



CODIGO DA PROVA:

RP 001 / 0011



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.

1- Controle da expressão gênica com ênfase em métodos computacionais

O controle da expressão gênica está relacionado as diferentes formas que molecular ou estruturalmente e há o controle nas etapas de transcrição, tradução ou produtos finais. O avanço nas técnicas de sequenciamento e no desenvolvimento de ~~metodos~~ metodologias e ferramentas computacionais favorecer a descoberta e compreensão do controle da expressão.

O principal dado biológico usado para o estudo nesta área são os dados de sequenciamento de RNA (RNA-seq) permitindo a identificação dos transcritos ~~dis~~ disponíveis disponíveis, o que permite a interpretação em diferentes estados celulares além da relação doença/saúde ou tratamento/controle.

O sequenciamento de RNA é feito usando plataformas de sequenciamento de segunda geração (mas também usado), ou sequenciamento de terceira geração para estudos mais complexos, além disso, o sequenciamento de 1ª geração pode ainda ser usado para validar/confirmar a descoberta específicas. Estes dados de sequenciamento em larga escala precisam ser ~~processados~~ processados utilizando ferramentas e metodologias computacionais que permitem a análise de grande volume de dados gerados a fim de permitir a interpretação biológica.

Tais ferramentas computacionais são frequentemente ~~compatíveis~~ compatíveis e desenvolvidas em ambiente Linux, com uso



de linguagens de programação como R, python e bash. Essas ferramentas podem ser ainda usadas em interfaces amigáveis que podem utilizar linguagens em java e C.

O desenvolvimento de metodologias computacionais para o estudo do controle da expressão gênica pode envolver o desenvolvimento de algoritmos próprios que organizem as informações e usem a estatística para interpretações dos dados. Algumas ferramentas ~~utiliza~~ que compõem pipelines de análise de RNA-seq incluem HISAT2, STAR, Bowtie, Bowtie2 e BWA para alinhamento de sequências utilizando genomas de referência, ferramentas para análise de expressão diferencial como DESeq e EdgeR que usam as matrizes de contagem calculadas pelo featureCounts para identificar genes diferencialmente expressos. Genes regulados positivamente (up-regulados), ou regulados negativamente (down-regulados) podem estar sendo regulados por uma complexa rede regulatória com elementos como fatores de transcrição, miRNA, RNA não codificantes, fatores epigenéticos, eventos de splicing, além de retrovírus.

O estudo desses elementos regulatórios pode ser feito com uso de redes regulatórias com ferramentas como FIMO usando dados disponíveis em bancos de dados públicos com BioGRID, a fim de compreender quais fatores de transcrição são ativados e quais miRNAs seus alvos. Análises de redes de co-expressão usando ferramentas como WGCNA e plataformas como Fluxon base adicionam camadas informativas que ajudam a elucidar essas mecânicas e prever funções regulatórias e alvos adicionais de elementos ainda não classificados.

Dentre os elementos regulatórios que podem ser identificados no transcriptoma estão os microRNA e RNA não codificantes. Ferramentas e bancos de dados como miRDB ajudam a entender e agrupar informações experimentais e preditivas ~~sobre~~ sobre os miRNA e seus alvos permitindo



compreender esse comando regulatório, adicionalmente o perfil de expressão obtido pelas análises de expressão diferencial supracitadas ajudam a relacionar a expressão do miRNA e seu alvo ou RNA não codificante e seu alvo, identificando o potencial regulatório por tais elementos. Em muitos casos a validação experimental ou computacional que permite, por exemplo, modelar a ligação entre o elemento regulatório e seu alvo são usadas para confirmar os dados obtidos. Neste sentido o uso da inteligência artificial e aprendizado de máquinas auxiliam as etapas de predição e modelagem aumentando a acurácia e diminuindo tempo de análise e alguns casos diminuindo ou usando de forma mais eficaz ~~o~~ a estrutura computacional requerida. Neste sentido a IA e ML podem auxiliar em casos onde não existe disponibilidade de informações sobre elementos regulatórios com fatores de transcrição TF ou CTCF, por exemplo, permitindo que eles sejam preditos, por exemplo, por reconhecimento de padrões.

Métodos de controle de expressão gênica incluem ~~em~~ eventos epigenéticos com metilação do DNA obtida pelo sequenciamento de bisulfito e analisado com ~~programas~~ <sup>programas</sup> ferramentas como bismark para identificação de sítios metilados ~~e~~ <sup>isto</sup> auxiliando a identificação de regiões de metilação diferencial, normalmente relacionados ao controle da expressão gênica; outros eventos epigenéticos ~~envolvem~~ <sup>envolvem</sup> a modificação de histonas estudadas a partir da análise de dados de ChIP-seq além dos ~~umodelamento~~ <sup>umodelamento</sup> da cromatina ~~esta~~ <sup>esta</sup> com dados de ATAC-seq e dobramento da cromatina por Hi-C. ~~Estes~~

Estes dados ajudam a modular a acessibilidade à cromatina, além da conexão entre partes diferentes da cromatina que permitem, por exemplo, a ligação entre um gene de transcrição e um promotor ou enhancer.

Além disso, o estudo de eventos de splicing com



Ferramentas com RNAi ajudam a compreender o nível de regulação pós-transcricional e possíveis informações gênicas expostas.

É importante obter o ~~estudo~~ entendimento integrado de diversos camadas ômicas, como transcriptômica, proteômica e metabolômica a fim de obter uma visão holística que agregue dados de ~~camadas regulatórias~~ regulatórias para identificação de ~~este~~ controle da expressão gênica.

Aliado a isto ~~o~~ infraestrutura computacional com uso de servidores, ~~use~~ ~~de~~ ~~nuvem~~ ~~e~~ auxiliam as análises computacionais. Vale ressaltar que isto é uma limitação para as análises que podem ser atendidas com uso de bancos de dados públicos e repositórios de dados que podem ser executados não só por linha de comando ~~e~~ como ~~em~~ muitas vezes em interface amigável.

## 7- Bioinformática aplicada a análise epigenômica.

As análises epigenômicas incluem ~~o~~ o estudo da metilação do DNA, modificação de histonas, ~~como~~ ~~o~~ remodelamento da cromatina e ~~em~~ o dobramento da cromatina. A metilação do DNA envolve a metilação de citosinas no contexto dinucleotídico CpG, ~~o~~ estes são ~~utilizados~~ ~~para~~ ~~análise~~ ~~de~~ ~~sequências~~ ~~longas~~ ~~com~~ ~~plataformas~~ ~~como~~ ~~PacBio~~ ~~ou~~ ~~Nanopore~~ ~~que~~ ~~permitem~~ ~~o~~ ~~sequenciamento~~ ~~em~~ ~~tratamento~~ ~~prévio~~ ~~do~~ ~~del~~ ~~molécula~~ ~~de~~ ~~DNA~~, ~~permitindo~~ ~~compreender~~ ~~melhores~~ ~~momentos~~ ~~específicos~~ ~~que~~ ~~precisam~~ ~~ser~~ ~~estudados~~. Ferramentas como biomark ajudam a analisar dados de sequenciamento de bisulfite e ~~identifi~~ ~~permi~~ ~~tindo~~ a identificação de



regiões diferencialmente metiladas. Além disso, há possibilidade de estudar a metilação de DNA a partir de sequenciamento de pares de CpG 450K ou mais diminuindo o custo e focando em regiões ~~que~~ que precisam ser estudadas. ~~A metil~~ A metilção do DNA pode ser acessada ~~através de análises epigenômicas~~ e analisada com ferramentas e metodologias computacionais que favorecem o desenvolvimento de bancos de dados como Encode e Roadmap epigenômics que apresentam informações de metilação em diversos tecidos, tratamentos e estado do doente. Tais dados podem ainda ser visualizados em Genome Browser como UCSC genome browser que permite integrar diversos camadas epigenômicas de forma gráfica.

Bancos de dados como os supracitados permitem o download de dados brutos ou parados em análises específicas, dessa forma, é possível ~~desenvolver~~ desenvolver algoritmos e pipelines computacionais próprios que processam os dados e ajudam na interpretação biológica ~~de~~ associada ao objetivo da pesquisa.

Outra análise epigenômica envolve a identificação de modificações de histonas que podem estar ligadas a grupos de nucleossomos que favorecem a acetilação, metilação, fosforilação e ubiquitinação. Essas modificações podem ser estudadas a partir do sequenciamento ChIP-seq que usa anticorpos obtidos a partir de imunoprecipitação. Diversos fragmentos em bioinformática permitem identificar os picos de enriquecimento de histonas associados a localizações gênicas (promotor, exons, intron, 3' ou 5' UTR) para compreender ~~o~~ o perfil de acessibilidade a cromatina em estado de repressão ou estímulo da expressão.

Neste sentido, o estudo epigenômico é ~~uma~~ muitas vezes associado ao sequenciamento de RNA, permitindo associar dados de regulação epigenômica ao perfil de expressão gênica.



Considerando a avaliação da acessibilidade à cromatina dados de ATAC-seq permitem verificar o enrolamento da cromatina e a acessibilidade a ~~varios~~ mesma. Além disso, dados de Hi-C permitem avaliar o dobramento da cromatina que permite a ligação ~~de~~ ou interação de pontos distintos do genoma e permite compreender não somente interações de fatores de transcrição ao alvo, com compreender a organização do material genético em diferentes estágios e ~~em~~ processos celulares, ~~por~~ por exemplo.

Neste sentido, a bioinformática ~~é~~ aplicada a análise epigenômica permite o estudo do desenvolvimento embrionário, diferenciação celular, além da influência ambiental em diferentes estados e doenças biológicas. Estudos de estados saudáveis e em doenças permitem compreender como a doença/patologia influencia no estado epigenômico e como ela pode ser afetada. Dessa forma, a bioinformática ~~aliada a~~ ~~é~~ aplicada nesta área ~~o~~ ajuda a entender os duas vias como os sistemas biológicos respondem ou são influenciados epigenômica pela doença, ou determinadas condições ambientais além de ajudar na descoberta de drogas para o combate a ~~estas~~ condições.

Limitações neste tema incluem a ~~simples~~ infraestrutura computacional e experimental para estudar e gerar uma quantidade massiva de dados; variabilidade em técnicas, protocolos e metodologias de sequenciamento e análise computacional, além da própria heterogeneidade dos processos biológicos que podem ser influenciados por diversos, por exemplo, durante o preparo de amostras e armazenamento de dados.



## 9- Aplicação da bioinformática no estudo das patologias humanas.

A bioinformática vem auxiliando o estudo das patologias humanas e ampliando as informações e precisão para diagnóstico e tratamento. A bioinformática pode ser aplicada à medicina de precisão, diagnóstico genético, estudo epidemiológico além da vigilância genômica e produção de vacinas.

~~Na medicina~~ Na medicina de precisão a bioinformática permite a descoberta de variantes de risco a partir de dados genômicos ou sequenciamento de exoma total, além de fatores moleculares que podem ser usados como biomarcadores com elementos regulatórios ou genes com perfil de expressão modificada em estados patológicos, estes últimos utilizam dados de sequenciamento de RNA, miRNA além de análise de dados epigenômicos.

Para dar suporte a essas análises de bancos de dados públicos e ferramentas computacionais (~~disponibilizam~~ ~~informações~~ e metodologias de análise) disponibilizam informações e metodologias de análise que integradas à análise estatística permitem a interpretação de dados biológicos e a compreensão da patologia.

O diagnóstico genético e descoberta de fatores moleculares podem ser usados na farmacogenética no estudo da influência genética no uso de drogas medicamentosas. Neste sentido, uso de ferramentas computacionais que usam inteligência artificial e algoritmos de máquinas aumentam a precisão de análise permitindo, por exemplo, desambiguar análise que modulam a ligação do fármaco com o alvo.

Considerando o uso de IA e aprendizado de máquinas na modulação de ligação molecular, existe um avanço na sua aplicação em estudos que requerem a ligação de ~~epitopos~~ epitopos com moléculas



de MHC ou ligação antígeno - anticorpo que ajudam na descoberta de vacinas e desenvolvimento de vacinas. Neste contexto, o custo computacional pode ser maior com requerimento de uma infraestrutura computacional mais robusta.

A análise de imunopeptídeos ou epítopos capazes de ativar a resposta imune é uma área extremamente abrangente. O uso de dados de sequenciamento de genoma, exoma, transcriptoma e proteoma podem ser usados de forma única ou integrados, ajudando na identificação e análise da análise.

É possível usar ferramentas computacionais para fazer a predição de epítopos como por exemplo NetMHCpan, NetCTLpan, NetMHCIIpan, para descoberta de epítopos que são reconhecidos pelos moléculas de MHC, assim como ferramentas como BepiPred para identificação de epítopos que se ligam a receptores de célula B. Uma variedade de ferramentas existem para avaliar a imunogenicidade de um epítopo e seu potencial vacinal. Os dados de sequenciamento podem ainda ser usados para predição de ~~estes~~ alelos HLA ~~Human~~ (Antígeno leucocitário humano) um grupo de genes que codificam as moléculas de MHC citadas acima. Tal predição ajudam a compreender o puzzle da resposta imune mediada por célula T nos pacientes, ajudando a compreender o perfil HLA de pacientes ~~que~~ susceptíveis a determinadas doenças, por exemplo, ou aqueles pacientes ~~que~~ que respondem de forma satisfatória a vacinação, comparados aqueles que apresentam baixa cobertura vacinal.

A bioinformática ~~ajuda~~ auxilia nos estudos de doenças como câncer ou doenças neurológicas, permitindo não só a descoberta de alvos moleculares como permitindo a identificação, por exemplo, e a localização espacial de transcritos na célula, sem a





a necessidade de cultivo celular. O estudo do transcrito-toma espacial e do transcrito-toma de célula única permitem localizar os transcritos regulados positivamente ou negativamente dentro da célula. Este estudo pode ser aplicado na compreensão de patologias complexas ou ~~tais~~ aquelas em tecidos de difícil acesso, como cânceres e doenças imunológicas.

O estudo epigenômico auxilia na compreensão de doenças que sofrem influência ambiental como dieta, tabagismo, estilo de vida. Doenças crônicas e metabólicas como diabetes, hipertensão, doenças cardiovasculares podem não somente ser melhor compreendidas com sequenciamento de RNA e DNA como com a adição de dados epigenéticos e descodificação de ~~transcritos~~ elementos regulatórios.

Dessa forma, a bioinformática auxilia ~~na saúde coletiva~~ na saúde coletiva. A aplicação pode ainda ser vista no estudo epidemiológicos com organização e modulação e predições de padrões de dados demográficos, perfil populacional, padrões de distribuição de patógenos ou vetores de doenças. Assim, a bioinformática aplicada a epidemiologia permite identificar e associar doenças, gravidade de ~~doenças~~ quadro clínico, susceptibilidade com focos endêmicos e epidêmicos, ajudando nas medidas profiláticas, diagnósticas e tratamento. Além disso, é possível manter a vigilância genômica, usando a bioinformática para identificação de cepas ou variantes virais ~~em~~ emergentes, por exemplo.

Em estudos populacionais dados como os disponíveis no SUS, ~~as~~ análises estatísticas aliadas ~~à~~ a bioinformática, ajudam a compreender o ~~o~~ perfil de determinada doença na população, assim como em caso de doenças específicas de grupos populacionais distintos.



Por fim a bioinformática ~~ajuda~~ amplia o estudo de ~~gen~~ doenças genéticas como síndrome de Down, triplo X, entre outras que são foco da ~~genética~~ genética clássica. Eventos epigenômicos & relacionados ao desenvolvimento embrionário, por exemplo, ajudam a compreender tais doenças. Outra comada ~~genética~~ informática diz respeito, ~~as~~ por exemplo, ao perfil de expressão gênica da cópia extra do cromossomo, ~~ou a~~ ~~origem~~ fornecendo clues & biomarcadores para terapia. Há ainda a análise da origem parental do cromossomo extra apoiada pelas análises & epigenômicas & transcriptômicas e realizadas por sequenciamento de Sanger, isso auxilia em casos onde há possibilidade de desenvolvimento genético.

Dessa forma, ~~desde~~ desde doenças infecciosas até as doenças genéticas podem ser melhor compreendidas com o uso da bioinformática.

|