O artigo "Next-Generation Sequencing Technology: Current Trends and Advancements" explica como as tecnologias de sequenciamento de nova geração (NGS) mudaram a forma como pesquisamos genética e genômica. Essa tecnologia permite sequenciar muitos pedaços de DNA ao mesmo tempo, o que torna o processo mais rápido e barato do que as técnicas antigas. Isso ajudou a avançar em áreas como a análise de diferenças genéticas, estudos de como os genes se expressam e mudanças químicas que controlam os genes.

Os autores falam sobre melhorias recentes no NGS, como métodos que aumentam a precisão e reduzem erros. Além disso, novos programas de computador ajudam a organizar e analisar a grande quantidade de dados gerados. Apesar dessas melhorias, ainda existem desafios, como o alto custo dos equipamentos e a dificuldade em analisar os dados de forma confiável.

O NGS é usado em várias áreas, como diagnóstico de doenças raras, desenvolvimento de tratamentos personalizados e estudos sobre bactérias em diferentes ambientes. Também ajuda a entender melhor a diversidade genética e como ela afeta a saúde humana e o meio ambiente, com impacto na ciência, na medicina e na saúde pública.

O UCSC Table Browser é uma ferramenta online que ajuda pesquisadores a acessar informações genômicas. Com ela, é possível procurar dados sobre genes, mutações e outras características do DNA de forma fácil e personalizada.

A ferramenta permite filtrar dados com critérios específicos ou usar comandos mais avançados para fazer buscas detalhadas. Os resultados podem ser salvos em vários formatos, como tabelas que funcionam em planilhas ou arquivos que podem ser usados em outros programas. Também dá para baixar sequências de DNA de partes específicas do genoma. Isso tudo facilita o trabalho dos pesquisadores, sem que precisem ter cópias locais dos grandes bancos de dados.

O white paper da Twist Bioscience explica como a empresa facilita o estudo de partes específicas do DNA no sequenciamento NGS. A tecnologia da Twist ajuda cientistas a analisar regiões importantes do DNA de forma mais eficiente, economizando tempo e dinheiro. O segredo está em sondas de DNA muito precisas, que selecionam as partes certas para serem estudadas. Isso evita erros e reduz a quantidade de dados inúteis. A tecnologia da Twist é versátil e pode ser usada em estudos sobre mutações raras, genética personalizada e pesquisas sobre doenças como o câncer. Por ser precisa e confiável, é uma ferramenta importante para áreas como o desenvolvimento de novos tratamentos médicos.

Next-Generation Sequencing (NGS) é uma tecnologia revolucionária que permite o sequenciamento rápido e preciso de DNA e RNA, transformando a pesquisa genômica e suas aplicações práticas.

A tecnologia NGS utiliza plataformas automatizadas que sequenciam milhões de fragmentos de DNA simultaneamente. Essa abordagem paralela permite obter grandes quantidades de dados em um curto período. Existem dois métodos principais:

* **Single-Read Sequencing**: Sequencia apenas uma extremidade de cada fragmento de DNA. É mais rápido e utilizado para estudos como RNA-Seq.
* **Paired-End Sequencing**: Sequencia ambas as extremidades de cada fragmento, proporcionando maior precisão para análise de variantes estruturais e montagens de genomas complexos.

A **preparação da biblioteca** é uma etapa crítica, envolvendo a fragmentação do DNA, a ligação de adaptadores e a amplificação por PCR. A escolha dos kits de preparo depende do objetivo do experimento e das plataformas utilizadas.

**Aplicações do NGS**

1. **Targeted Resequencing**: Permite investigar genes específicos associados a doenças ou traços. Usado em medicina personalizada para identificar mutações genéticas.
2. **Deep Sequencing**: Gera dados de alta profundidade, permitindo a detecção de variantes raras e a análise de populações microbianas complexas.
3. **Mate Pair Sequencing**: Ideal para identificar grandes reordenamentos genômicos e mapear regiões repetitivas.
4. **Sample Multiplexing**: Usa índices para sequenciar várias amostras em uma única corrida, reduzindo custos e aumentando a eficiência.

**Bases Moleculares do Sequenciamento**

O sequenciamento baseia-se na detecção de nucleotídeos incorporados durante a síntese de DNA. As principais plataformas (Illumina, Ion Torrent) diferem nos métodos de detecção:

* **Illumina**: Utiliza a síntese reversível de terminadores fluorescentes, gerando alta precisão e rendimento.
* **Ion Torrent**: Detecta mudanças no pH geradas pela incorporação de nucleotídeos, sendo uma alternativa mais acessível.

A cobertura de sequenciamento é essencial para garantir a qualidade dos dados. Uma maior cobertura aumenta a confiança na detecção de variantes, especialmente em regiões de baixa expressão.

**Design**

* **Seleção de Alvos**: Escolha de genes ou regiões a serem sequenciados.
* **Preparação de Biblioteca**
* **Multiplexação e Índices**: Planejamento para combinar amostras sem perda de integridade.

**Ferramentas e Plataformas NGS**

**Illumina:**

* **Sequencing Platforms**: Sistemas como NovaSeq, NextSeq e MiSeq atendem diferentes demandas, desde estudos de expressão gênica até análise de genomas completos.
* **Experiment Planning Tools**: Ferramentas que auxiliam no planejamento e escolha dos kits de biblioteca mais adequados.
* **NGS Library Preparation**: Oferece kits de alta eficiência para uma ampla variedade de aplicações.

**Ion Torrent:**

* **Construct Library**: Kits otimizados para fragmentação e amplificação do DNA.
* **NGS Instruments**: Sistemas como Ion GeneStudio S5 são projetados para estudos direcionados e sequenciamento de baixo custo.

**Comandos :**

O pipeline processa dados de sequenciamento para encontrar mutações genéticas e analisar sua relevância clínica. Esse fluxo é usado para identificar variantes somáticas em câncer e associá-las a possíveis impactos funcionais.

**1. Download de arquivos**

!wget -c https://hgdownload.soe.ucsc.edu/goldenPath/hg19/chromosomes/chr9.fa.gz

Baixa o arquivo **chr9.fa.gz**, que contém a sequência do cromossomo 9 da referência do genoma **hg19**. O -c permite continuar um download interrompido.

**2. Visualização do conteúdo**

!zcat chr9.fa.gz | head

* zcat chr9.fa.gz: Descompacta e exibe o conteúdo do arquivo comprimido **chr9.fa.gz** sem salvar.
* head: Mostra as primeiras 10 linhas do arquivo.

**3. Processamento da sequência**

!zcat chr9.fa.gz | sed -e "s/chr//g" > chr9.fa

* zcat chr9.fa.gz: Descompacta o arquivo.
* sed -e "s/chr//g": Remove todas as ocorrências da palavra **"chr"** do arquivo.
* > chr9.fa: Salva a saída no arquivo **chr9.fa**.

**4. Conferir as primeiras linhas do arquivo**

!head chr9.fa

* Exibe as primeiras 10 linhas do arquivo **chr9.fa** para confirmar que a modificação foi aplicada corretamente.

**5. Instalação do Samtools**

!sudo apt-get install samtools

* Instala o **Samtools**, um conjunto de ferramentas para manipulação de arquivos de sequenciamento de DNA.

**6. Indexação do genoma**

!samtools faidx chr9.fa

* Cria um **índice** para o arquivo **chr9.fa**, permitindo buscas rápidas na sequência.

**7. Extração e configuração do GATK**

!unzip gatk-4.2.2.0.zip

* Descompacta o **GATK 4.2.2.0**, uma ferramenta usada para chamar variantes genéticas.

!./gatk-4.2.2.0/gatk

* Executa o GATK para verificar se está funcionando corretamente.

**8. Criar um dicionário para o genoma**

!./gatk-4.2.2.0/gatk CreateSequenceDictionary -R chr9.fa -O chr9.dict

* Cria um **dicionário** do genoma **chr9.fa**, necessário para várias ferramentas do GATK.

**9. Criar intervalos de referência**

!./gatk-4.2.2.0/gatk ScatterIntervalsByNs -R chr9.fa -O chr9.interval\_list -OT ACGT

* Gera uma lista de intervalos de regiões do genoma para processamento paralelo.

**10. Identificação de variantes somáticas (câncer)**

!./gatk-4.2.2.0/gatk Mutect2 \

-R chr9.fa \

-I somatico/tumor\_JAK2.bam \

-I somatico/normal\_JAK2.bam \

-normal WP044 \

--germline-resource somatico/af-only-gnomad-chr9.vcf.gz \

-O somatic.vcf.gz \

-L chr9.interval\_list

* **Mutect2** é uma ferramenta do GATK usada para detectar variantes somáticas (câncer).
* -R chr9.fa: Usa **chr9.fa** como referência.
* -I tumor\_JAK2.bam e -I normal\_JAK2.bam: Amostras de DNA do tumor e do tecido normal.
* --germline-resource: Usa um banco de variantes germinativas (herdadas).
* -O somatic.vcf.gz: Salva o resultado em **somatic.vcf.gz**.
* -L chr9.interval\_list: Define intervalos do cromossomo 9.

**11. Coleta de estatísticas de variantes**

!./gatk-4.2.2.0/gatk GetPileupSummaries \

-I somatico/tumor\_JAK2.bam \

-V somatico/af-only-gnomad-chr9.vcf.gz \

-L chr9.interval\_list \

-O tumor\_JAK2.table

* Coleta estatísticas de cobertura da amostra tumoral.

!./gatk-4.2.2.0/gatk GetPileupSummaries \

-I somatico/normal\_JAK2.bam \

-V somatico/af-only-gnomad-chr9.vcf.gz \

-L chr9.interval\_list \

-O normal\_JAK2.table

* Coleta estatísticas da amostra normal.

**12. Anotação funcional das variantes com Ensembl VEP**

!./ensembl-vep/vep \

-i filtered.vcf.gz \

-o filtered.vep.tsv \

--database --assembly GRCh37 --refseq \

--pick --pick\_allele --force\_overwrite --tab --symbol --check\_existing \

--fields "Location,SYMBOL,Consequence,Feature,Amino\_acids,CLIN\_SIG" \

--fasta chr9.fa \

--individual all

* Usa o **Ensembl Variant Effect Predictor (VEP)** para **anotar** as variantes detectadas.
* -i filtered.vcf.gz: Arquivo de variantes filtradas.
* -o filtered.vep.tsv: Salva a saída no formato **TSV**.
* --fields: Campos de interesse (posição, gene, consequência funcional, impacto clínico, etc.).
* --fasta chr9.fa: Usa **chr9.fa** como referência.

**13. Exibir as variantes anotadas**

!cat filtered.vep.tsv

* Exibe o conteúdo do arquivo com as variantes anotadas.

Observações ao longo dos estudos:

Genética estuda partes específicas do DNA (genes) e como elas influenciam características ou doenças.

Genoma é o estudo do DNA inteiro, analisando como todos os genes e sequências trabalham juntos para formar um organismo.

O site da Illumina apresenta uma visão geral sobre as tecnologias de sequenciamento genético que a empresa oferece. Ele destaca como o sequenciamento de DNA revolucionou áreas como pesquisa biomédica, diagnósticos e medicina personalizada. A Illumina utiliza tecnologias avançadas para ler o material genético com alta precisão, permitindo identificar mutações, estudar doenças genéticas e explorar a biodiversidade.

O conteúdo explica os dois métodos principais de sequenciamento: **sequenciamento de leitura curta (short-read)** e **sequenciamento de leitura longa (long-read)**, enfatizando que a abordagem da Illumina é baseada em leituras curtas, que são rápidas, acessíveis e confiáveis. Essa tecnologia é usada para estudos como análise do genoma completo, exoma e RNA, oferecendo dados detalhados para entender a biologia em nível molecular.

Além disso, o site destaca aplicações práticas do sequenciamento, como na pesquisa do câncer, saúde reprodutiva e doenças infecciosas, ajudando cientistas e médicos a encontrar soluções inovadoras para problemas complexos de saúde e biologia.

O genoma de referência é um modelo padrão usado para comparar sequências de DNA, como o da Universidade da Califórnia (UCSC Genome Browser). Sequenciadores leem fitas de DNA e geram dados que podem ser alinhados com esse genoma de referência. A diferença entre genoma e gene é que o genoma é todo o DNA de um organismo, enquanto um gene é uma parte específica que contém instruções para funções biológicas.

No Linux, comandos como cd (navegar entre diretórios) e ls (listar arquivos) são úteis para organizar e acessar os arquivos de sequenciamento.

Cromossomos são grandes estruturas de DNA organizadas dentro do núcleo celular, compostos por nucleotídeos, que são as unidades básicas do DNA.

O **Ion Torrent** é uma tecnologia prática e acessível para sequenciamento, ideal para análises direcionadas. Após o sequenciamento, os dados são analisados em um pipeline que inclui etapas como pré-processamento, alinhamento e chamada de variantes. A escolha de softwares e ferramentas depende do objetivo do experimento, sendo essencial garantir qualidade e precisão nos resultados.

**Aplicações Preferenciais**

**Ion Torrent**

* Ideal para estudos de menor escala ou direcionados, como:
  + **Targeted resequencing**: Investigação de genes específicos ou pequenos painéis genômicos.
  + **Diagnósticos clínicos rápidos**: Doenças infecciosas, câncer, ou variantes conhecidas.
  + Experimentos que requerem menor investimento inicial.

**Illumina**

* Excelente para projetos de maior escala e abrangência, como:
  + **Sequenciamento de genomas inteiros**: Genoma humano ou organismos complexos.
  + **RNA-Seq**: Estudos de expressão gênica com alta precisão.
  + **Metagenômica e microbioma**: Análise de comunidades microbianas com diversidade genética alta.
  + **Sequenciamento de exomas**: Detecção de variantes em regiões codificantes do genoma.

**Tempo de Sequenciamento**

**Ion Torrent**

* Sequenciamento mais rápido, devido à simplicidade do método de detecção (mudanças de pH).
* Menor tempo de preparação e processamento, ideal para resultados urgentes.

**Illumina**

* Sequenciamento mais demorado, já que utiliza ciclos de síntese com detecção fluorescente e múltiplas etapas de processamento.
* Maior tempo para estudos complexos, mas oferece maior profundidade de dados.

**Custo e Acessibilidade**

**Ion Torrent**

* Geralmente mais acessível em termos de custo inicial e manutenção de equipamentos.
* Kits de preparo e reagentes podem ser mais econômicos, mas geram menor volume de dados por corrida.

**Illumina**

* Equipamentos têm custo inicial mais alto.
* Kits de preparo de biblioteca são mais caros, mas oferecem maior rendimento e precisão.

**Qualidade dos Dados**

**Ion Torrent**

* Pode apresentar maior taxa de erro, especialmente em sequências ricas em homopolímeros (repetições da mesma base, como AAAAA).
* Dados são adequados para análises direcionadas, mas menos indicados para genomas grandes ou complexos.

**Illumina**

* Alta precisão e baixa taxa de erro.
* Melhor escolha para análises que exigem qualidade e profundidade, como genomas completos ou regiões de difícil sequenciamento.