which of the following technique is used to study DNA-protein interaction
    1. RNA-Seq
    2. DNA-Seq
    3. **Chip-Seq**
    4. Methyl-Seq
46. Genes from different species that evolved from a common ancestral which retain the same function in the course of evolution is know as.
    1. Homologous
    2. **Orthologs**
    3. Paralogs
    4. Xenolog
47. Mitochondrial DNA is
    1. Double stranded linear DNA molecule
    2. Single stranded circular DNA molecule
    3. Single stranded linear DNA molecule
    4. **Double stranded circular DNA molecule**
48. The options to check the syntax of a perl code is
    1. perl –t
    2. perl –w
    3. **perl –c**
    4. perl -X
49. In a family, the father is affected but the mother is normal. The disease is transmitted only to the daughters and not to sons. The type of inheritance is :
    1. Autosomal dominant
    2. Autosomal recessive
    3. **X-linked dominant**
    4. X-linked recessive
50. Human mitochondrial DNA encodes for
    1. 10-20 genes
    2. **30-40 genes**
    3. More than 100 genes
    4. More than 1000 genes