

## ORIENTAÇÕES GERAIS AO CANDIDATO

1. Verifique se este caderno contém a quantidade correta de questões e se não há falhas de impressão. Em caso de irregularidade, comunique imediatamente ao fiscal de sala.
2. Os candidatos deverão marcar suas respostas no **cartão-resposta** de forma apropriada e entregá-lo juntamente com o caderno de provas ao término da aplicação.
3. **O cartão-resposta é o único documento válido para correção da prova.** O preenchimento é de inteira responsabilidade do candidato.
4. Preencha o cartão-resposta com **caneta esferográfica de tinta azul ou preta**, fabricada em material transparente.
5. As marcações no cartão-resposta deverão ser feitas cobrindo completamente o círculo correspondente à alternativa escolhida, sem ultrapassar seus limites.
6. **O tempo total de realização da prova será de 2 horas**, incluído o tempo destinado ao preenchimento do cartão-resposta.
7. O candidato somente poderá **deixar o local de prova após decorrida 1 hora do início da aplicação.**
8. O candidato que necessitar ausentar-se da sala para ir ao banheiro deverá solicitar autorização ao fiscal e será acompanhado, conforme normas do certame.
9. **Os 03 (três) últimos candidatos deverão permanecer em sala e somente poderão sair juntos**, após a assinatura da ata de encerramento dos trabalhos.
10. Não será permitida qualquer forma de consulta a materiais impressos, anotações ou comunicação entre candidatos, sob pena de eliminação do processo seletivo.
11. **Sobre a carteira do candidato são permitidos apenas o documento de identificação, o caderno de provas e a caneta esferográfica.** São vedados: bonés, chapéus, óculos escuros (exceto por prescrição médica devidamente comprovada), alimentos, bebidas (exceto água em garrafa transparente e sem rótulo), bolsas ou qualquer outro material não autorizado.
12. **É expressamente proibido, durante toda a realização da prova, o uso de quaisquer aparelhos eletrônicos**, tais como telefones celulares, relógios digitais ou smartwatches, fones de ouvido ou dispositivos similares. Os aparelhos deverão permanecer desligados e acondicionados conforme orientação da fiscalização. O toque ou a vibração de aparelho eletrônico durante a realização da prova, independentemente de uso intencional, implicará a eliminação imediata do candidato e lavratura de ocorrência em ata.
13. A folha de rascunho para anotar as respostas assinaladas poderá ser levada pelo candidato, quando permitido, não possuindo qualquer validade para fins de correção.
14. O descumprimento das normas estabelecidas poderá implicar eliminação do candidato, conforme regras do edital.

# Prova 1 - Categoria Bioinformática

Identificador:  
WSG412953

## Variante A | 35 perguntas (50 p.)

Atenção! Certifique-se de que recebeu uma folha de respostas com o mesmo identificador.

### Pergunta 1 | escolha única

A diferença entre cobertura horizontal e cobertura vertical é:

- A. Enquanto a horizontal representa loci (registrável em arquivo bed) cobertos pelo ensaio, a vertical representa a quantidade de reads sobre um dado locus ou média entre loci
- B. A horizontal representa loci (registrável em arquivo fastq) cobertos pelo ensaio e a vertical representa a profundidade derivada do número de reads alinhados a um locus ou média entre loci
- C. A horizontal representa a profundidade derivada do número de reads alinhados a um locus ou média entre loci e a vertical representa loci (em arquivo bam) cobertos pelo ensaio
- D. A horizontal permite um aumento da acurácia e a vertical representa a possibilidade de cobrir poucos genomas inteiros ou mais exomas em um único ensaio de corrida
- E. A horizontal é representada por X (por exemplo, quando falamos em um genoma completo 30X) enquanto a vertical é representada em megabases

### Pergunta 2 | escolha única

Em um ensaio Paired-end as reads represent:

- A. Em um mesmo cluster, ciclos de sequenciamento diferentes com a mesma fita sendo sequenciada no sentido 5'→3' na primeira etapa de ciclos e no sentido 3'→5' na segunda etapa de ciclos
- B. Em um mesmo cluster, ciclos de sequenciamento diferentes com as fitas complementares sendo sequenciadas no sentido 5'→3', sendo cada fita em uma das etapas dos ciclos
- C. Em clusters independentes, ciclos de sequenciamento diferentes com as fitas complementares sendo sequenciadas no sentido 5'→3', sendo cada fita em uma das etapas dos ciclos
- D. Em clusters independentes, ciclos de sequenciamento diferentes com a mesma fita sendo sequenciada no sentido 5'→3' na primeira etapa de ciclos e no sentido 3'→5' na segunda etapa de ciclos
- E. Tanto em um mesmo cluster quanto em clusters independentes (reunidos pelo mesmo par de índices na etapa de demultiplex), ciclos de sequenciamento diferentes com as fitas complementares sendo sequenciadas no sentido 5'→3', sendo cada fita em uma das etapas dos ciclos

### Pergunta 3 | escolha única

Demultiplexar significa:

- A. Agrupar reads de indivíduos diferentes sequenciados numa mesma corrida utilizando a informação dos adaptadores
- B. Agrupar clusters de um mesmo indivíduo sequenciados em corridas diferentes utilizando a informação dos índices
- C. Agrupar regiões de interesse do genoma quando múltiplas corridas são realizadas utilizando a informação das sondas
- D. Agrupar reads de um mesmo indivíduo sequenciados em uma mesma corrida utilizando a informação dos índices
- E. Separar clusters diferentes de indivíduos diferentes, utilizando uma combinação de adaptadores e índices

**Pergunta 4** | escolha única

A Figura 1 corresponde a:

- A. Recorte do software Dragen indicando um trecho de exoma de um paciente
- B. Recorte do software IGV indicando um trecho de genoma de dois pacientes
- C. Recorte do software IGV indicando dois trechos de exoma de um paciente
- D. Recorte do software Dragen, indicando um trecho de genoma de dois pacientes
- E. Recorte do software IGV indicando um trecho de exoma de dois pacientes

**Pergunta 5** | escolha única

Na Figura 1 é possível observar claramente uma:

- A. Duplicação que inclui exon, 5'UTR e região 5' do gene PIGL em um paciente
- B. Duplicação que inclui exon, 3'UTR e região 3' do gene PIGL em um paciente
- C. Deleção que inclui exon, 5'UTR e região 5' do gene PIGL em um paciente
- D. Deleção que inclui exon, 3'UTR e região 3' do gene PIGL em um paciente
- E. Deleção que inclui exon, 5'UTR e região 5' do gene PIGL em dois pacientes

**Pergunta 6** | escolha única

Na Figura 1, as observações que fornecem evidência para a resposta da questão 5 são:

- A. Split reads, distância de paired reads e cobertura vertical
- B. Distância de paired reads, cobertura vertical e flags de reads
- C. Split reads, cobertura vertical e flags de reads
- D. Cobertura horizontal e cobertura vertical
- E. Split reads, comprimento de reads e cobertura vertical

**Pergunta 7** | escolha única

Sobre as colunas da Figura 2: I. A-E, H-J estão presentes em qualquer VCF enquanto F e G indicam que o VCF foi anotado II. É confiável dizer que um indivíduo tem um genótipo com alelo alternativo e dois indivíduos tem três genótipos com alelos alternativos, assumindo um critério de no mínimo 3 reads por alelo III. A coluna F sugere que o método utilizado foi o sequenciamento do genoma completo e não do exoma completo

- A. I, II e III estão corretas
- B. I e III estão corretas
- C. I e II estão corretas
- D. II e III estão corretas
- E. Apenas I está correta

**Pergunta 8** | escolha única

Na Figura 2, o sexo provável das amostras é:

- A. Se FOXP3 estiver em região pseudoautosômica, há um homem, uma mulher e uma amostra sem sexo determinável
- B. Se FOXP3 estiver em região não-pseudoautosômica, há ao menos uma mulher (J)
- C. Se FOXP3 estiver em região não-pseudoautosômica, há ao menos uma mulher (I)
- D. Se FOXP3 estiver em região pseudoautosômica, J deve ser um homem
- E. Se FOXP3 estiver em região não-pseudoautosômica, J deve ser um homem

**Pergunta 9**

| escolha única

Sobre o Joint Genotyping, assinale a alternativa correta:

- A.** Representa uma etapa computacional crítica, uma vez que um arquivo populacional ocupa mais espaço em memória do que um conjunto de gVCFs individuais de mesmo tamanho amostral.
- B.** À medida que o número de indivíduos cresce e a representatividade da diversidade genética aumenta, o número de novas variantes descobertas com a adição de um novo indivíduo tende a diminuir, uma vez que a taxa de identificação de novas variantes no dataset tende a diminuir, o que pode ser aproximado com a fórmula de amostragem de Ewens.
- C.** À medida que o número de indivíduos cresce e a representatividade da diversidade genética aumenta, o número de novas variantes descobertas com a adição de um novo indivíduo tende a aumentar, uma vez que a descoberta de variantes cresce de forma exponencial, independentemente do número de indivíduos já adicionados.
- D.** Representa uma etapa computacional crítica, uma vez que a geração do arquivo populacional exige alto desempenho em processamento, mas representa um processo com consumo relativamente baixo de memória RAM, devido ao processamento sequencial dos dados.
- E.** O joint genotyping pode ser substituído pela junção dos arquivos vcf individuais, desde que as posições não chamadas para o indivíduo sejam completadas com 0/0, garantindo a homozigose do alelo de referência.

**Pergunta 10**

| escolha única

Sobre GWAS, avalie se as afirmações são verdadeiras ou falsas:

- ( ) É sempre importante avaliar as frequências alélicas dos sinais observados e compará-las com as frequências em outras populações ao redor do mundo para garantir que não é um sinal provocado por estruturação populacional.
- ( ) O uso de indivíduos aparentados, sem utilização de matriz de parentesco como covariável, pode levar à descoberta confiável de variantes associadas ao fenótipo de interesse, uma vez que incorpora a informação de segregação familiar das variantes.
- ( ) Dados sociodemográficos podem ser usados como covariáveis em diversos desenhos experimentais, uma vez que doenças complexas podem ocorrer devido a fatores ambientais, como a falta de acesso à alimentação de qualidade.
- ( ) O uso de informação de ancestralidade local como covariável para populações miscigenadas pode revelar variantes associadas a fenótipos que não puderam ser descobertas em estudos com amostras majoritariamente europeias.
- ( ) Ao avaliar sinais de associação, não é esperado que haja diferença de frequência alélica entre o grupo caso e o grupo controle, uma vez que isso provocaria um desvio no equilíbrio de Hardy-Weinberg, o que deveria ser removido no controle de qualidade do banco de dados. Assinale a sequência correta:

- A.** VVVFV
- B.** VFVWV
- C.** VFVFV
- D.** VFVVF
- E.** VFFVF

**Pergunta 11**

| escolha única

Sobre o sequenciamento por NGS de regiões do cromossomo X em humanos, escolha a alternativa correta relativa à chamada de variantes:

- A.** A chamada de variantes para o cromossomo X segue os mesmos critérios que para os autossomos, pelo que não é necessário nenhum cuidado especial.
- B.** Para a maior parte dos softwares de chamada de variantes, é suficiente indicar o sexo dos indivíduos analisados.
- C.** Além do sexo dos indivíduos, é necessário usar a informação relativa à ploidia específica das regiões pseudo-autossômicas e não pseudo-autossômicas dos cromossomos sexuais.
- D.** Além do sexo dos indivíduos, é necessário usar a informação relativa à ploidia específica das regiões pseudo-autossômicas e não pseudo-autossômicas entre o cromossomo X e os autossomos.
- E.** Em geral, a melhor solução é limitar as análises a indivíduos de sexo masculino pela sua natureza hemizigótica.

**Pergunta 12**

| escolha única

O SNV intrônico rs6683977 no gene PDE4B tem dois alelos: G (o alelo de referência em dbSNP, correspondente à fita (+) do genoma humano) e C (o alelo alternativo). Um GWAS publicado reportou que o alelo C está associado a um maior risco de recidiva em hispanos dos Estados Unidos, após tratamento da leucemia linfóide aguda, sem o cuidado de especificar que se referia à fita (-) do cromossomo 1, e não, como é convencional, à fita (+). Indique a afirmação correta:

- A.** Em relação a fita (+) o alelo de risco de recidiva é G, o alelo de referência, e o alelo alternativo é C.
- B.** Em relação a fita (+) o alelo de risco de recidiva é C, o alelo de referência, e o alelo alternativo é G.
- C.** Em relação a fita (-) o alelo de risco de recidiva é G, o alelo alternativo, e o alelo de referência é C.
- D.** O alelo C refere-se a fita complementar da fita (+), pelo que não corresponde a outro alelo.
- E.** As variantes com alelos G e C, ou A e T devem ser excluídas de todas as análises.

**Pergunta 13**

| escolha única

Para dois SNVs distantes no genoma humano aproximadamente 20kb: rs100 (com alelos C e T) e rs101 com alelos (A e G) se observam os seguintes genótipos obtidos por microarranjos para rs100 e rs101, respectivamente:

Indivíduo1 CC AA

Indivíduo2 CC AG

Indivíduo3 CT AG

Indivíduo4 CT GG

Indivíduo5 TT GG

Para que indivíduos existe incerteza em relação aos haplótipos que carrega:

- A.** Indivíduo1
- B.** Indivíduo2 e indivíduo4
- C.** Indivíduo3
- D.** Indivíduo5
- E.** Em nenhum caso.

**Pergunta 14** | escolha única

rs269868 (Ser1067Leu, g.G>A) é um SNV no gene DUOX2 para a qual a figura 3 mostra a representação para o genoma e o transcrito (em verde).

Figura 3. Variation Viewer do NCBI

Qual das seguintes opções, que representam os códons mutados é a correta para o SNV indicado acima?

- A. AGC > AAC
- B. CGA > CAA
- C. UCA > UUA
- D. UCG > UUG
- E. Não é possível definir o códon mutado com as informações disponíveis

**Pergunta 15** | escolha única

Enquanto analisava um arquivo VCF, você executou o comando:

```
> grep "chr1" YourFile.vcf.gz
```

e obteve a seguinte mensagem de erro:

```
Binary file YourFile.vcf.gz matches
```

ou saída ilegível com caracteres corrompidos.

Considerando a situação, qual é a melhor prática para manipular corretamente esse arquivo?

- A. Converter o arquivo para formato BCF, pois o VCF compactado não pode ser processado com ferramentas de linha de comando
- B. Reprocessar o arquivo, pois esse erro indica corrupção no arquivo compactado
- C. Descompactar o arquivo ou utilizar ferramentas apropriadas para arquivos .gz
- D. Atualizar a versão do tabix, já que esse erro ocorre em versões anteriores à tabix/0.2.6
- E. Nenhuma das opções anteriores.

Sobre o Manhattan plot (parte A) da Figura 4 e sobre GWAS, indique a afirmação correta:

- A.** Dado um SNV, o eixo vertical indica os valores de significância (p values) em escala de  $-\log_{10}$  para cada teste estatístico correspondente a um SNV. O eixo horizontal representa a posição física de cada SNV nos diferentes autossomos. Cada ponto representa o valor de  $-\log_{10}(p)$  do teste de associação entre os números de cópias de um dos alelos do SNV específico (variável independente) e as concentrações plasmáticas de R-metadona (variável dependente) nos indivíduos participantes no GWAS. Para os testes de associação, a hipótese nula é que existe associação entre genótipos e fenótipos.
- B.** Dado um SNV, o eixo vertical indica os valores de significância (p values) em escala de  $-\log_{10}$  para cada teste estatístico correspondente a um SNV. O eixo horizontal representa a posição física de cada SNV nos diferentes autossomos. Cada ponto representa o valor de  $-\log_{10}(p)$  do teste de associação entre os números de cópias de um dos alelos do SNV específico (variável independente) e as concentrações plasmáticas de R-metadona (variável dependente) nos indivíduos participantes no GWAS. Para os testes de associação, a hipótese nula é a ausência de associação entre genótipos e fenótipos
- C.** Dado um SNV, o eixo vertical indica os valores de significância (p values) em escala de  $-\log_{10}$  para cada teste estatístico correspondente a um SNV. O eixo horizontal representa a posição física de cada SNV nos diferentes autossomos. Cada ponto representa o valor de  $-\log_{10}(p)$  do teste de associação entre os números de cópias de um dos alelos do SNV específico (variável dependente) e as concentrações plasmáticas de R-metadona (variável independente) nos indivíduos participantes no GWAS. Para os testes de associação, a hipótese nula é a ausência de associação entre genótipos e fenótipos.
- D.** Dado um SNV, o eixo vertical indica os valores de significância (p values) em escala de  $-\log_{10}$  para cada teste estatístico correspondente a um SNV. O eixo horizontal representa a posição física de cada SNV nos diferentes autossomos. Cada ponto representa o valor de  $-\log_{10}(p)$  do teste de associação entre os números de cópias de um dos alelos do SNV específico (variável dependente) e as concentrações plasmáticas de R-metadona (variável independente) nos indivíduos participantes no GWAS. Para os testes de associação, a hipótese nula é que existe associação entre genótipos e fenótipos.
- E.** Dado um SNV, o eixo vertical indica os valores de significância (p values) em escala de  $-\log_{10}$  para cada teste estatístico correspondente a um SNV. O eixo horizontal representa a posição física de cada SNV nos diferentes autossomos. Cada ponto representa o valor de  $-\log_{10}(p)$  do teste de associação entre os números de cópias de um dos alelos do SNV específico (variável independente) e as concentrações plasmáticas de R-metadona (variável dependente) nos indivíduos participantes no GWAS. Para os testes de associação, a hipótese nula a ser testada é o equilíbrio de Hardy-Weinberg.

**Pergunta 17** | escolha única

Em relação à figura 4, indique a afirmação correta.

- A.** A significância (p-valor) se define como a probabilidade de observar um resultado igual ao observado ou outro ainda mais diferente quando comparado com os valores esperados sob a hipótese nula, assumindo que a hipótese nula é falsa. Nos GWAS, se usa como limiar de significância um p-valor muito baixo, normalmente da ordem de  $10^{-8}$ , para evitar falsos negativos produto da realização de um número muito elevado de testes.
- B.** A significância (p-valor) se define como a probabilidade de observar um resultado igual ao observado ou outro ainda mais diferente quando comparado com os valores esperados sob a hipótese nula, assumindo que a hipótese nula é falsa. Nos GWAS, se usa como limiar de significância um p-valor muito baixo, normalmente da ordem de  $10^{-8}$ , para evitar falsos positivos produto da realização de um número muito elevado de testes.
- C.** A significância (p-valor) se define como a probabilidade de observar um resultado igual ao observado ou outro ainda mais diferente quando comparado com os valores esperados sob a hipótese nula, assumindo que a hipótese nula é verdadeira. Nos GWAS, se usa como limiar de significância um p-valor muito baixo, normalmente da ordem de  $10^{-8}$ , para evitar falsos negativos produto da realização de um número muito elevado de testes.
- D.** A significância (p-valor) se define como a probabilidade de observar um resultado igual ao observado, ou outro ainda mais diferente quando comparado com os valores esperados sob a hipótese nula, assumindo que a hipótese nula é verdadeira. Nos GWAS, se usa como limiar de significância um p-valor muito baixo, normalmente da ordem de  $10^{-8}$ , para evitar falsos positivos produto da realização de um número muito elevado de testes.
- E.** A significância (p-valor) se define como a probabilidade de observar um resultado igual ao observado, ou outro mais similar quando comparado com os valores esperados sob a hipótese nula, assumindo que a hipótese nula é verdadeira. Nos GWAS, se usa como limiar de significância um p-valor muito baixo, normalmente da ordem de  $10^{-8}$ , para evitar falsos positivos produto da realização de um número muito elevado de testes.

**Pergunta 18** | escolha única

Indique a afirmação incorreta em relação ao resultado do GWAS da figura 4.

- A.** Foi encontrado um SNV significativamente associado com o desfecho (rs17180299), sem outras variantes em linkage disequilibrium com rs17180299.
- B.** Foi encontrado um SNV significativamente associado com o desfecho (rs17180299). Algumas variantes próximas no genoma estão em linkage disequilibrium com rs17180299 e mostram algum nível de associação com o desfecho.
- C.** rs17180299 tem dois alelos, A e G, sendo que o alelo G está associado positivamente com o desfecho.
- D.** As frequências alélicas de A e G são 0,91 e 0,09, respectivamente.
- E.** As frequências genotípicas de AA, AG e GG são 0,83, 0,16 e 0,01, respectivamente.

**Pergunta 19** | escolha única

Sobre vocabulários fenotípicos padronizados, assinale a alternativa correta:

- A.** Termos HPO (Human Phenotype Ontology) são destinados a descrever anomalias fenotípicas, portanto possuem uma boa sobreposição com o CID (Classificação Internacional de Doenças)
- B.** O OMOP (Observational Medical Outcomes Partnership) suporta um conjunto de vocabulários mais abrangente do que o CID (Classificação Internacional de Doenças), uma vez que também padroniza fenótipos não clínicos.
- C.** O uso de vocabulários padronizados garante que as métricas de completude, confiabilidade e coerência sejam cumpridas nos dados fenotípicos.
- D.** O OMOP (Observational Medical Outcomes Partnership) é considerado um modelo de vocabulário padronizado menos assertivo do que o CID (Classificação Internacional de Doenças), uma vez que padroniza fenótipos não clínicos, enquanto o CID exige um diagnóstico.
- E.** O uso de vocabulários padronizados não influencia a possibilidade de replicabilidade dos resultados de um estudo.

**Pergunta 20** | escolha única

Assinale a alternativa correta sobre o pseudocódigo da Figura 5

- A. Quando o vetor atingir 100 itens, o código finaliza a execução.
- B. O código faz com que o vetor “sequencia” seja preenchido de acordo com o que é digitado pelo usuário. O vetor chegará a 100000000000 itens.
- C. Se o usuário digitar um stop-codon antes da 100ª posição do vetor, o código formará um loop infinito.
- D. O código faz com que o vetor “sequencia” seja preenchido de acordo com o que é digitado pelo usuário. O número de iterações que o código fará para isso é 100.
- E. A condição “ATÉ  $i > 100$ ” faz com que o código finalize mesmo que o usuário nunca digite um stop-codon.

**Pergunta 21** | escolha única

Sobre a complexidade do código da Figura 5, assinale a alternativa correta:

- A. A complexidade do código é linear, pois o vetor é percorrido uma única vez dentro do loop.
- B. A complexidade do código é logarítmica, pois a condicional reduz a quantidade de vezes que o código executa uma ação.
- C. A complexidade do código muda de acordo com a quantidade de códons inserida pelo usuário, sendo limitada ao tamanho do vetor.
- D. A complexidade do código é indeterminada, uma vez que depende do input do usuário.
- E. A complexidade do código no pior caso é constante, sendo definida por  $O(1)$ .

**Pergunta 22** | escolha única

Dado uma única Flow cell, o que permite a inclusão de múltiplos indivíduos ou amostras em uma única corrida de sequenciamento de nova geração é:

- A. Índices ligados aos fragmentos de cada indivíduo e concentração de clusters para aumento de cobertura vertical
- B. Sondas hibridizando aos fragmentos por indivíduo e densidade de clusters para aumento da cobertura vertical
- C. Índices ligados aos fragmentos de cada indivíduo e captura por hibridização de sondas para enriquecimento de loci de interesse com aumento da cobertura vertical e redução da horizontal
- D. Adaptadores ligados aos fragmentos de cada indivíduo e captura por hibridização com fragmentos para enriquecimento de loci de interesse com aumento da cobertura horizontal e redução da horizontal
- E. Índices ligados aos fragmentos com loci de interesse para redução da cobertura vertical e sondas hibridizando aos fragmentos por indivíduo

**Pergunta 23** | escolha única

No contexto de sequenciamento de nova geração, o substrato com sondas utilizado para sequenciar milhões de fragmentos em paralelo é:

- A. Limitado, nas flow cells, pela densidade de clusters bem sucedidos e pela capacidade da flow cell (output)
- B. Limitado, nas flow cells, pelo tamanho dos clusters bem sucedidos e pela capacidade da flow cell (throughput)
- C. Limitado, nas flow cells, pelo concentração da biblioteca e pela capacidade da flow cell (cobertura)
- D. Limitado, antes da corrida, pelo concentração da biblioteca e pela capacidade da flow cell (output)
- E. Ilimitado, dependendo apenas da densidade de sondas e da capacidade da flow cell (cobertura)

**Pergunta 24** | escolha única

Em relação a Figura 6, considerando que a doença diagnosticada clinicamente é autossômica recessiva com heterogeneidade alélica, qual seria o resultado esperado em relação a zigosidade do paciente? Sinalize a resposta correta.

- A. O paciente tem que apresentar a mesma mutação em homozigose.
- B. É esperado que o paciente apresente duas variantes patogênicas em heterozigose no mesmo gene.
- C. É esperado que o paciente apresente duas variantes patogênicas em heterozigose em genes diferentes.
- D. É esperado que o paciente apresente duas variantes patogênicas no mesmo cromossomo.
- E. Todas as opções a-e são erradas.

**Pergunta 25** | escolha única

A variante reportada no laudo da Figura 6, Ile396Lys, do ponto de vista da anotação funcional, é classificada como:

- A. de perda de função
- B. de sentido trocado (missense)
- C. sinônima
- D. frameshift ou de mudança de fase de leitura
- E. stop-codon

**Pergunta 26** | escolha única

A variante reportada na Figura 6 é produto de uma mutação de T>A na posição 1187 do transcrito. Indique a resposta correta:

- a) O códon mutado é AUU
  - b) O códon mutado é AUC
  - c) O códon mutado é AUA
  - d) O códon mutado é AUG
  - e) As opções a, b e c estão erradas.
- A. O códon mutado é AUU
  - B. O códon mutado é AUC
  - C. O códon mutado é AUA
  - D. O códon mutado é AUG
  - E. As opções a, b e c estão erradas.

**Pergunta 27** | escolha única

A partir desse laudo, indique a afirmação incorreta em relação à classificação da variante reportada de acordo com os critérios da ACMG e as implicações clínicas.

- A. Dado que a variante envolve uma mudança de aminoácido na proteína codificada, podemos concluir que a variante patogênica foi encontrada no paciente.
- B. Embora a variante encontrada envolva uma mudança de aminoácido num gene envolvido no metabolismo de carboidratos, não existe informação suficiente para inferir que essa variante seja de natureza patogênica, nem para afirmar que é benigna, pelo que a variante foi classificada como de significado incerto. Se recomenda revisar a classificação da variante em meses ou anos, porque podem aparecer novas evidências que levem a mudar a classificação da variante.
- C. A ausência de outras variantes candidatas a patogênicas no painel de 30 genes estudado, sugere que pode ser necessário a análise de outros genes, pelo que se recomenda uma análise de todo o exoma do paciente.
- D. rs866682996 corresponde à nomenclatura mais usada na variante no banco de dados dbSNP.
- E. No caso do banco de dados gnomAD, o nome é um acrônimo para Genome Aggregated Database, o que significa que o banco não contém informação genômica de indivíduos, mas dados agregados.

**Pergunta 28** | escolha única

Use o código genético (Figura 7) para completar o quadro que segue. Na figura as colunas representam alinhamentos transcricionais e traducionais. Na primeira linha, a extremidade 5' está à esquerda e a 3' à direita. Indique o polipeptídeo codificado

C Dupla hélice de DNA

T G A

C A U mRNA transcrito

G C A Anticódon no tRNA

Trp Aminoácidos

incorporados à proteína

- A. NH<sub>3</sub> - Cys - Ser - Trp - Thr - COO-
- B. NH<sub>3</sub> - Thr - Trp - Ser - Cys - COO-
- C. NH<sub>3</sub> - Ala - Trp - COO-
- D. NH<sub>3</sub> - Thr - Ser - Thr - Cys - COO-
- E. Nenhuma das opções anteriores.

**Pergunta 29** | escolha única

Fenótipos monogênicos (especificamente de doenças) são aqueles onde

- A. Alterações patogênicas em apenas um gene são necessárias para determinar a ocorrência do fenótipo (quem tem a alteração certamente terá o fenótipo)
- B. Alterações deletérias em apenas um gene são suficientes para determinar a ocorrência do fenótipo (quem tem a alteração certamente terá o fenótipo)
- C. Alterações clinicamente relevantes em apenas um gene são suficientes para aumentar significativamente o risco de ocorrência do fenótipo (quem tem a alteração provavelmente terá o fenótipo)
- D. Alterações patológicas em um gene são causadoras definitivas do fenótipo
- E. Alterações tipo VUS são importantes para influenciar o fenótipo

**Pergunta 30** | escolha única

Penetrância, no âmbito de doenças monogênicas e multifatoriais, pode ser definida como

- A. o tamanho do efeito de uma variante genética, medido como proporção de afetados entre um conjunto de pessoas numa família com a variante ou proporções de pessoas com e sem a variante em conjuntos caso e controle
- B. a proporção de pessoas com a variante entre o total de afetados numa família
- C. a proporção de afetados numa população com uma mesma variante candidata
- D. o efeito deletério de um gene associado a uma doença
- E. o efeito deletério de uma variante genética sobre o fenótipo

**Pergunta 31** | escolha única

A frequência esperada de um genótipo com variante classificada como patogênica, com número de alelos compatível com o padrão de herança e penetrância completa, é sempre

- A. tão ou menos rara que a própria prevalência da doença
- B. proporcional à prevalência da doença
- C. menos rara que a própria prevalência da doença
- D. tão ou mais rara que a própria prevalência da doença
- E. mais comum que a própria prevalência da doença, desde que tenha penetrância completa

**Pergunta 32** | escolha única

A Figura 8 apresenta projeções bidimensionais (PC1 vs PC2 e PC3 vs PC4) inferidas a partir de dados genéticos obtidos por array de genotipagem. O método estatístico utilizado para gerar esta representação visual é a Análise de Componentes Principais (PCA). É correto afirmar que o PCA neste contexto resulta na redução da dimensionalidade:

- A. da informação dos marcadores genéticos genotipados no banco de dados.
- B. da informação do número de indivíduos amostrados no banco de dados.
- C. da frequência de cada alelo agregada por população do banco de dados.
- D. do FST de cada posição genômica medido entre os grupos.
- E. do coeficiente de parentesco (kinship) calculado entre cada par de indivíduos.

**Pergunta 33** | escolha única

Ainda com base na Figura 8, avalie as afirmações:

- () A separação gráfica entre os Nativos Peruanos e os Ibéricos no PC1 afirma a não ocorrência de eventos de miscigenação entre as duas populações.
- () As variantes genéticas que possuem maior influência na inferência do PC1 são aquelas mais diferenciadas entre as populações Nativas Peruanas e a população Ibérica.
- () A maior dispersão dos indivíduos Ibéricos nos eixos PC3 e PC4 sugere maior variabilidade genética intrapopulacional capturada por esses Componentes Principais em comparação aos indivíduos Nativos Peruanos.
- () O PC2 revela uma separação em gradiente entre os indivíduos nativos peruanos, com os indivíduos dos Andes Áridos tendo valores de PC2 mais próximos a zero que os indivíduos da costa. Assinale a sequência correta:

- A. VVVF
- B. FFFV
- C. FFVV
- D. FVVF
- E. FVFE

**Pergunta 34** | escolha única

De acordo com a Lei Geral de Proteção de Dados (Lei Nº 13.709, de 14 de agosto de 2018), que visa proteger e garantir a segurança de dados pessoais, o dado anonimizado é definido como:

- A. dado relativo a titular que não possa ser identificado, considerando a utilização a limitação de indivíduos que possam identificar o titular;
- B. dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento;
- C. dado relativo a titular que não possa ser identificado, considerando a utilização de padronização no chaveamento na reidentificação;
- D. dado relativo a titular que não possa ser identificado, considerando o uso de ferramenta específica aprovada pela Lei Nº 13.709, de 14 de agosto de 2018;
- E. dado relativo a titular que não possa ser identificado, considerando a autorização do mesmo para reidentificação;

**Pergunta 35**

| escolha única

Um aluno de doutorado trabalhando com dados sensíveis de fenótipos de uma coorte de idosos teve seu computador pessoal hackeado, o que levou ao vazamento desses dados. Considerando que a comunicação do incidente foi realizada em período razoável determinado pela autoridade nacional, é correto dizer que no mínimo ele deve ter mencionado na comunicação:

- A.** a descrição da natureza dos dados pessoais afetados; as informações sobre os titulares envolvidos; o relato completo do incidente e as condições em que o mesmo aconteceu; os riscos relacionados ao incidente; as medidas que foram ou serão tomadas para reverter ou mitigar o prejuízo.
- B.** a descrição da natureza dos dados pessoais afetados; as informações sobre os titulares envolvidos; quem teve acesso aos dados e a amplitude do vazamento; os riscos relacionados ao incidente; as medidas que foram ou serão tomadas para reverter ou mitigar o prejuízo.
- C.** a descrição da natureza dos dados pessoais afetados; as informações sobre os titulares envolvidos; a indicação das medidas técnicas e de segurança utilizadas para a proteção dos dados; os riscos relacionados ao incidente; as medidas que foram ou serão tomadas para reverter ou mitigar o prejuízo.
- D.** a descrição da natureza dos dados pessoais afetados; as informações sobre os titulares envolvidos; a indicação das medidas técnicas e de segurança utilizadas para a proteção dos dados; o relato completo do incidente e as condições em que o mesmo aconteceu; as medidas que foram ou serão tomadas para reverter ou mitigar o prejuízo.
- E.** a descrição da natureza dos dados pessoais afetados; as informações sobre os titulares envolvidos; a indicação das medidas técnicas e de segurança utilizadas para a proteção dos dados; o relato completo do incidente e as condições em que o mesmo aconteceu; quem teve acesso aos dados e a amplitude do vazamento; as medidas que foram ou serão tomadas para reverter ou mitigar o prejuízo.