

CADERNO DE QUESTÕES



HOSPITAL DE
CLÍNICAS
PORTO ALEGRE RS

EDITAL Nº 07/2023

DE PROCESSOS SELETIVOS (PS)

Cargo de Nível Superior

PS 49 - ASSISTENTE I

(Assistente de Pesquisa em Bioinformática)

	MATÉRIA	QUESTÕES	PONTUAÇÃO	
	Conhecimentos Específicos	01 a 25	0,40 cada	

ATENÇÃO

Transcreva no espaço apropriado da sua FOLHA DE RESPOSTAS (Folha Óptica), com sua caligrafia usual, considerando as letras maiúsculas e minúsculas, a seguinte frase:

Não há amargo mais doce que o do chimarrão.

Nome do Candidato: _____

Inscrição nº: _____

DIREITOS AUTORAIS RESERVADOS. PROIBIDA A REPRODUÇÃO, AINDA QUE PARCIAL, SEM A PRÉVIA AUTORIZAÇÃO DA FAURGS E DO HCPA.



FAURGS

Fundação de Apoio da Universidade Federal do Rio Grande do Sul





EDITAL Nº 07/2023 DE PROCESSOS SELETIVOS

GABARITO APÓS RECURSOS

PROCESSO SELETIVO 49

ASSISTENTE I (Assistente de Pesquisa em Bioinformática)

01.	C	11.	A	21.	D
02.	A	12.	C	22.	A
03.	D	13.	C	23.	B
04.	C	14.	A	24.	E
05.	C	15.	B	25.	B
06.	E	16.	D		
07.	A	17.	E		
08.	E	18.	E		
09.	D	19.	B		
10.	B	20.	C		

INSTRUÇÕES

- 1 Verifique se este CADERNO DE QUESTÕES corresponde ao Processo Seletivo para o qual você está inscrito. Caso não corresponda, solicite ao Fiscal da sala que o substitua.
- 2 Esta PROVA consta de **25** (vinte e cinco) questões objetivas.
- 3 Caso o CADERNO DE QUESTÕES esteja incompleto ou apresente qualquer defeito, solicite ao Fiscal da sala que o substitua.
- 4 Para cada questão objetiva, existe apenas **uma** (1) alternativa correta, a qual deverá ser assinalada na FOLHA DE RESPOSTAS.
- 5 O candidato que comparecer para realizar a prova **não deverá, sob pena de ser excluído do certame**, portar armas, malas, livros, máquinas calculadoras, fones de ouvido, gravadores, *paggers*, *notebooks*, telefones celulares, *pen drives* ou quaisquer outros tipos de aparelhos eletrônicos, nem utilizar véus, bonés, chapéus, gorros, mantas, lenços, aparelhos/próteses auditivas, óculos escuros, ou qualquer outro adereço que lhes cubra a cabeça, o pescoço, os olhos, os ouvidos ou parte do rosto, **exceto em situações autorizadas pela Comissão do Concurso e/ou determinadas em lei. Os relógios de pulso serão permitidos, desde que permaneçam sobre a mesa, à vista dos fiscais, até a conclusão da prova.** (conforme subitem 7.10 do Edital de Abertura)
- 6 **É de inteira responsabilidade do candidato comparecer ao local de prova munido de caneta esferográfica, preferencialmente de tinta azul, de escrita grossa, para a adequada realização de sua Prova Escrita. Não será permitido o uso de lápis, marca-textos, régua, lapiseiras/grafites e/ou borrachas durante a realização da prova.** (conforme subitem 7.15.2 do Edital de Abertura)
- 7 Não será permitida nenhuma espécie de consulta em livros, códigos, revistas, folhetos ou anotações, nem o uso de instrumentos de cálculo ou outros instrumentos eletrônicos, exceto nos casos em que forem pré-estabelecidos no item 13 do Edital. (conforme subitem 7.15.3 do Edital de Abertura)
- 8 Preencha com cuidado a FOLHA DE RESPOSTAS, evitando rasuras. Eventuais marcas feitas nessa FOLHA a partir do número **26** serão desconsideradas.
- 9 Ao terminar a prova, entregue a FOLHA DE RESPOSTAS ao Fiscal da sala.
- 10 A duração da prova é de **duas horas e trinta minutos (2h30min)**, já incluído o tempo destinado ao preenchimento da FOLHA DE RESPOSTAS. Ao final desse prazo, a FOLHA DE RESPOSTAS será **imediatamente** recolhida.
- 11 **O candidato somente poderá se retirar da sala de prova uma hora (1h) após o seu início. Se quiser levar o Caderno de Questões da Prova Escrita, o candidato somente poderá se retirar da sala de prova uma hora e meia (1h30min) após o início. O candidato não poderá anotar/copiar o gabarito de suas respostas de prova.**
- 12 **Após concluir a prova e se retirar da sala, o candidato somente poderá utilizar os sanitários nas dependências do local de prova se for autorizado pela Coordenação do Prédio e se estiver acompanhado de um fiscal.** (conforme subitem 7.15.6 do Edital de Abertura)
- 13 Ao concluir a Prova Escrita, o candidato deverá devolver, obrigatoriamente, ao fiscal da sala a Folha de Respostas (Folha Óptica). Se assim não proceder, será excluído do Processo Seletivo. (conforme subitem 7.15.8 do Edital de Abertura)
- 14 A desobediência a qualquer uma das recomendações constantes nas presentes instruções poderá implicar a anulação da prova do candidato.



01. Em um estudo, foram obtidos dados de expressão gênica de 37.420 genes para um conjunto de 500 amostras de células. O objetivo do estudo é identificar genes que estão envolvidos na regulação de um processo biológico específico e posteriormente induzir um modelo preditivo. As amostras de células foram coletadas de dois grupos de indivíduos, um grupo saudável (400 amostras) e um grupo com uma doença específica (100 amostras). Qual dos seguintes algoritmos de aprendizado de máquina é o mais adequado para extração de um modelo lógico-matemático e indução de um modelo preditivo a partir desse conjunto de dados?

- (A) Redes neurais convolucionais.
- (B) Regressão logística multinomial.
- (C) Redes neurais artificiais *feedforward*.
- (D) DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*).
- (E) *Random Forest Feature Elimination*.

02. Considere um arquivo em formato *Extensible Markup Language (XML)* (arquivo.xml) contendo as seguintes informações sobre uma rede regulatória:

```
<network>
  <gene id="OCA2">
    <regulacao>
      <tipo>Ativacao</tipo>
      <fator_transcricional>FT1</fator_transcricional>
    </regulacao>
  </gene>
  <gene id="HERC2">
    <regulacao>
      <tipo>Repressao</tipo>
      <fator_transcricional>FT2</fator_transcricional>
    </regulacao>
  </gene>
</network>
```

Qual dos seguintes comandos BASH utiliza expressões regulares e que, ao ser executado, em um terminal Linux, retorna apenas o nome dos genes (OCA2 e HERC2) do referido arquivo XML?

- (A) `grep -oE "\\(\\w+\\)" arquivo.xml`
- (B) `grep "\\(\\w+\\)" arquivo.xml`
- (C) `grep -o "\\(\\w+\\)" arquivo.xml`
- (D) `grep -P "\\(\\w+\\)" arquivo.xml`
- (E) `grep -oS "\\(\\w+\\)" arquivo.xml`

03. Considere um arquivo no formato VCF (*Variant Call Format*) que contém informações sobre SNPs (*single nucleotide polymorphisms*) de um conjunto de 1.000 amostras. Cada SNP é representado por uma linha no arquivo, e cada linha contém informações sobre a posição do SNP, o identificador do SNP (rsID), os alelos referência e os alelos alternativos, e as informações genotípicas para cada amostra. Ao usar a linguagem de programação Python para implementar uma estrutura de dados que permita a manipulação das informações, qual das seguintes alternativas apresenta a implementação de uma estrutura de dados mais adequada para armazenar todas as informações?

- (A) Uma lista de dicionários, onde cada dicionário representa um SNP e contém as informações relevantes como chaves.
- (B) Um dicionário onde as chaves são as posições dos SNPs e os valores são os alelos alternativos.
- (C) Uma tupla para cada amostra contendo as informações de posição do SNP, o identificador do SNP (rsID) e os alelos referência.
- (D) Uma classe chamada SNP com atributos para cada informação presente em uma linha do arquivo VCF e uma lista de instâncias dessa classe para representar todos os SNPs.
- (E) Uma estrutura de dados baseada em grafos para armazenar todas as informações das amostras.

04. Foi realizado um experimento cujo principal objetivo era avaliar a composição do microbioma e do resistoma no trato intestinal de modelos animais submetidos a tratamentos com probióticos associados ou não a drogas antimicrobianas convencionais. Duas diferentes drogas foram testadas associadas ou não a um produto probiótico. Ao todo, cinco grupos experimentais foram formados, sendo que cada grupo continha 20 animais. Amostras de fezes, no momento da eutanásia, foram coletadas individualmente, processadas, e o DNA metagenômico foi sequenciado em uma plataforma de NGS (*Next Generation Sequencing*). A fim de se atingir o principal objetivo do experimento, a estratégia metodológica de sequenciamento mais adequada

- (A) é a realização do sequenciamento das amostras pela estratégia de *Shotgun*, em uma profundidade a partir de 10.000.000 de *reads*, pois possibilitará, com alta confiabilidade, a taxonomia em nível de espécie e a identificação precisa da coleção de genes de resistência a antimicrobianos.
- (B) é a realização do sequenciamento das amostras pela estratégia de amplificação e sequenciamento de *amplicons*, pois permitirá a identificação do microbioma presente nas fezes dos animais apresentando a taxonomia em nível de espécie para procariotos e eucariotos, assim como a identificação de vírus e a determinação dos genes de resistência a antimicrobianos.
- (C) é o sequenciamento de amostras pela estratégia de *Shotgun*, a depender da profundidade e da qualidade das *reads* geradas, pois possibilitarão a determinação taxonômica do microbioma em nível de espécie e a identificação de genes de resistência a antimicrobianos.
- (D) é o sequenciamento de *amplicons*, levando em conta o custo e o tempo dispendidos nas análises, construindo quatro bibliotecas para cada amostra: uma com *amplicons* bacterianos, uma com *amplicons* eucarióticos, uma com *amplicons* virais e uma com *amplicons* de genes marcadores de resistência a antimicrobianos.
- (E) é o emprego do sequenciamento por *Shotgun*, para assim identificar o microbioma, associado ao sequenciamento de *amplicons*, para identificar o resistoma.

05. Ao realizar operações avançadas de manipulação e análise de grandes conjuntos de dados, qual das alternativas apresenta a abordagem mais eficiente e eficaz para processar dados na linguagem de programação R?

- (A) Utilizar apenas estruturas de dados padrão, como *data frames* e listas, para armazenar e manipular os dados.
- (B) Utilizar o Pacote *h2o* para manipulação de dados, aplicando funções como *filter*, *mutate*, e *group_by*.
- (C) Utilizar o Pacote *data.table* para realizar operações de manipulação de dados, aproveitando a eficiência e a velocidade oferecidas por essa ferramenta.
- (D) Utilizar laços de repetição *for* e *while* para percorrer os dados e realizar as operações necessárias.
- (E) Utilizar o Pacote *caret* para manipulação e análise dos dados, incluindo Pacotes como *dplyr*, *ggplot2* e *purrr*.

06. Considere um projeto de pesquisa que envolve o processamento de dados WGS (*Whole Genome Sequencing*) de 100.000 indivíduos (genoma humano). Todas as amostras foram sequenciadas usando tecnologia WGS Illumina. O arquivo bruto (*raw data*) de uma única amostra possui aproximadamente 10 GB (gigabytes) de tamanho. Os dados são complexos, e as análises a serem realizadas possuem alto custo computacional em termos de processamento e uso de grandes quantidades de memória RAM. Qual das seguintes arquiteturas de computação em nuvem (*cloud computing*) é a mais adequada para execução do projeto e obtenção dos resultados?

- (A) Uma arquitetura de computação em nuvem baseada em contêineres, por conta do alto desempenho.
- (B) Uma arquitetura de computação em nuvem baseada em microsserviços, por conta da alta flexibilidade.
- (C) Uma arquitetura de computação em nuvem baseada em virtualização, por conta da possibilidade de criação de máquinas heterogêneas.
- (D) Uma arquitetura de computação em nuvem baseada em GPU (*graphics processing unit*), por conta do alto desempenho e da flexibilidade.
- (E) Uma arquitetura de computação em nuvem baseada em computação distribuída, por conta da escalabilidade.

07. O GenBank é um banco de dados de sequências de nucleotídeos mantido pelo *National Center for Biotechnology Information* (NCBI), dos Estados Unidos. É o maior banco de dados de sequências biológicas do mundo, armazenando sequências de nucleotídeos de uma ampla variedade de organismos, incluindo bactérias, archaea, fungos, plantas, animais e humanos. Qual dos seguintes modelos de banco de dados é mais adequado para o armazenamento de sequências biológicas organizadas de forma estruturada?

- (A) Modelo relacional.
- (B) Modelo NoSQL baseado em grafos.
- (C) Modelo hierárquico.
- (D) Modelo NoSQL baseado em família de colunas.
- (E) Modelo NoSQL baseado em chave-valor.

08. A Lei Geral de Proteção de Dados (LGPD) (Lei nº 13.709/2018) determina o modo e as condições em que os dados pessoais podem ser tratados, define hipóteses de tratamento, os direitos dos titulares, as obrigações e as consequências. É permitido o tratamento de dados pessoais sensíveis sem fornecimento de consentimento do titular, nas hipóteses em que for indispensável para as situações nas alternativas abaixo, **EXCETO** em uma delas. Assinale-a.

- (A) Tratamento compartilhado de dados necessários à execução, pela administração pública, de políticas públicas previstas em leis ou regulamentos.
- (B) Tutela da saúde, exclusivamente, em procedimento realizado por profissionais de saúde, serviços de saúde ou autoridade sanitária.
- (C) Realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais sensíveis.
- (D) Proteção da vida ou da incolumidade física do titular ou de terceiro.
- (E) Cumprimento de deveres legais, morais ou regulatórios pelo controlador.

09. Sobre os conceitos em epigenética, assinale com **V** (verdadeiro) ou **F** (falso) as seguintes afirmações.

- () Uma variante epigenética não é uma variação genética, porém pode ser herdada pelas próximas gerações.
- () Uma variante epigenética é uma variação genética, pois altera a sequência de DNA, sendo herdada pelas próximas gerações.
- () Uma variante epigenética determinará de modo irreversível uma mudança fenotípica.
- () Uma variante epigenética não está relacionada à expressão de fenótipos, mas sim a mudanças genéticas herdáveis.

A sequência correta de preenchimento dos parênteses, de cima para baixo, é

- (A) V – F – V – F.
- (B) F – F – V – V.
- (C) F – V – V – F.
- (D) V – F – F – F.
- (E) V – V – F – V.

10. Considerando a Lei nº 8.080/1990, que dispõe sobre as condições para a promoção, proteção e recuperação da saúde, a organização e o funcionamento dos serviços correspondentes e dá outras providências, no que se refere à assistência terapêutica e da incorporação de tecnologia em saúde, considere os itens abaixo.

- I - Os critérios para o diagnóstico da doença ou do agravo à saúde.
- II - O tratamento preconizado, sem referência a posologias.
- III- Os mecanismos de controle clínico.
- IV - O acompanhamento e a verificação dos resultados terapêuticos a serem seguidos pelos pacientes.

Quais dos itens acima um protocolo clínico e diretriz terapêutica deverão conter?

- (A) Apenas III.
- (B) Apenas I e III.
- (C) Apenas II e IV.
- (D) Apenas III e IV.
- (E) Apenas I, II e IV.

11. Um estudo randomizado buscando identificar o perfil microbiano da pele de colaboradores de diferentes setores de um hospital público foi realizado. Ao todo foram coletados swabs de pele de 300 colaboradores, sendo três sítios de coleta, totalizando 900 amostras. Considere os itens abaixo sobre o processamento dos dados obtidos, a partir de sequenciamento por *Shotgun*.

- I - A determinação no *raw dataset* da qualidade das bases sequenciadas, seguida da identificação e eliminação de contaminantes.
- II - A determinação nos *contigs* da qualidade das bases sequenciadas, seguida da identificação e eliminação de contaminantes.
- III - A classificação taxonômica das sequências contra um conjunto de dados privado construído pelo próprio pesquisador.
- IV - A classificação taxonômica da montagem, para posterior quantificação das ASVs (*amplicon sequence variant*), e observação dos índices de diversidade para cada amostra.

Quais esse processamento de dados deve prever?

- (A) Apenas I.
- (B) Apenas IV.
- (C) Apenas I e IV.
- (D) Apenas II e III.
- (E) Apenas I, II e III.

12. Um pesquisador pretende utilizar o banco de dados Ensembl para estudar a evolução de um gene. O objetivo é comparar a sequência de genes em diferentes espécies. O que pode ser feito para melhorar a precisão das comparações?

- (A) Usar uma ferramenta de busca de sequências para comparar as sequências dos genes entre as diferentes espécies.
- (B) Usar uma ferramenta de bioinformática para alinhar as sequências dos genes em diferentes espécies utilizando um *cutoff* de 25% de identidade.
- (C) Usar uma ferramenta de bioinformática para remover as mutações silenciosas das sequências dos genes.
- (D) Usar uma ferramenta de bioinformática para excluir as mutações nos genes das diferentes espécies.
- (E) Usar uma ferramenta de bioinformática para construir uma sequência média (*average sequence*) de cada gene entre as diferentes espécies.

13. Os estudos de sequenciamento genético iniciaram-se na década de 1970. Desde então, as tecnologias de sequenciamento vêm evoluindo e permitindo expressivos avanços nas mais diversas áreas do conhecimento. Analise as seguintes afirmativas referentes às metodologias de sequenciamento de ácidos nucleicos atualmente disponíveis.

- I - As tecnologias de sequenciamento de ácidos nucleicos denominadas de NGS (*Next Generation Sequencing*) apresentam como grande diferencial o processamento massivo de amostras, que é possível pelo uso de adaptadores específicos e determinados para cada tecnologia.
- II - Metodologias de segunda e terceira geração de NGS são baseadas na amplificação dos ácidos nucleicos a serem sequenciados, sendo essa amplificação essencial para a captação do sinal de leitura em equipamentos de ambas as gerações.
- III - A tecnologia de sequenciamento por terminadores reversíveis tem como principal característica, frente a outras tecnologias, o fundamento da fragmentação de DNA com alta integridade e posterior ligação de terminadores reversíveis à porção 5' e 3' dos fragmentos gerados.
- IV - A base do processo, em todas as metodologias de sequenciamento, para a obtenção da sequência de um genoma bacteriano completo é a fragmentação do DNA, a construção das bibliotecas e a leitura das bases. Posteriormente, o sucesso na montagem do conjunto de dados obtido, com fechamento do genoma bacteriano, está primariamente vinculado à qualidade das sequências geradas.

Quais estão corretas?

- (A) Apenas I.
- (B) Apenas I e III.
- (C) Apenas I e IV.
- (D) Apenas II e III.
- (E) Apenas II e IV.

14. Uma enfermeira de um hospital, no seu penúltimo dia de férias, sofreu um acidente com uma faca, o que gerou uma lesão na mão direita, que infeccionou. Para o retorno ao trabalho, conforme a Norma Regulamentadora nº 32 (NR-32) – Segurança e saúde no trabalho em serviços de saúde –, que prevê medidas de proteção contra riscos biológicos, a enfermeira deve

- (A) buscar avaliação médica para a emissão de documento de liberação para o trabalho, caso o médico julgar viável o retorno imediato.
- (B) comunicar à chefia imediata que tem uma lesão na mão direita e se comprometer a usar luvas durante todos os procedimentos laborais.
- (C) lavar as mãos antes de qualquer procedimento que possa trazer riscos de contaminação para a lesão em sua mão direita e não usar luvas.
- (D) manter-se afastada de suas atividades laborais até que a ferida esteja completamente cicatrizada e, após a cicatrização, buscar avaliação médica para a emissão de documento de liberação para o trabalho.
- (E) comunicar à chefia imediata que tem uma lesão na mão direita e se comprometer, em documento por escrito, a usar luvas durante todos os procedimentos laborais.

15. Um bioinformata está utilizando um *pipeline* para mineração de dados para identificar genes que estão associados a um determinado tipo de câncer. O *pipeline* utilizado, neste estudo, consiste em cinco etapas. Em qual delas o bioinformata está mais propenso a cometer um erro que pode levar a resultados incorretos?

- (A) Etapa 1 – coleção de dados: são coletados dados de expressão gênica de pacientes com câncer e pacientes saudáveis.
- (B) Etapa 2 – pré-processamento dos dados: os dados de expressão gênica são normalizados.
- (C) Etapa 3 – redução de dimensionalidade: a dimensionalidade dos dados é reduzida para torná-los mais gerenciáveis.
- (D) Etapa 4 – aplicação de técnicas de agrupamento: são aplicados algoritmos de agrupamento para identificar grupos de genes que são coexpressos.
- (E) Etapa 5 – aplicação de algoritmos baseados em regras de associação: são utilizadas regras de associação para identificar genes que estão associados à patologia.

16. Um pesquisador decidiu usar técnicas estatísticas para identificar genes que estão associados a um determinado tipo de doença. O pesquisador obteve dados de expressão gênica de pacientes com a doença e de uma amostra de pacientes saudáveis. Ele usou um teste de hipótese para comparar a expressão gênica nas duas condições. O teste de hipótese fornece um p-valor de 0,05. Sobre uma possível conclusão desse teste, assinale a alternativa correta.

- (A) O pesquisador pode concluir que os genes com p-valores menores que 0,05 estão associados com a doença.
- (B) O pesquisador pode concluir que os genes com p-valores maiores que 0,05 não estão associados com a doença.
- (C) O pesquisador pode concluir que os genes com p-valores iguais a 0,05 estão associados com a doença.
- (D) O pesquisador não pode concluir nada sobre a associação entre os genes e a doença com base no p-valor.
- (E) O pesquisador pode concluir que apenas 5% dos genes estão associados com a doença.

17. Um bioinformata recebeu dados de polimorfismos de nucleotídeo único de 300 pacientes saudáveis e de 1.200 pacientes com uma doença neurodegenerativa. Foi utilizado o algoritmo *Random Forest* para indução de um modelo lógico-matemático de classificação. Qual das seguintes métricas é a mais confiável para avaliar a qualidade do classificador induzido?

- (A) Acurácia.
- (B) Precisão.
- (C) Erro médio absoluto (MAE).
- (D) Erro médio quadrado (MSE).
- (E) F1-Score.

18. Considere as seguintes afirmações sobre redes biológicas.

- I - Redes de coexpressão representam as interações entre genes com base na correlação entre seus níveis de expressão.
- II - Redes de interações proteína-proteína representam as interações diretas entre proteínas.
- III- Redes de interações entre genes e proteínas representam as interações indiretas entre genes, mediadas por proteínas.
- IV - Análises de redes biológicas podem ser usadas para identificar genes ou proteínas com funções importantes em um sistema biológico.

Quais estão corretas?

- (A) Apenas III.
- (B) Apenas I e III.
- (C) Apenas II e IV.
- (D) Apenas III e IV.
- (E) I, II, III e IV.

19. Considere uma rede de interações proteína-proteína (PPI) com 200 proteínas. As proteínas são representadas por vértices, e as interações, por arestas. Considere as seguintes afirmações sobre a análise topológica das redes.

- I - A análise de centralidade (*centrality analysis*) fornece uma estimativa da importância de um vértice para a conectividade ou fluxo de informações da rede.
- II - O grau das arestas pode ser usado como uma estimativa aproximada da centralidade.
- III- A centralidade de proximidade (*closeness centrality*) é uma medida que estima a rapidez com que o fluxo de informações passa por um determinado vértice para outros vértices.
- IV - A rede obtida é necessariamente acíclica.
- V - Os vértices com alta centralidade de intermediação (*betweenness centrality*) oferecem informações sobre o controle do fluxo de informações. Esses nós podem representar proteínas importantes nas vias de sinalização e podem formar alvos para a descoberta de medicamentos.

Quais estão corretas?

- (A) Apenas I, II e V.
- (B) Apenas I, III e V.
- (C) Apenas II, III e IV.
- (D) Apenas II, IV e V.
- (E) Apenas I, II, III e IV.

20. A triagem virtual de fármacos (*virtual screening*) é uma técnica que utiliza modelos computacionais para identificar compostos químicos com potencial para atuar como fármacos. A simulação de atracamento molecular é um processo computacional que simula a interação entre o composto químico e o alvo molecular. Sobre esta etapa, qual das seguintes afirmações é **INCORRETA**?

- (A) O algoritmo de atracamento molecular prevê a orientação do ligante em relação ao receptor.
- (B) Funções de pontuação (*scoring*) empíricas ou baseadas em conhecimento são utilizadas pelos métodos de atracamento buscando identificar a orientação de maior afinidade entre o ligante e o receptor.
- (C) Em uma função de pontuação (*scoring*) baseada em conhecimento, os fenômenos físicos, ou seja, a entropia e as interações eletrostáticas, são considerados durante o processo de atracamento.
- (D) Um método de busca por gradiente explora o espaço de busca a partir de uma pose inicial, seguindo o gradiente da função de pontuação (*scoring*).
- (E) Métodos de busca heurística podem ser utilizados para explorar o espaço de busca de afinidade entre o ligante e o receptor.

21. Um pesquisador recebeu dados de um projeto contendo *short*, *long* e *ultra-long reads* advindas de sequenciamentos de genomas humanos. O primeiro objetivo do pesquisador é empregar um montador híbrido para a montagem dos genomas. Para isso, o pesquisador deve observar

- (A) que, após análise de controle de qualidade, o próximo passo no processo é a trimagem das *short*, *long* e *ultra-long raw reads*, usando o *software* Trimmomatic com valor de Phred33.
- (B) que os montadores híbridos levam em conta, para o mapeamento das *long* e *ultra-long reads*, o tamanho dos contíguos, mas não os k-mers compartilhados entre eles.
- (C) que, empregando a abordagem de montagem híbrida, utilizando exclusivamente grafos de Bruijn, ele obterá os resultados mais robustos.
- (D) que abordagens de montagem por grafos de Bruijn ou *overlap-layout-consensus*, ou a combinação delas, podem ser empregadas por montadores híbridos.
- (E) que, para a obtenção de resultados confiáveis, a estratégia de montagem deverá prever o processamento das *short*, *long* e *ultra-long reads* com valor inferior a Phred10.

22. Em um conjunto de genomas totais sequenciados de pacientes com diferentes tipos de tumores, será realizada a análise de eventos epigenômicos presentes nesses genomas. Para essas buscas, é correto afirmar

- (A) que, de acordo com a tecnologia de sequenciamento, os dados não são adequados para a busca se não foram, previamente ao sequenciamento, realizados tratamentos específicos nos ácidos nucleicos para a marcação dos possíveis eventos epigenômicos.
- (B) que a identificação de eventos epigenômicos em genomas humanos deve, exclusivamente, ser realizada em sequenciamentos *single cell*.
- (C) que metilação de DNA, modificação de histonas, mudanças ou perdas de *imprinting* genômico são marcas genéticas que o pesquisador poderá identificar. Somente com a identificação concomitante dos três tipos de marcas genéticas, será possível afirmar que há eventos epigenômicos nos genomas estudados.
- (D) que eventos de hipometilação ou hipermetilação, se identificados nos genomas, deverão ser confirmados por reações de PCR quantitativo.
- (E) que, para a posterior busca por eventos epigenômicos, no início do processo de montagem genômica, devem ser filtradas e usadas *reads* com tamanhos maiores que 700 pares de bases (pb), quando advindas de sequenciadores geradores de *short reads*, e *reads* maiores que 2.000 pb, quando advindas de sequenciadores geradores de *long reads*.

23. Montagens de genomas bacterianos, a partir de dados de NGS (*Next Generation Sequencing*), podem ser realizadas por duas principais abordagens: *de novo* ou mapeamento por referência. Abaixo são apresentadas possíveis etapas de um *pipeline* para montagem de um genoma bacteriano, a partir de sequências provenientes de um sequenciamento em uma plataforma NGS. Sobre esse tema, assinale a afirmativa correta.

- (A) Montagem do genoma com abordagem *de novo* tem a seguinte ordem de execução no seu *pipeline*: análise de qualidade, trimagem, ordenamento, montagem, alinhamento e fechamento de *gaps*.
- (B) Montagem do genoma com abordagem *de novo* tem a seguinte ordem de execução no seu *pipeline*: análise de qualidade, trimagem, alinhamento, montagem, ordenamento e fechamento de *gaps*.
- (C) Montagem do genoma com abordagem de mapeamento por referência tem, no seu *pipeline*, após a escolha e a montagem da sequência referência, na ordem de execução: análise de qualidade, alinhamento na referência e ordenamento.
- (D) Montagem do genoma com abordagem de mapeamento por referência tem, no seu *pipeline*, após a escolha da sequência referência, na ordem de execução: análise de qualidade, trimagem, alinhamento *de novo* na referência, fechamento de *gaps* e ordenamento.
- (E) Montagem do genoma com abordagem de mapeamento por referência tem, no seu *pipeline*, após a escolha e a montagem da sequência referência, na ordem de execução: análise de qualidade, trimagem, alinhamento na referência e compactação dos grafos e visualização da montagem.

24. O diretor de uma Unidade de Pronto-Atendimento (UPA) pediu de forma cortês ajuda a um colega, responsável pelos agendamentos e encaminhamentos de exames da UPA, para um favorecimento na fila de espera para realização de um exame de imagem de um amigo, o qual está com urgência na obtenção deste para continuidade de tratamento médico. Baseando-se no Decreto nº 1.171/1994, que descreve sobre a Ética Profissional do Servidor Público Civil do Poder Executivo Federal, considere as condutas abaixo.

- I - Atender imediatamente ao pedido do diretor, pois é um dos deveres do servidor público ter respeito à hierarquia.
- II - Atender ao pedido do diretor, pois é dever do servidor público ser probo, reto, leal e justo, e o atendimento do pedido será benéfico à saúde de um cidadão.
- III- Não atender ao pedido do diretor, pois trata-se de uma vantagem indevida, e denunciá-lo.
- IV - Não atender ao pedido do diretor, pois é um dos deveres do servidor público exercer com estrita moderação as prerrogativas funcionais que lhe sejam atribuídas, abstendo-se de fazê-lo contrariamente aos legítimos interesses dos usuários do serviço público.

Quais são condutas corretas desse servidor diante do pedido do diretor?

- (A) Apenas III.
- (B) Apenas I e II.
- (C) Apenas II e III.
- (D) Apenas II e IV.
- (E) Apenas III e IV.

25. Considere um conjunto de dados de expressão gênica de 12.000 genes em 100 amostras de tecido, de indivíduos-controle e doentes. Qual das seguintes técnicas de análise de dados biológicos multidimensionais é a mais adequada para identificar os genes que estão diferencialmente expressos entre as duas condições?

- (A) Análise de componentes principais (PCA).
- (B) Análise discriminante linear (LDA).
- (C) Análise de coordenadas principais (PCO).
- (D) Regressão linear múltipla (RLM).
- (E) Regressão de ângulo mínimo (LARS).