

CODIGO DA PROVA: RP 001 / 0017UNIVERSIDADE 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.

TEMA 02: Controle da expressão gênica com ênfase em métodos computacionais

O controle da expressão é o processo fundamental para determinar e ajustar a produção de proteínas, em especial nos diferentes estágios do desenvolvimento ~~celular~~ celular.

De forma simples, a expressão gênica é o processo que a troca de informações ocorre através de diferentes moléculas tais como: DNA, RNA e aminoácidos. O controle destas trocas segue o fluxo específico e ordenado de acordo com o estágio de desenvolvimento do organismo do tecido ou da condição do ambiente. No manuscrito o foco do controle da expressão gênica será em organismos eucariotas.

Há três etapas fundamentais para o controle da expressão gênica: Transcrição, pós-transcrição e Tradução. Na transcrição ocorre o processo de transformação da informação genética (genes) armazenada em moléculas de DNA para moléculas de RNA. Nesse momento diferentes moléculas de RNA são geradas desde de moléculas codificadoras de proteínas (RNA mensageiro - mRNA) até moléculas não codificadoras de proteínas como microRNA, longos RNA não codificantes, ~~outros~~.

Apesar da importância do RNA não codificante recentemente o número mais elevado destas moléculas

01/02



na célula do organismo a sua função e conhecimento
relação com a regulação gênica ainda está evoluindo.
Em contrapartida o mRNA é menos abundante
e tem sido o foco de diferentes estudos associados à
transcriptômica. Essa importância é justificada pois
estas moléculas são o molde para produção de proteínas
do organismo. Apesar de nem todo mRNA gerar
uma sequência de aminoácidos.

A molécula de mRNA é instável e pode ser degra-
dada rapidamente. Por isso é fundamental que essa
molécula esteja protegida até que cumpra o seu
papel seja este a produção de proteínas ou degra-
dação caso não seja mais necessário a sua manu-
tenção.

Após as etapas de Transcrição e pós-Transcrição
a molécula de mRNA pode servir de molde para
produção de uma sequência de aminoácidos que será
a base para produção de proteínas. As proteínas
são moléculas fundamentais para diferentes processos
bioquímicos, de desenvolvimento celular e estrutural
do organismo.

De fato estes processos são dinâmicos e podem ocorrer
simultaneamente, por isso o processo de regulação ocorre
em diferentes etapas de processamento da expressão gênica.

Em resumo, a transcrição é regulada principalmente
através da interação proteína-DNA e regiões específicas
da fita DNA como por exemplo na região promotora.
Essa região fica upstream da sequência de DNA do gene e
sua principal função é reatrar a RNA polimerase
para iniciar a transcrição do gene para mRNA. A região
promotora assim como outras regiões regulatórias (sequên-
cias de DNA que interagem com proteínas) podem sofrer modifi-
cação na sua capacidade de ligar com a proteína
alvo. Esse mecanismo pode ser positivo quando estimula
a expressão do genes ou negativa - quando reprimem



a expressão do gene.

Superada a transcrição a regulação gênica pode ser realizada através de outros mecanismos como a ligação de CAP 5 na extremidade 5' da molécula de RNA e sua complementação com a cauda poliA nos sítios de degradação. Esses mecanismos são fundamentais a estabilidade do mRNA e evitar a degradação. No entanto o efeito do controle da expressão gênica está no processamento do mRNA primário em diferentes moléculas de mRNA através do procedimento chamado splicing. Ou seja, transformações do mRNA em diferentes mRNA com a retirada de íntrons e reorganização de éxons.

Após transcrição e o processamento do mRNA, a molécula estará "madura" e apta para ser traduzida em sequência de aminoácidos. No momento da tradução o controle da expressão gênica pode ser efetuada através dos sítios de tradução que podem impedir ou estimular a tradução da sequência de aminoácidos.

Esse conjunto de controle é necessário para determinar qual gene será expresso e qual produto será gerado.

A biologia sintética tem apresentado maior destaque nestes últimos anos devido ao seu potencial de gerar modelos artificiais ajustados a uma demanda específica. Por exemplo biosensor de poluentes ambientais no meio ambiente.

Com a biologia sintética e métodos computacionais mais eficientes podemos desenvolver modelos biológicos similares aos circuitos lógicos eletrônicos. Nestes circuitos o controle da expressão gênica é realizado por processos matemáticos, em especial álgebra de Boole.

No geral a regulação gênica nestes circuitos



é realizado na etapa de transcrição gênica através da manipulação da região promotora. Nestas construções podemos gerar sequências de DNA sintética que sejam responsáveis em ativar quando determinado estímulo (proteína/molécula) estiver no meio. Com a sistematização booleana podemos incluir processos mais refinados para o controle de expressão gênica incluindo portas lógicas como "AND" fazendo que a expressão ocorra somente quando ~~do~~ quando dois produtos específicos estiverem no meio.

Por fim, a modelagem por aprendizado de máquinas e a inteligência artificial podem contribuir com o desenho de circuitos biológicos mais robustos e eficientes. O controle de expressão gênica ainda está em desenvolvimento, em especial no Brasil. No entanto, o desenvolvimento das tecnologias de sequenciamento e os avanços na área computacional são possíveis - uma futura próxima - construir circuitos biológicos mais sofisticados para atuarem nas diferentes etapas do controle da expressão gênica e poderá abrir novos horizontes para terapias genéticas e manipulação de genes.

Tema 09: Aplicações da Bioinformática no estudo das patologias humanas.

Atualmente não há dúvidas sobre o impacto que a bioinformática tem nos estudos das patologias humanas.

A pandemia de SARS-CoV-2 evidenciou e reforçou a importância da sua aplicação para resolver problemas de saúde coletiva, neste caso em escala mundial.



A atuação da bioinformática ocorreu desde da resolução de milhares de genomas de SARS-CoV-2 em tempo recorde nas mais diferentes partes do globo. Na vigilância epidemiológica através do monitoramento de variantes em escala regional e global assim como na atuação direta na produção e desenvolvimento de vacinas.

A análise do genoma possibilitou a descoberta de genes alvo para produção de vacinas mais eficazes. Além disso, foi possível verificar em curto espaço de tempo a variação genética do vírus. ~~proporção~~ Esse conhecimento foi fundamental para categorizar em indicadores genéticos ~~patentes~~ variantes de risco ser monitoradas. Ademais, a aplicação da bioinformática possibilitou a organização e acessibilidade dos dados genômicos entre diferentes grupos de pesquisas. Essa interação e acessibilidade dos dados possibilitou o desenvolvimento de testes mais eficientes para diagnóstico em larga.

Com a epidemiologia genômica foi possível monitorar o fluxo de variantes do vírus em escala regional e global. A descoberta de novas variantes e seus respectivos centros de origem possibilitou a implementação de medidas sanitárias mais eficientes.

A integração de dados clínicos e os "dados omics" foi possível classificar o potencial patogênico do vírus (variantes) em seus possíveis hospedeiros (comorbidades).

Através da vacinologia reversa foi possível desenvolver vacinas em tempo recorde. De fato a bioinformática possibilitou esse cenário devido às diferentes simulações e modelagens de ~~testes~~ ensaios clínicos. O ponto alto foi o desenvolvimento de vacinas baseada em RNA por duas empresas Pfizer e Moderna. A geração destes produtos foi possível graças à aplicação da bioinformática.



e o seu amadurecimento ao longo dos eixos anteriores.

Tema 07: Bioinformática aplicada à análise epigenômica.

A epigenética é o processo de regulação de genes mais elaborado e complexo pois a necessidade ~~de~~ de modificações ou mutações da sequência de DNA. Na epigenômica é possível identificar rotular e prever quais regiões do genoma estão expressas ou disponíveis para ser expressas. No entanto essa regulação não é simples e envolve vários fatores (como metilação de fita de DNA, acetilação e retilação de histonas e empacotamento da fita de DNA por nucleossomos).

Com o desenvolvimento das Tecnologias de sequenciamento foi possível capturar e estudar a influência que alguns sítios destes marcadores epigenéticos podem exercer no genoma do organismo. Tais Tecnologias que estão disponíveis incluem a CHIP-seq e ATAC-seq. A função do CHIP-seq é identificar regiões no genoma (sequências de DNA) onde há interação proteína-DNA podendo ser Fatores Transcritores, RNA polimerase e outras proteínas com potencial regulatório.

A ATAC-seq usa uma transposase para identificar sequências de DNA acessíveis no genoma para interação proteína-DNA ou em alguns casos para sítios de RNA não codificantes (no geral microRNAs).

Estas Tecnologias geram um volume de dados muito elevado e na maioria das vezes com vários ruídos.

Por isso a aplicação de ferramentas de bioinformáticas e diferentes métodos matemáticos são



são necessários para resolver este problema.

Alguns ferramentas já estão disponíveis para estas tecnologias: MAEST e HOMER.

Estas ferramentas usam máxima verossimilhança para identificar regiões com reads superadas ("Pico") com significância estatística em relação do background ou sinais de ruído.

A epigenômica está em evolução tanto na geração dos dados quanto na análise e interpretação dos mesmos.