



FIOCRUZ

Concurso Público Fiocruz 2023

Tecnologista em Saúde Pública

Prova Discursiva

TE42

Vigilância Molecular e Sequenciamento de Nova Geração com foco em Genômica e Transcriptômica

Espelho de Resposta

Pontuação da Questão Discursiva conforme Anexo II do Edital nº 2, de acordo com a Unidade detentora da vaga.

Espera-se que o candidato, no desenvolvimento do tema, tenha feito considerações técnicas adequadas sobre os seguintes pontos:

Primeira Geração: Método Sanger, síntese por PCR, maior precisão, maior custo por base,

Segunda Geração: leituras curtas (short reads), multiplexado ou em paralelo, ganho de velocidade de processamento; sem necessidade de eletroforese para detecção, custo por base menor do que o Sanger, requer etapa de amplificação (PCR)

Terceira Geração: leitura longa, tempo de resposta mais rápido, não necessitam da PCR, baixo custo por base, maior taxa de erro.

As técnicas de sequenciamento de DNA têm evoluído com o passar dos anos. O sequenciamento Sanger é denominado como sendo de primeira geração. Utiliza a PCR como metodologia básica. Envolve a utilização de didesoxinucleotídeos de terminação de cadeia marcados radioativamente ou com fluoróforos para realizar a síntese de uma fita de DNA complementar à fita modelo interrogada. Os fragmentos são então separados por tamanho e analisados por eletroforese em gel ou capilar para determinar a sequência.

A técnica de Sanger possibilita um sequenciamento preciso, com baixo erro, mas com um custo por base analisada elevado. Dado que apenas uma reação de sequenciamento pode ser analisada por vez, com isso o método tem rendimento limitado, mas continua sendo utilizado para sequências pequenas e mesmo para validar resultados de outros métodos.

Já no sequenciamento de segunda geração (*short-read*), com o uso de diversas plataformas (Roche/454, Ion torrent, Illumina/Solexa, ABI/SOLiD), o sequenciamento é clonal, multiplexado (em paralelo) e de alto rendimento. Assim, simultaneamente é analisado um gene, vários genes ou o genoma inteiro.

As abordagens geralmente iniciam com fragmentação de DNA, formação de biblioteca, ligação de adaptador, fixação de superfície e amplificação.

Envolve o sequenciamento massivo de *reads* de 250–800 bp, em que milhões de reações de sequenciamento individuais ocorrem em paralelo. Não há necessidade da fase de separação de fragmentos levando a um ganho de velocidade de processamento. A multiplexação de genes e amostras reduz drasticamente o custo por amostra.

Essa tecnologia pode resolver a maioria dos problemas de fase encontrados no sequenciamento Sanger, no entanto, apresenta um nível de erro maior que o método de Sanger.

O sequenciamento NGS de leitura curta domina o mercado atual de sequenciamento devido a sua relação custo/benefício. Propiciou o acesso para o diagnóstico de doenças genéticas nas várias modalidades oferecidas, que vai desde um NGS específico de um gene, até o NGS do genoma completo.

Surgiram então, as técnicas de sequenciamento de terceira geração (long-read) com plataformas PacBio e ONT, que ganharam mais atenção devido a comprimentos de reads mais longos (> 10 kb) em análises em tempo real, oferecendo assim um tempo de resposta mais rápido do que os métodos de segunda geração.

Essas plataformas não necessitam da PCR, eliminando por exemplo a dificuldade de amplificação de regiões ricas em AT e GC. Com isso, acabam por melhorar as análises do DNA repetitivo. Com a característica de *reads* mais longas, é possível resolver outro problema que o sequenciamento de *reads* curtas falha, o faseamento de genes polimórficos.

Em comparação com o sequenciamento de terceira geração, as limitações do sequenciamento de segunda geração incluem tempos de execução mais longos e dificuldades com montagem de novo, faseamento de haplótipos e identificação de isoformas de transcrição e variantes estruturais.

Uma desvantagem do sequenciamento de leitura longa é que muitas ferramentas computacionais são projetadas e dedicadas à mineração de dados de leitura curta. Um fato inegável é que as ferramentas de bioinformática se tornaram fundamentais para qualquer sequenciamento.

O sequenciamento de leitura curta também tem a vantagem de ser mais preciso do que o sequenciamento de leitura longa. O nível de erro do sequenciamento de terceira geração é alto (~14%). As empresas fabricantes dessas plataformas têm tentado diminuir essa taxa por várias estratégias e com muito investimento. É apenas uma questão de tempo para se solucionar diversos problemas e diminuir essa taxa de erro. De qualquer forma, o Método de Sanger ainda continua sendo o que tem a menor taxa de erro.

De maneira geral, o desenvolvimento das gerações de sequenciamento, tem ocorrido rapidamente, com diminuição dos custos por base sequenciada, mas, infelizmente com uma taxa de erro significativamente maior. Vale lembrar também, que os custos de equipamentos continuam altos nas várias gerações.