

- O que é RNA Seq?

É uma abordagem desenvolvida para analisar o perfil de transcriptoma* usando tecnologias de sequenciamento next generation.

*Conjunto completo de transcritos (RNAs mensageiros, RNAs ribossômicos, RNAs transportadores e os microRNAs) de um dado organismo, órgão, tecido ou linhagem celular.

- Aplicações no estudo de doenças genéticas:

1. Detecta alterações na molécula de RNA em células cancerosas
2. Permite identificar níveis de expressão e de novos transcritos
3. Identifica mecanismo de splicing alternativo e fusões gênicas
4. Detecta mutações, entre elas rearranjos cromossômicos, que levam a transcritos quiméricos

- Ferramentas de alinhamento:

1. BFast: facilita o mapeamento rápido e preciso de sequências curtas com sequências de referência. Vantagens: velocidade e precisão.
2. Bowtie: alinha sequências curtas de DNA com genoma humano.
3. BWA: sequenciamento rápido, preciso, eficiente em relação ao uso de memória; para sequências curtas e longas.
4. Kallisto: é um programa para quantificar a abundância de transcrições a partir de dados de RNA-Seq. Pseudo Alinhamento de sequências preserva a informação chave necessária para a quantificação e Kallisto é, portanto, não apenas rápido, mas também tão preciso como ferramentas de quantificação existentes.

- Projeto:

O projeto consiste em realizarmos a análise de gene diferencialmente expresso, de forma que, para o usuário comum, esse processo seja simplificado ao máximo. Evitando que ele se 'perca' em meio aos processos intermediários até o resultado da análise final.

Análise de RNA Seq baseada em genoma referência, utilizando TopHat para alinhamento, CuffLinks para montagem do transcriptoma.

Após todo esse processo de limpeza, alinhamento e montagem, utilizaremos um conjunto de programas para alinhamento múltiplo de sequência e análise e geraremos o GDE (gene diferencialmente expresso).

A partir desse estudo, poderemos encontrar variantes que poderam determinar enfermidades, como no caso do câncer.

Por fim, utilizaremos:

- GO, Gene Ontology, permitindo a interpretação funcional de dados experimentais, através de análise de enriquecimento
- KEGG é um recurso de banco de dados para a compreensão de funções de alto nível e utilitários do sistema biológico, como o celular, o organismo e o ecossistema, a partir de informações de nível molecular, especialmente conjuntos de dados moleculares em larga escala gerados pelo sequenciamento do genoma e outros de alto rendimento tecnologias experimentais.(Fonte:<http://www.genome.jp/kegg/>)
- COGS: O banco de dados de Clusters de Orthologous grupos de proteínas (CPV) é uma tentativa de classificação filogenética das proteínas codificadas em genomas completos. Cada COGs inclui proteínas que são inferidas a ser orthologs (homólogos evolutivos diretos).