

RP 001/0009



CODIGO DA PROVA:

RP 001/0009



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS
CONCURSO:

FOLHA DE RESPOSTA

Importante: O código da prova só será colocado na entrega da prova ao fiscal. As provas serão escaneadas e enviadas aos membros da banca avaliadora sem o nome do candidato.

Ponto 02. Controle da expressão gênica com ênfase em métodos computacionais.

A expressão gênica é um processo biológico no qual uma sequência nucleotídica que codifica um gene é copiada em uma molécula de RNA. Esta, então, é traduzida em aminoácidos que compõem uma proteína - produto final da expressão. Assim como outros processos celulares, a expressão gênica passa por diversas etapas de controle, as quais possuem a função de modificar ou silenciar a síntese da proteína.

Modificações na expressão gênica podem estar ligadas a processos celulares importantes durante todo o ciclo de vida de um organismo. Por exemplo, durante a embriogênese, genes específicos devem ser ativados e silenciados de forma orquestrada para o correto desenvolvimento do embrião; ou, ainda, a formação de células convergentes em indivíduos adultos. Essas modificações incluem:

a) Modificações pré-transcricionais: incluem mudanças epigenéticas, as quais impedem o acesso da maquinaria de transcrição à sequência nucleotídica. Ferramentas computacionais incluem a análise de dados de sequenciamento de imunoprecipitação de cromatina (ChIP-seq) e de ligação de bisulfita em citosinas não-metiladas (BS-seq), para reconhecer sítios de ligação proteica ativos (como fatores de transcrição) e identificação de regiões epigenéticas, respectivamente. Ambos requerem o emparelhamento de sequências provenientes de sequenciamento de alto desempenho contra um genoma referência (podem também ser realizados com programas como "BWA-MEM") - No caso do ChIP-seq, são analisados sítios no sequenciamento que indicam, por exemplo, a presença de



Comissão Organizadora do Concurso
Gabinete da Direção
ICB - UFRJ

histonas. Os dados de BS-seq, por outro lado, não mapeiam contra regiões heterocromáticas (não-condensadas), indicando as regiões de interesse por seu potencial para expressão.

b) Modificações pós-transcricionais: mudanças que ocorrem após a síntese do RNA mensageiro (mRNA), como o splicing alternativo (formação de isoformas pela junção de íons distintos). Essas modificações não podem ser analisadas diretamente através do sequenciamento de transcritos (RNA-seq), onde todos os mRNAs presentes em uma amostra são sequenciados. Com o uso de ferramentas como "StringTie", é possível reconstruir as diferentes isoformas produzidas por um gene em um determinado tratamento. Com PoDe-se, por exemplo, associa a presença de determinadas isoformas com o desenvolvimento de uma patologia específica.

c) Modificação pré-translacional: incluem a ligação de miRNAs (miRNA) em mRNAs, para que esse molécula não seja traduzida, ocasionando na diminuição do gene. Essas modificações não podem ser caracterizadas também via dados de RNA-seq. As leituras são alinhadas contra um genoma de referência, a partir do qual são realizadas contagens na profundidade do sequenciamento de um mRNA. Assim, através de análises estatísticas com auxílio de pacotes como "DESeq2" e "edgeR", pode-se analisar a expressão diferencial de genes de interesse em diversos tratamentos; assim como no caso anterior, é possível discernir genes negativamente ou positivamente regulados em diferentes estágios de uma patologia, se comparar indivíduos saudáveis versus indivíduos afetados.

d) Modificações pós-translacionais: incluem mudanças na estrutura da proteína sintetizada, como a ubiquitinação. Podem ser estudadas através da modelagem estrutural de proteínas a partir da sequência de aminoácidos, dados de cristalografia e simulação de dinâmica de proteínas em alvos.

Outros estudos incluem a reconstrução de redes de co-expressão (onde grupos de genes influenciam a expressão de outros) e redes de co-regulação (grupos de genes regulam a expressão de outros). Essas análises podem ser realizadas com o programa "dePTools", integrando os dados de RNA-seq com CHIP-seq.

Ponto 07: Bioinformática aplicada à análise epigenômica

A epigenômica inclui o mapeamento, em nível de genoma completo, as mudanças no padrão de acessibilidade da maquinaria de transcrição à sequência nucleotídica presente no genoma. Assim, modificações em histonas, metilação e compactação da cromatina acarretam mudanças no perfil transcripcional de determinada célula sem modificar, contudo, a sequência nucleotídica em si. Além disso, essas modificações não herdam-se junto às informações genômicas.

O fluxo de trabalho ~~inicia~~ inicia com o sequenciamento de alto rendimento através de plataformas como "Illumina" - next-gen, em leituras curtas e paralelas. Há três principais métodos de sequenciamento: CHIP-seq, no qual é realizada imunoprecipitação da cromatina e sites de ligação de proteínas não identificadas através de picos em diferentes frequências luminosas; BS-seq, onde é utilizado bisulfite para converter citosinas metiladas em uracila e, assim, mapear regiões heterometiladas ao nível nucleotídico; e ATAC-seq, onde é utilizada a transposase Tn5 para mapear regiões cromáticas.

A partir do CHIP-seq, é possível detectar mudanças em fatores de transcrição, como a presença de modificações em histonas. ~~Órtas acarretam~~ acarretam no silenciamento ou na super-expressão de genes, ao possibilitar ou impedir fatores de compactação e a transcrição gênica. Além disso, esses dados permitem a reconstrução de redes de regulação - ~~possibilitando~~ possibilitando entender diferentes perfis de regulação gênica.

O impacto de BS-seq facilita a detecção de regiões heterometiladas, como seqs que sinalizam citosinas metiladas. A análise deve iniciar com a preparação do genoma de referência, no qual as leituras devem ser convertidas para uracila. As leituras são, então, mapeadas contra o genoma; aquelas que alinharem com o genoma metilado do organismo tanto a região heterometilada quanto quasi citosinas estão metiladas na amostra sequenciada.

O método de ATAC-seq, por sua vez, utiliza a transposase Tn5 como uma máquina de corte/insertão ~~em~~ em regiões cromáticas. Ou seja, uma seq que as regiões heterometiladas não estão disponíveis para a inserção da transposase, esse método acaba por restringir o sequenciamento às regiões com potencial de transcrição/transcrição.

Combinados, uns mitoses possibilitam marcar modificações epigenéticas ao nível de nucleotídeos em diferentes condições ou tratamentos. Por exemplo, é possível associar modificações epigenéticas à doença ~~ou~~ ~~condição~~ ou patologias, e em diferentes graus ou estágios.

Ponto 09 - Aplicação de bioinformática no estudo de patologias humanas.

A bioinformática permite o estudo de patologias humanas desde o nível genômico, transcritômico e proteômico até o modelagem estrutural de proteínas e a simulação de ligações destas com seus alvos.

Com análises genômicas, é possível mapear variantes genéticas associadas a determinada condição ou patologia; por exemplo, inserções ou deleções nucleotídicas e ~~variações~~ rearranjos cromossômicos. Aqui, leituras provenientes de sequenciamento de alto rendimento são alinhadas contra um genoma de referência. No caso de mutações e inserções/deleções, deve-se priorizar o uso de plataformas de sequenciamento em leituras curtas, pois, de uma forma geral, estas possuem qualidade e fidelidade superior às longas. No caso destas, seu uso ~~está~~ é recomendado para o estudo de rearranjos cromossômicos grandes, pois, no caso de uma inserção, por exemplo, suas leituras ~~podem~~ completamente ~~perder~~ pelos pontos de quebra. Além, a identificação desse tipo de variante genética se torna mais preciso. Ferramentas como BWA-MEM e Bowtie2 não são utilizadas para mapear esses dados, os quais são posteriormente ~~o~~ analisados com o pacote GATK.

Patologias humanas também podem ser estudadas ao nível transcritômico. Dados de RNA-seq podem revelar genes diferencialmente expressos em indivíduos ~~com~~ saudáveis e em pacientes afetados. Além, as leituras são mapeadas contra um genoma de referência utilizando programas específicos para lidar com o natureza repetitiva desses dados e com a possibilidade de ocorrência de "splicing" - como os programas "STAR" e "TopHat". A contagem de leituras associadas a determinados transcritos é feita ~~por~~ também por programas específicos (como o "featureCounts"); e, então, as análises estatísticas para estimar a expressão diferencial são realizadas por pacotes ~~de~~ linguagem R, como "DESeq2" e "edgeR". Além disso, é possível associar mudanças epigenéticas às diferentes patologias. Por exemplo, ao estudar as modificações epigenéticas durante a neurogênese, pode-se encontrar marcadores para doenças como Alzheimer e compreender o desenvolvimento do autismo.

As diferentes conformações que uma proteína pode apresentar podem ser estimadas computacionalmente. A modelagem estrutural de proteínas pode auxiliar no entendimento de patologias que envolvem

envolvem a produção de proteínas defuncionais, por exemplo. Programas como o "Rosetta" modelam proteínas a partir de uma sequência de aminoácidos conhecida. ~~Dessa~~ Diferenciar as diferentes formas que uma proteína pode assumir auxilia no desenvolvimento de fármacos específicos; um exemplo são fármacos que impedem a entrada de vírus em células saudáveis. Nesse caso, além da modelagem estrutural, também é realizada a modelagem da ligação entre a proteína e o alvo - contacto físico conhecido como "docking".