



INCA INSTITUTO NACIONAL DO CÂNCER

CONCURSO PÚBLICO

CARGO 43:
TECNOLOGISTA JÚNIOR

ÁREA:
BIOMEDICINA OU CIÊNCIAS DA COMPUTAÇÃO

ESPECIALIDADE:
BIOINFORMÁTICA

CADERNO DE PROVAS – PARTE II
Conhecimentos Específicos e Discursiva

MANHÃ

LEIA COM ATENÇÃO AS INSTRUÇÕES ABAIXO.

- 1 Nesta parte II do seu caderno de provas, confira atentamente se os seus dados pessoais e os dados identificadores do seu cargo transcritos acima estão corretos e coincidem com o que está registrado em sua folha de respostas e em sua folha de texto definitivo da prova discursiva. Confira também o seu nome e o nome de seu cargo em cada página numerada desta parte de seu caderno de provas. Em seguida, verifique se o seu caderno de provas (partes I e II) contém a quantidade de itens indicada em sua folha de respostas, correspondentes às provas objetivas, e a prova discursiva, acompanhada de espaço para rascunho. Caso o caderno esteja incompleto, tenha qualquer defeito ou apresente divergência quanto aos seus dados pessoais ou quanto aos dados identificadores do seu cargo, solicite ao fiscal de sala mais próximo que tome as providências cabíveis, pois não serão aceitas reclamações posteriores nesse sentido.
- 2 Quando autorizado pelo chefe de sala, no momento da identificação, escreva, no espaço apropriado da **folha de respostas**, com a sua caligrafia usual, a seguinte frase:

A arte e a ciência têm o seu ponto de encontro no método.

OBSERVAÇÕES

- Não serão objeto de conhecimento recursos em desacordo com o estabelecido em edital.
- Informações adicionais: telefone 0(XX) 61 3448-0100; Internet – www.cespe.unb.br.
- É permitida a reprodução deste material apenas para fins didáticos, desde que citada a fonte.

CONHECIMENTOS ESPECÍFICOS

Acerca da bioinformática em geral, julgue os itens subsequentes.

- 41 A bioinformática pode ser definida como uma área multidisciplinar que envolve principalmente a Biologia, a Ciência da Computação, a Matemática e a Estatística.
- 42 O projeto Genoma Humano tinha como objetivo a análise dos genes que compõem o DNA humano. Concluído em 2001, esse projeto contou com o apoio decisivo da bioinformática.
- 43 A biologia computacional é um campo de pesquisa muito próximo ao da bioinformática.
- 44 Enquanto a biologia computacional geralmente trata de ferramentas para organização e recuperação de informações biológicas, a bioinformática geralmente está relacionada ao desenvolvimento de algoritmos.
- 45 Dayhoff, Waterman e Altschul são pesquisadores de relevante importância para o desenvolvimento da bioinformática.

Com relação aos bancos de dados genômicos, julgue os itens a seguir.

- 46 O formato FASTA é composto por uma linha de comentário iniciada por “>”; pela sequência biológica; e por um “*” opcional, que marca o final da sequência.
- 47 O MSS (*multiple sequence standard*) é um formato comum para alinhamentos múltiplos de sequências.
- 48 O AceDB foi um dos primeiros bancos de dados genômicos.
- 49 O NCBI (National Center for Biotechnology Information) mantém o GenBank, que é uma coleção anotada de sequências de DNA disponíveis publicamente.
- 50 A base de dados nr é mantida pelo NCBI, e sua sigla significa *new and representative*.
- 51 O DDBJ, o EMBL e o GenBank são os mais importantes bancos de dados públicos de informação genômica e são sincronizados diariamente.

Quanto à comparação de sequências biológicas, julgue os seguintes itens.

- 52 O objetivo principal da comparação de sequências biológicas é encontrar o(s) alinhamento(s) de maior escore de similaridade.
- 53 O algoritmo de Smith-Waterman permite encontrar o alinhamento ótimo entre duas sequências, porém, devido à sua complexidade exponencial de tempo, sua execução é muito demorada.
- 54 O BLAST é executado em duas grandes fases: sementeira (*seeding*) e extensão.
- 55 O programa ssearch permite implementar uma variação do BLAST.
- 56 O BLAST é um conjunto de programas que podem ser executados via sítio do NCBI.

No que diz respeito ao alinhamento múltiplo de sequências, julgue os próximos itens.

- 57 O alinhamento múltiplo de sequências consiste no alinhamento de mais de duas sequências.
- 58 O alinhamento múltiplo de sequências é bastante utilizado para obtenção das estruturas terciárias das proteínas que compõem o mesmo.
- 59 O CLUSTALW é uma das ferramentas heurísticas mais comuns para alinhamento múltiplo de sequências.
- 60 O HMMER usa modelos coloridos de Markov para alinhar uma sequência de aminoácidos com um determinado perfil (*profile*).
- 61 Os métodos progressivos de alinhamento múltiplo de n sequências possibilitam calcular, na última fase, o alinhamento de todos os possíveis pares de sequências.
- 62 A análise filogenética usa geralmente o alinhamento múltiplo de sequências.
- 63 O alinhamento múltiplo de sequências pode ser local ou global.

Acerca de sequências de macromoléculas aplicadas ao câncer, julgue os itens a seguir.

- 64 O formato de sequências FASTA obedece a um padrão que é definido por uma linha que se inicia pelo símbolo >, que determina uma definição curta sobre a sequência.
- 65 O sequenciamento genômico estrutural de célula extraída de câncer utiliza bibliotecas de cDNA, submetidas a processo de *shot gun*.
- 66 O formato FASTA de sequências nucleotídicas é sempre mostrado em letras minúsculas.
- 67 O método SAGE provê uma análise quantitativa da expressão gênica, com a vantagem de conseguir detectar transcritos desconhecidos em uma hiperplasia.

Considerando o alinhamento de uma sequência nucleotídica desconhecida em relação a um banco de dados de genes associados ao câncer, julgue os itens seguintes.

- 68 No resultado de alinhamento, por meio do programa BLAST, de gene associado ao câncer de mama, o *e-value* corresponde ao inverso da probabilidade resultante de um alinhamento que seja obtido ao acaso com uma sequência do banco de dados.
- 69 Duas sequências similares e com a mesma função são ditas homólogas e não têm ancestral comum.
- 70 A matriz de penalidade BLOSUM62 considera os blocos de sequência com, no máximo, 62% de identidade.
- 71 O programa BLASTp permite a comparação de uma sequência de aminoácidos com todas as proteínas de cada sequência de aminoácido possível, calculada a partir de cada sequência de nucleotídeos depositada no banco, nos 6 quadros de leitura, além das sequências depositadas diretamente como *string* (sequência) de aminoácidos.
- 72 Parálogos são genes que possuem a mesma função na mesma espécie, em consequência de fenômenos de duplicação gênica.
- 73 Ortólogos são genes que possuem funções diferentes em espécies diferentes.

Uma revolução está em curso: a abordagem ao estudo da base genética do câncer. Enormes quantidades de dados são geradas por meio de técnicas de alto desempenho, como a tecnologia de microarranjos de DNA, e novos algoritmos computacionais têm sido desenvolvidos para auxiliar a análise desses dados. Ao mesmo tempo, as normas de repositórios de base, incluindo o banco de dados de Microarranjo de Stanford e da expressão do gene *omnibus*, têm sido desenvolvidas para armazenar e divulgar os resultados de experimentos de microarranjo.

Adaptado de D. A. Hanauer *et al.* *Bioinformatics Approaches in the Study of Cancer*. p.133-41.

Tendo o texto apresentado acima apenas como referência inicial, julgue os itens subsequentes.

- 74** A tecnologia de microarranjos baseia-se na hibridização entre as proteínas derivadas de um espécime biológico que são marcadas por fluorocromos e uma sonda de DNA de uma amostra desconhecida.
- 75** Em um experimento de microarranjo, as imagens constituem os dados da análise de expressão gênica, sendo diretamente obtidas no processo de endereçamento das regiões, sem a influência dos *spots*.
- 76** Na quantificação dos *spots* de um microarranjo, a intensidade é igual à diferença entre a mediana da intensidade de luz (vermelha ou verde) e a mediana da intensidade de luz do *background*, multiplicada pela área do círculo.

Cientistas têm observado que os avanços tecnológicos no sequenciamento do DNA se espelharam na lei de Moore para microprocessadores, isto é, o número de elementos funcionais que podem ser colocados em um *chip* por custo unitário tem dobrado a cada dezoito meses, nas últimas décadas. De fato, as novas gerações de sequenciadores de DNA estão aumentando a velocidade de leitura muito mais rapidamente que o previsto pela lei de Moore. O primeiro genoma humano sequenciado demorou três ou quatro anos para ficar pronto e custou cerca de US\$ 300 milhões. Acredita-se que, em dez anos, uma sequência do genoma humano individual vai custar menos de US\$ 1 mil – 300 mil vezes menos do que custou a primeira –, e poderá ser feita em apenas um dia. Na próxima década, avanços similares em outras tecnologias biomédicas relevantes possibilitarão o aparecimento de uma medicina previsiva e personalizada. Com relação ao sequenciamento de genoma, julgue os próximos itens.

- 77** No pirosequenciamento, ocorre a amplificação do DNA por PCR em emulsão, e a incorporação de cada nucleotídeo acarreta a liberação de pirofosfato.
- 78** O sequenciador Illumina utiliza a técnica de sequenciamento por síntese, em que um nucleotídeo é incorporado por vez, em uma reação catalizada pelo DNA polimerase fixada em uma nanoesfera.
- 79** O formato padrão dos arquivos produzidos pelo sequenciador SOLID foi nomeado *Standard Flowgram File* (SFF) e pode conter dados de uma ou várias sequências.
- 80** A partir de um genoma de referência, pode-se mapear as sequências produzidas pelas novas tecnologias de sequenciamento e agrupá-las conforme sua posição relativa, e, a partir desse mapeamento, identificar polimorfismos responsáveis pela manifestação de doenças como o câncer.
- 81** O sequenciador 454 produz sequências de cerca de 250 bases de comprimento, enquanto o Illumina obtém fragmentos de DNA com 45 bases, no máximo.

Julgue os itens que se seguem, acerca da organização de bancos de dados de genes.

- 82** Anotação genômica é o processo de predição funcional de genes com base em comparações com bancos de dados.
- 83** No banco de dados de ESTs (etiquetas de sequências expressas) armazenado no GenBank, estão estocadas sequências curtas de cerca de 200 a 500 nucleotídeos e identificadas como únicas em um genoma resultante de um amplicon pela reação da polimerase em cadeia (PCR).
- 84** As bases de dados COG e KOG são bases de dados de proteínas ortólogas de organismos procariotos (COG) e eucariotos (KOG).

O microarranjo é uma técnica que busca medir os níveis de expressão de transcritos em larga escala. A tecnologia dos microarranjos tem impulsionado, de maneira importante, a pesquisa de genômica funcional de diferentes organismos, desde a bactéria até o ser humano, em situações normais e patológicas (câncer, doenças autoimunes, doenças degenerativas, entre outras). Considerando os processamentos bioinformáticos dos experimentos de microarranjo, julgue os itens a seguir.

- 85** De modo geral, os passos para a análise de microarranjos são: preparação, aquisição, processamento, análise da imagem, quantificação e normalização e, finalmente, a identificação dos genes diferencialmente expressos.
- 86** As medidas de intensidade de fluorescência, em cada canal de cor da imagem, podem ser distorcidas durante os passos de preparação de dados. A normalização de dados desconsidera condições experimentais como as variações de intensidade e hibridização.
- 87** A normalização utiliza descritores estatísticos que incluem a média, mediana, moda ou percentil da distribuição da intensidade das amostras no arranjo.
- 88** O desenho experimental de um microarranjo permite a identificação de genes diferencialmente expressos e o desenvolvimento de novas ferramentas de diagnóstico.
- 89** A técnica de microarranjo permite a detecção de centenas e milhares de alterações em genes codificadores de proteínas e anormalidades em genes não codificadores.
- 90** A análise não supervisionada define classes de tumores por meio do agrupamento de genes ou espécimes, com base nas semelhanças dos perfis de expressão gênica, sem qualquer organização inicial.
- 91** Em câncer de mama, os experimentos de microarranjo têm, consistentemente, separado tumores com base na presença ou ausência da expressão do receptor de estrogênio (ER).

A tecnologia SAGE está baseada no sequenciamento de pequenos fragmentos de cDNA de uma região relativa do transcrito, cujo princípio se fundamenta no pressuposto de que o pequeno fragmento (*tag*) contém informação suficiente para identificar, univocamente, um gene. A respeito das aplicações de SAGE nas pesquisas do câncer, julgue os itens a de 92 a 100.

92 O método DDD (*digital differential display*) determina a significância da diferença de frequência de ESTs de um mesmo conjunto de bibliotecas de cDNA.

93 As sequências derivadas de bibliotecas de cDNA são processadas para identificar e eliminar as regiões de baixa qualidade, regiões de vetores, *primers* e outros contaminantes.

94 No banco de dados Genbank, a subdivisão RefSeq é uma coleção curada de sequências redundantes de dados de DNA genômico, transcritos e proteínas.

95 Os microRNAs (miRNAs) podem estar localizados em regiões intergênicas do genoma e possuir sua própria unidade de transcrição.

96 Os miRNA são genes RNA que são o complemento inverso de porções de transcritos de outros genes.

97 Os polimorfismos de nucleotídeos únicos (SNPs) são o tipo de variação muito rara no genoma humano.

98 Na identificação de polimorfismos de nucleotídeos únicos, para entender as bases moleculares do câncer, são usadas apenas as sequências derivadas de DNA genômico.

99 As aplicações de identificação de SNPs auxiliam o mapeamento genômico e estudos funcionais e farmacogenéticos.

100 Um banco de dados pode ser considerado uma coleção de dados inter-relacionados, projetado para suprir as necessidades de um grupo específico de aplicações e usuários.

PROVA DISCURSIVA

- Nesta prova, faça o que se pede, usando o espaço para rascunho indicado no presente caderno. Em seguida, transcreva o texto para a **FOLHA DE TEXTO DEFINITIVO DA PROVA DISCURSIVA**, no local apropriado, pois **não serão avaliados fragmentos de texto escritos em locais indevidos**.
- Qualquer fragmento de texto além da extensão máxima de linhas disponibilizadas será desconsiderado.
- Na **folha de texto definitivo**, identifique-se apenas no cabeçalho da primeira página, pois **não será avaliado** texto que tenha qualquer assinatura ou marca identificadora fora do local apropriado.

As novas tecnologias de sequenciamento genômico, como o pirosequenciamento, a tecnologia Illumina e os sequenciadores SOLID, reduziram os custos e incrementaram a velocidade de aquisição de dados em uma magnitude nunca antes imaginada. O uso clínico dos dados de sequenciamento de alto desempenho trazem um novo modo de identificar as causas de muitas doenças de etiologia desconhecida por meio da varredura simultânea de milhares de loci em busca de mutações associadas a patogênese e do sequenciamento de espécies de novo. Adicionalmente, as novas tecnologias de sequenciamento poderão substituir as tecnologias de microarranjos e do sequenciamento de Sanger na identificação de sequências diferencialmente expressas, bem como nas aplicações clínicas e no desenvolvimento de ferramentas de diagnóstico.

Considerando que o fragmento de texto acima tem caráter unicamente motivador, redija um texto dissertativo acerca do seguinte tema.

SEQUENCIAMENTO DE NOVA GERAÇÃO APLICAÇÃO AO TRANSCRITOMA DO CÂNCER

Ao elaborar seu texto, atenda, necessariamente, as seguintes determinações:

- ▶ discorra a respeito das vantagens e desvantagens tecnológicas do sequenciamento 454;
- ▶ descreva os princípios da metodologia Illumina;
- ▶ descreva uma aplicação para a detecção de polimorfismos em regiões codificadoras no genoma humano.

RASCUNHO

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	