



UNIVERSIDADE ESTADUAL PAULISTA
"Júlio de Mesquita Filho"
Instituto de Biociências de Botucatu

Nicole Malachias Nacarato

**SEQUENCIAMENTO DE NOVA GERAÇÃO COMO FERRAMENTA DE
DETECÇÃO DE VARIANTES GENÉTICAS DE ORIGENS SOMÁTICAS
E GERMINATIVAS**

BOTUCATU - SP

2023



Nicole Malachias Nacarato

SEQUENCIAMENTO DE NOVA GERAÇÃO COMO FERRAMENTA DE DETECÇÃO DE VARIANTES GENÉTICAS DE ORIGENS SOMÁTICAS E GERMINATIVAS

Trabalho de Conclusão de Curso (TCC) apresentado ao Instituto de Biociências, campus de Botucatu, UNESP, para obtenção do título de Bacharel em Ciências Biomédicas.

Orientadora: Prof. Dra. Cláudia Aparecida Rainho

Coorientadora: Dra. Dirce Maria Carraro

BOTUCATU - SP

2023

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: MARIA CAROLINA ANDRADE CRUZ E SANTOS-CRB

Nacarato, Nicole Malachias.

Sequenciamento de nova geração como ferramenta de detecção de variantes genéticas de origens somáticas e germinativas / Nicole Malachias Nacarato. - Botucatu, 2023

Trabalho de conclusão de curso (bacharelado - Ciências Biomédicas) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Cláudia Aparecida Rainho

Coorientador: Dirce Maria Carraro

Capes: 20205007

1. Sequenciamento de nucleotídeos. 2. Oncologia.
3. Variação genética.

Palavras-chave: NGS; Oncologia personalizada; Variantes genéticas.

AGRADECIMENTOS

Agradeço à minha mãe, Luciana, minha melhor amiga e parceira desde minhas primeiras memórias. Mãe, essa conquista também é sua! Ao meu pai, Orlando, que descobriu e nos apresentou um novo lado da vida e, mesmo com as dificuldades, sempre me faz sorrir. À minha família, tia Carol, tio Lu, Ravi, vó Maria, pelos conselhos, apoio e cuidado.

Agradeço aos meus amigos de turma e agora de profissão, Bruna, Debora, Guilherme, Mai e Poliana, não tenho palavras para dizer o quanto vocês são especiais. À minha amiga e primeira colega de quarto, Mariana (Ita), vamos juntas sempre, por favor! Aos meus amigos Samuel, Paola, “error 404”, que mesmo longe sempre estiveram presentes.

Agradeço todos que, durante a graduação e estágios, fizeram com que as respectivas cidadesse tornassem uma casa para mim. À República VaiDaNada, minha segunda família, obrigada por me ensinarem tanto sobre a vida. À República Medusa, obrigada por apresentarem uma realidade nova à essa unespiana perdida na Unicamp. Aos amigos Juliana e Victor, obrigada pelo acolhimento e por tornarem a cidade de São Paulo menos difícil.

Agradeço à Dra. Dirce Maria Carraro pela oportunidade de realizar o estágio curricular e iniciação científica no CIPE - A.C. Camargo, bem como ao Grupo de Genômica Clínica e Funcional. À Prof. Dra. Célia Regina Garlipp pela oportunidade de estágio na Divisão de Patologia Clínica do HC Unicamp. À Prof. Dra. Cláudia Aparecida Rainho pela disponibilidade e orientação neste trabalho.

Agradeço à comunidade da UNESP Botucatu, em especial do Instituto de Biociências e a coordenação do curso de Ciências Biomédicas, por proporcionarem um ambiente tão rico para nós. Sou muito grata por ter vivido essa experiência num lugar tão bom. Obrigada.

RESUMO

As metodologias de sequenciamento de DNA têm sofrido grandes transformações nas últimas duas décadas com o advento do Sequenciamento de Nova Geração (NGS). Esse método tem revolucionado a pesquisa básica e as aplicações clínicas na área da genética e genômica. A técnica é robusta, permitindo a obtenção da sequência de múltiplos fragmentos de DNA simultaneamente, tornando a sua execução mais rápida e acessível. O NGS consiste em um conjunto de diferentes tecnologias baseadas na incorporação de nucleotídeos à fita em síntese, permitindo a determinação da sequência de interesse. Existe uma variedade de análises que pode ser feita a partir do NGS: sequenciamento de genoma completo (WGS – *whole genome sequencing*), exoma completo (WES – *whole exome sequencing*), painéis multigênicos e sequenciamento alvo específico (*targeted sequencing*), que inclui o sequenciamento de *amplicons*, ou seja, pequenas regiões genômicas amplificadas por PCR (*targeted amplicon sequencing*). O NGS de *amplicons* com alta cobertura permite uma análise mais direcionada e sua implementação tem se expandido cada vez mais, sendo uma das aplicações mais importantes a estimativa da ploidia baseada nas variantes genéticas, tanto germinativas como somáticas, além de permitir a detecção de variantes em baixa frequência. Assim, apresenta grande aplicabilidade no campo da oncologia, tanto em pesquisa científica como na prática clínica da oncologia personalizada. O NGS de *amplicon* de alta cobertura é bastante utilizado para avaliação de variantes patogênicas conhecidas em pacientes suspeitos de síndromes de predisposição hereditária ao câncer e avaliação de variantes acionáveis que respaldam a prescrição de terapias dirigidas.

Palavras-chave: NGS; variantes genéticas; oncologia personalizada.

ABSTRACT

DNA sequencing methodologies have gone through substantial transformation in the last two decades with the advent of Next Generation Sequencing (NGS). This method revolutionized scientific research and clinical practice in the area of genomics and genetics. The technique is robust, allowing to obtain the sequence of multiple DNA fragments simultaneously, making its execution faster and more accessible. NGS consists of a group of different technologies which all work after nucleotide incorporation into the DNA template strand, allowing the determination of the sequence of interest. There is a variety of analysis that can be done through NGS: whole genome sequencing (WGS), whole exome sequencing (WES), multigene panels and targeted sequencing, which includes amplicon sequencing, i.e small genomic regions amplified by PCR (targeted amplicon sequencing). High-coverage amplicon-based sequencing permits a more directed analysis and its implementation has been expanding more and more, with one of the most important applications being the estimation of ploidy based on genetic variants, both germline and somatic, allowing for the detection of low frequency variants. Therefore, NGS presents important applicability in oncology, both in research and clinics, especially in personalized oncology. Amplicon-based NGS is highly used in the evaluation of known pathogenic variants in patients suspected of hereditary cancer syndromes and evaluation of actionable variants which support the prescription of targeted therapies.

Keywords: NGS; genetic variants; personalized oncology.

SUMÁRIO

| | |
|--|----|
| 1 Introdução..... | 8 |
| 2 Metodologia..... | 9 |
| 3 Sequenciamento de nova geração (NGS)..... | 10 |
| 3.1 Sequenciamento de genoma completo (WGS)..... | 12 |
| 3.2 Sequenciamento de exoma completo (WES)..... | 13 |
| 3.3 Painéis multigênicos..... | 14 |
| 3.4 Sequenciamento alvo específico (TS)..... | 14 |
| 4 Detecção e análise de variantes..... | 16 |
| 5 NGS na oncologia..... | 19 |
| 5.1 Síndromes de predisposição hereditária..... | 20 |
| 5.2 Oncologia personalizada..... | 21 |
| 6 Conclusão..... | 23 |
| 7 Referências Bibliográficas..... | 24 |

1 INTRODUÇÃO

O Projeto Genoma Humano, finalizado em 2002 após 12 anos de pesquisa e o investimento de aproximadamente 3 bilhões de dólares, foi um marco na história da ciência e da humanidade, a partir do qual foi obtido uma primeira referência do “genoma humano normal” (MUZZEY et al., 2015; MORGANTI et al., 2018). Os resultados foram alcançados graças a aplicação do sequenciamento de Sanger, um método considerado da “primeira geração” de sequenciamento de DNA, desenvolvido nos anos 1970 (MUZZEY et al., 2015; SANGER et al., 1977). Desde então, as ferramentas de sequenciamento de ácidos nucleicos avançaram rapidamente em conjunto com o desenvolvimento tecnológico para atender a demanda crescente em volume, especificidade e complexidade nas análises, considerando a limitação do método de Sanger ao sequenciamento de uma única amostra de até 1000 pares de base por leitura (PETERSEN et al., 2017). Como principais vantagens quando comparado ao método de Sanger, o sequenciamento de nova geração (NGS) possibilita o sequenciamento paralelo de várias moléculas de DNA ou RNA, otimizando tempo e recursos, além de possuir maior sensibilidade para a detecção de variantes com baixa frequência alélica (ZHONG et al., 2021). Existe uma variedade de análises que pode ser feita a partir do NGS: sequenciamento de genoma completo (*WGS – whole genome sequencing*), exoma completo (*WES – whole exome sequencing*), painéis multigênicos e sequenciamento alvo específico (*targeted sequencing*) baseado em *amplicons* ou captura híbrida. Há ainda o sequenciamento de RNA, técnica mais recente que vem sendo implementada gradualmente, explorada com mais detalhes por STARK et al., 2019 e HRDLICKOVA et al., 2016.

Na pesquisa científica e na rotina clínica, os principais usos do NGS estão relacionados com a detecção, identificação e classificação de variantes genéticas de origens somáticas (adquiridas durante as divisões celulares dos órgãos e tecidos) e germinativas (presentes nos gametas). Alguns dos primeiros estudos aplicando NGS no campo da oncologia tinham como objetivo compreender o perfil mutacional somático de tumores, esses trabalhos mostraram a necessidade de um entendimento mais aprofundado da genômica do câncer, abrindo espaço para novos estudos mais abrangentes como aqueles baseados em WGS e WES, abordagens que compreendem um número muito maior de regiões do genoma. Por meio desses estudos, foi possível a classificação de mutações somáticas em *driver* e *passenger*, bem como a caracterização de variantes germinativas causadoras das síndromes de predisposição hereditária ao câncer (MARTÍNEZ-JIMÉNEZ et al., 2020; KAMPS et al., 2017). A implementação do NGS na prática clínica é relativamente recente e ainda precisa ser

aprimorada tanto no âmbito ético como técnico, porém já permite avanços importantes no tratamento dos pacientes, especialmente através da medicina personalizada e de precisão - neste trabalho com destaque à área da oncologia -, modelo de atendimento no qual o perfil genético do paciente é considerado na tomada de decisões profiláticas e terapêuticas, de forma a aumentar a eficácia e diminuir os efeitos tóxicos do tratamento, auxiliar na caracterização da doença e identificação de biomarcadores como potenciais alvos terapêuticos, predição mais assertiva de prognóstico e de possíveis mecanismos de resistência (JACKSON, CHESTER, 2014; BERGER, MARDIS, 2018). A preocupação ética na clínica relaciona-se ao respeito e garantia da privacidade do paciente, considerando o armazenamento e compartilhamento de dados e à forma como os resultados são repassados para o paciente e família no que se refere à mutações encontradas por acaso, em especial variantes clinicamente relevantes porém que não são o motivo pelo qual o paciente realizou a análise inicialmente (KAMPS et al., 2017) - existem recomendações pelo *American College of Medical Genetics and Genomics (ACMG)* acerca do repasse da informação para o paciente na publicação “*DNA-based screening and population health: a points to consider statement for programs and sponsoring organizations from the American College of Medical Genetics and Genomics (ACMG)*” (MURRAY et al., 2021).

2 METODOLOGIA

A proposta deste trabalho é uma revisão narrativa da literatura com o objetivo de apresentar o princípio de funcionamento e aplicações do NGS na pesquisa básica e prática clínica. Para isso, foi realizada busca de artigos na plataforma PubMed, seguindo principalmente os termos “*Next generation sequencing*” ou “NGS”, “*variant calling*” e “*precision oncology*”.

3 SEQUENCIAMENTO DE NOVA GERAÇÃO (NGS)

O NGS consiste em um conjunto de tecnologias que permitem o sequenciamento de múltiplos fragmentos de DNA e RNA simultaneamente e funcionam a partir da incorporação de nucleotídeos à fita original, permitindo a determinação da sequência de interesse. Suas principais vantagens estão relacionadas ao baixo custo em relação à grande quantidade de dados que produz em um curto período de tempo e à possibilidade de identificar e quantificar regiões e variantes ainda pouco exploradas, além de permitir a investigação de informações adicionais, como *splicing* alternativo (METZKER, 2009). As aplicações na pesquisa e na clínica são diversas, incluindo testes pré-natais para detecção de aneuploidias cromossômicas, diagnóstico de doenças monogênicas, identificação de biomarcadores que auxiliam no diagnóstico e monitoramento da progressão de doenças e tratamentos, caracterização de células e tecidos (MELDRUM et al., 2011; XUAN et al., 2013), além de ser um pilar essencial na medicina personalizada. A realização de “estudos de associação genômica ampla (*genome wide association studies*, GWAS) iniciou e cresceu gradualmente com a disponibilidade do NGS, elucidando o surgimento e predisposição à doenças complexas de herança multifatorial e/ou condições a partir de suas bases genéticas, entre elas o câncer, asma, esquizofrenia, diabetes, doenças cardíacas, entre outras (PROKOP et al., 2018).

Para a realização do sequenciamento, é necessário seguir uma série de protocolos de preparo prévio que variam de acordo com o método a ser utilizado, porém de forma geral possuem uma matriz similar: preparo do DNA de interesse (incluindo extração e enriquecimento), preparo de biblioteca, sequenciamento e análise dos resultados (ZHONG et al., 2021; HU et al., 2021). O preparo de biblioteca é um passo essencial e consiste na fragmentação mecânica ou enzimática do DNA de acordo com a tecnologia utilizada (*short* ou *long-read*), seleção dos fragmentos e adição de adaptadores, que podem ser *primers*, sequências-alvo específicas já conhecidas pelo equipamento ou *barcodes/indexes*, que permitem a montagem de *pools* de amostras para realização do chamado sequenciamento multiplex (YOHE, THYGARAJAN, 2017; HU et al., 2021). Após o preparo da biblioteca, há ainda uma próxima etapa de enriquecimento que pode variar para cada equipamento, sendo realizado um PCR em emulsão para Ion Torrent e PCR em ponte diretamente no flow cell para Illumina (MELDRUM et al., 2011). A análise dos resultados inclui uma etapa importante de controle de qualidade, que conta com um relatório geral normalmente oferecido pelo próprio *software* do sequenciador (MELDRUM et al., 2011).

O primeiro sequenciador iterativo, cujo método se baseava no pirosequenciamento, foi lançado em 2004 pela 454 Life Sciences (MORGANTI et al., 2018). Atualmente, existem muitas tecnologias de NGS disponíveis no mercado, divididas em dois grandes grupos: sequenciadores de fragmentos curtos (100 a 600 pares de base) e de fragmentos longos (até 900 mil pares de base), sendo considerados respectivamente tecnologias de segunda e terceira geração, sendo que a última não necessita da etapa de amplificação (MORGANTI et al., 2018; ZHONG et al., 2021). Os sequenciadores de segunda geração são os mais amplamente utilizados devido ao menor custo e maior precisão nos resultados e as principais plataformas são Illumina (sequenciadores HiSeq, Miseq, NextSeq e NovaSeq) e Ion Torrent (sequenciadores IonPGM, IonProton e IonS5) (MORGANTI et al., 2018; ZHONG et al., 2021). Ambas as plataformas funcionam por meio de sequenciamento por síntese, ou seja, dependem da DNA polimerase para incorporação dos nucleotídeos (GOODWIN et al., 2016). No entanto, sua principal diferença está diretamente no princípio do método: em Ion Torrent, os dNTPs são liberados individual e sequencialmente, de forma que quando houver a incorporação de um dNTP à fita molde ocorrerá a liberação de um ion H^+ e a alteração de pH resultante será detectada; já para Illumina, um mix contendo os quatro dNTPs marcados com fluoróforos é adicionado de uma vez a cada ciclo e as bases não incorporadas são removidas, assim os dNTPs ligados à fita molde são detectados pelo equipamento ao final do ciclo (GOODWIN et al., 2016). As figuras 1 e 2 ilustram o funcionamento das tecnologias Illumina e Ion, mais tecnologias são apresentadas com detalhes e esquemas de ilustração por METZKER, 2009 e GOODWIN et al., 2016.

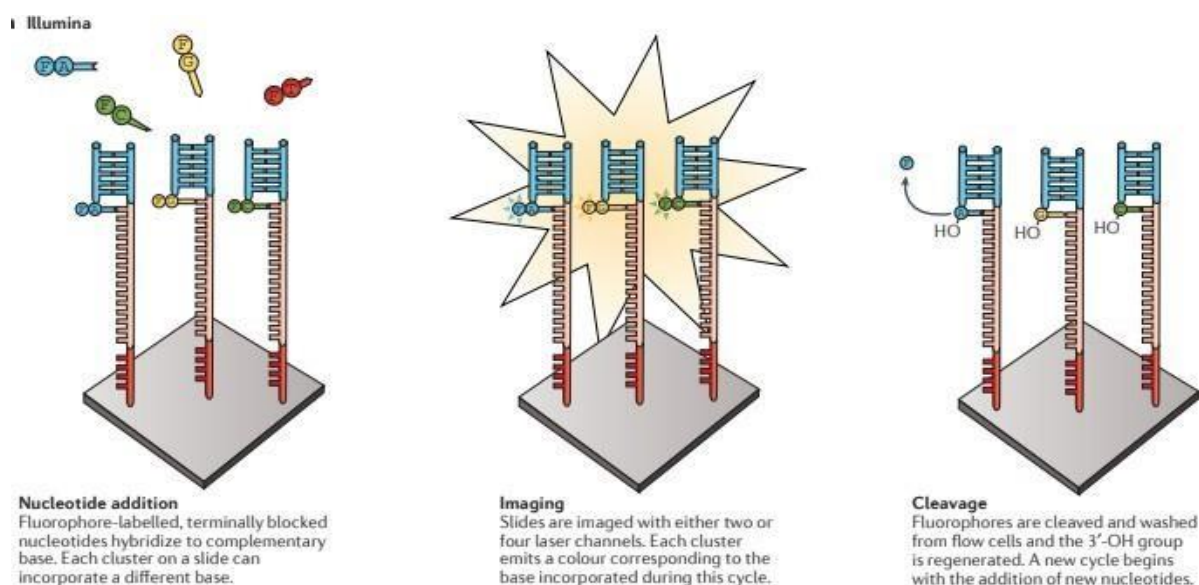


Figura 1. Princípio de funcionamento de sequenciadores Illumina. Adição de mix nucleotídeos marcados com fluoróforos (F). Os nucleotídeos livres são eliminados e os que foram incorporados são detectados pelo equipamento ao final do ciclo; depois, os fluoróforos são removidos e a terminação 3'OH é liberada para alongamento. Fonte: GOODWIN et. al, 2016.

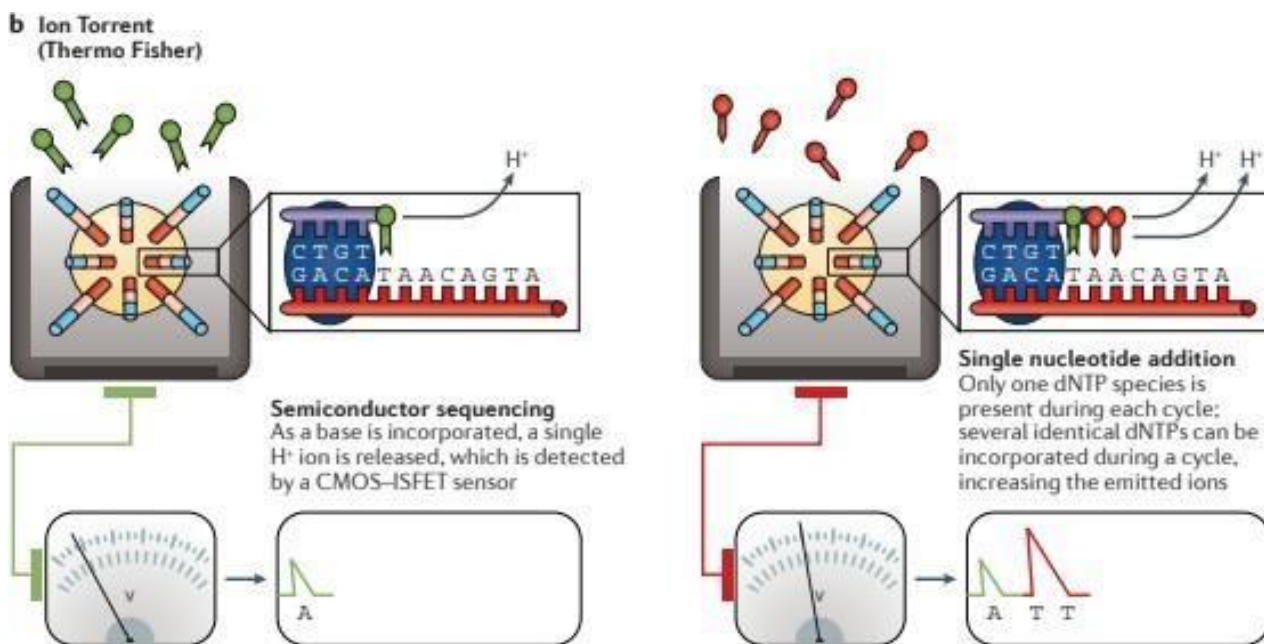


Figura 2. Princípio de funcionamento de sequenciadores Ion Torrent. Pools/amostras são adicionados à placa, que é conectada a um pHmetro; cada bead ocupa um micropoço. Os nucleotídeos são liberados individualmente e, quando são incorporados à fita, um íon H+ é liberado, alterando o pH. Fonte: GOODWIN et. al, 2016.

Alguns dos desafios quanto à aplicação do NGS envolvem o custo-benefício e limitações técnicas relacionadas à sensibilidade dos equipamentos para o sequenciamento de determinadas regiões do genoma, principalmente regiões de homopolímeros que podem gerar artefatos como falsas mutações indel (YOHE, THYGARAJAN, 2017; ZHONG et al., 2021).

3.1 Sequenciamento de genoma completo (WGS)

Dentre as tecnologias disponíveis para WGS, a mais utilizada é a de leitura-curta (*short-read*), devido ao menor custo, alta precisão e facilidade na interpretação dos resultados e uso da ferramenta (LAPPALAINEN et al., 2019). No entanto, não é a mais indicada para este tipo de análise, dificultando por exemplo a detecção de variantes em regiões de homopolímeros e duplicações segmentares, que por estarem presentes em diversas regiões do genoma apresentam maior risco de serem alinhadas erroneamente, considerando que as *short-reads* têm baixa cobertura horizontal e, portanto, a quantidade de bases lida num mesmo

fragmento é pequena (LAPPALAINEN et al., 2019; NAKAGAWA, FUJITA, 2018). Apesar de mais adequadas para WGS, as tecnologias de terceira geração ainda possuem um custo muito elevado, principalmente levando em conta sua alta taxa de erros (NAKAGAWA, FUJITA, 2018).

Atualmente, a maioria dos estudos em oncologia emprega WES em suas análises, porém uma migração ou integração do WGS seria vantajosa, visto a possibilidade de compreender de forma mais completa o processo de carcinogênese pelo estudo de regiões e mecanismos ainda pouco explorados, como as variantes localizadas em regiões não-codificantes, sítios regulatórios, mutações estruturais, grandes inserções e deleções, translocações, inversões, entre outras (NAKAGAWA, FUJITA, 2018). Numa visão mais geral, o WGS permite maior precisão na detecção e caracterização molecular de todos os tipos de variantes devido à sua alta cobertura e possibilita o aprimoramento de genomas de referência, uma demanda crescente para evitar vieses provenientes de genomas mais associados à determinadas populações (LAPPALAINEN et al., 2019; NAKAGAWA, FUJITA, 2018; GOODWIN et al., 2016). Além disso, é também uma ferramenta importante no âmbito da farmacogenômica ao compreender regiões não-codificantes do genoma, como regiões regulatórias e envolvidas no *splicing*, o que não é atendido pelo sequenciamento de exoma (KAMPS et al., 2017). A principal dificuldade na aplicação do WGS relaciona-se com o custo e armazenamento de dados; essa dificuldade estende-se na pesquisa especialmente para estudos populacionais, que podem envolver milhares de sujeitos da pesquisa (LAPPALAINEN et al., 2019).

3.2 Sequenciamento de exoma completo (WES)

O exoma compreende a região codificante do genoma, ou seja, aproximadamente 20.000 genes, dos quais por volta de 6.000 podem ter associação conhecida à doenças (AGGARWAL, 2021). Apresenta cobertura vertical de aproximadamente 100x, o que permite boa sensibilidade em relação à cobertura média de 30x de WGS, sendo suficiente para detecção de variantes em baixa frequência alélica, como em populações subclonais e mosaicismos em linhagens germinativas (NAKAGAWA, FUJITA, 2018; KOBOLDT, 2020). A utilização de WES é mais frequente na pesquisa, como no *Cancer Genome Project*, realizado pelo Instituto Sanger, que reúne cerca de 5 milhões de variantes somáticas relacionadas ao câncer por meio do sequenciamento do exoma de 7.000 tumores primários (KAMPS et al.,

2017). Na clínica, WES é indicado principalmente em casos nos quais uma análise prévia de painel multigênico não apresentou resultados (ZHONG et al., 2021)

3.3 Painéis multigênicos

Os painéis multigênicos configuram uma abordagem direcionada a genes específicos relacionados ao fenótipo de interesse (KOBOLDT, 2020), por exemplo genes de risco associados a determinados tipos de câncer. São bastante utilizados devido ao seu custo-benefício, processamento e interpretação relativamente simples por meio de ferramentas de bioinformática, além de fácil armazenamento de dados (KAMPS et al., 2017). Os painéis estão disponíveis comercialmente em diversos níveis de abrangência e versões direcionadas a diferentes aplicações, podendo incluir centenas de genes (KOBOLDT, 2020). Uma possibilidade interessante é a customização de painéis, que podem ser otimizados de forma a melhor atender o objetivo do estudo/análise por meio da seleção de genes específicos e/ou regiões de interesse como regiões promotoras, *hotspots* ou *breakpoints*, uma vez que os genes selecionados podem ser sequenciados em sua totalidade ou apenas parcialmente (BEWICKE-COPLEY et al., 2019; YOHE, THYGARAJAN, 2017).

A cobertura pode variar de acordo com o tipo de mutação sendo investigada, sendo que uma cobertura vertical de 80x é suficiente para detecção de mutações germinativas, enquanto que para mutações somáticas a quantidade de leituras deve ser maior, com o mínimo por volta de 500x visto que as mutações podem estar presentes em subclones, diminuindo sua frequência (KAMPS et al., 2017). Assim como WES, os painéis possuem boa sensibilidade para detecção de variantes em baixa frequência alélica (KOBOLDT, 2020).

3.4 Sequenciamento alvo-específico (TS)

O sequenciamento alvo-específico (*targeted sequencing*, TS) é o principal método de análise na maioria dos laboratórios devido ao seu custo-benefício, sendo utilizado com maior frequência na investigação de regiões e genes já bem caracterizados, de forma que a detecção de variantes é bastante confiável (ZHONG et al., 2021; BEWICKE-COPLEY et al., 2019). O TS é dividido em dois grupos: sequenciamento por hibridização e captura, e sequenciamento baseado em *amplicons*, cujas técnicas de enriquecimento consistem em PCR e captura híbrida, respectivamente (BEWICKE-COPLEY et al., 2019; MELDRUM et al., 2011). O TS baseado em *amplicons* permite a detecção de mutações pontuais e indels, e a depender do

número de amplicons avaliados a avaliação de CNVs (*copy-number variants*) e até mesmo funções gênicas em RNA, além de fornecer a frequência alélica da variante na amostra, informação importante no monitoramento do tratamento (WADDEN et al., 2022).

A alta cobertura, sensibilidade e baixo custo do TS são suas principais vantagens comparado aos outros métodos, considerando que o material analisado não precisa estar em sua maior quantidade ou qualidade, bem como a possibilidade de detecção de variantes com baixa frequência alélica, com um limite de até 0,1% (BEWICKE-COPLEY et al., 2019).

4 DETECÇÃO E ANÁLISE DE VARIANTES

A detecção de variantes está entre as principais utilidades do NGS na clínica. Esse processo é realizado com o uso de *software* que alinham e comparam os dados de sequenciamento do paciente a um genoma de referência, sendo que a variante, por definição, é quando determinada sequência de bases do DNA do paciente diverge da referência (MUZZEY et al., 2015). Alguns *softwares* frequentemente utilizados em cada etapa da análise estão dispostos na tabela 1. Como mencionado anteriormente, o controle de qualidade é indispensável no diagnóstico de variantes. Avalia-se o índice de qualidade (*quality score*), relacionado com a confiabilidade da leitura, ou seja, com a probabilidade de erro ou acerto do equipamento na detecção de cada base (MELDRUM et al., 2011). A cobertura vertical também está diretamente relacionada com a confiabilidade do resultado: quanto maior a cobertura vertical, maior a sensibilidade do ensaio, assim o número de vezes que determinada base será lida em diferentes moléculas será maior, tornando o resultado mais assertivo (LAPPALAINEN et al., 2019). Se possível, é recomendado que seja feita confirmação da variante por meio de outra plataforma de sequenciamento disponível, principalmente no caso de mutações raras (CHANG, LI, 2013).

Para o diagnóstico, é recomendado que haja, além da amostra de interesse, uma amostra de DNA constitutivo do mesmo paciente (frequentemente sangue periférico ou *swab* bucal), de forma que a classificação da variante em somática ou germinativa torna-se mais confiável (KOBOLDT, 2020; BERGER, MARDIS, 2018).

Após a detecção e confirmação de variantes, o que se segue é uma etapa para caracterização funcional, ou seja, classificação do tipo de variante e identificação dos efeitos no organismo (MELDRUM et al., 2011). No geral, as variantes são classificadas em duas categorias: alterações na sequência de nucleotídeos e variantes estruturais. As alterações na sequência incluem os polimorfismos de nucleotídeo único (SNPs, *single-nucleotide polymorphisms*) e pequenas inserções/deleções (indels), que ocorrem milhares ou milhões de vezes num genoma normal em comparação à referência e, na grande maioria dos casos, não apresentam nenhum efeito deletério para o indivíduo (MUZZEY et al., 2015; LAPPALAINEN et al., 2019). Já as variantes estruturais incluem as CNVs, rearranjos e elementos transponíveis e, apesar de presentes em frequência muito menor que SNPs/indels, comumente apresentam consequências importantes por afetarem regiões maiores (LAPPALAINEN et al., 2019). Quanto à interpretação dos efeitos propriamente ditos, o

profissional pode fazer uso de bancos de dados online, que normalmente apresentam classificação de variantes como preconizado pelo ACMG em provavelmente benignas, benignas, provavelmente patogênicas, patogênicas ou variantes de significado incerto (VUS, *variants of uncertain significance*) (RICHARDS et al., 2015), além disso, existem opções de software que podem ser utilizados como ferramenta para predição do efeito sobre a transcrição de proteínas e o *splicing*, alguns desses software utilizados em laboratórios clínicos são Alamut, Polyphen e Sift (KAMPS et al., 2017) e mais ferramentas são apresentadas na tabela 2.

Tabela 1. Exemplos de ferramentas disponíveis para cada etapa da análise de variantes.

Fonte: KOBOLDT, 2020. (Adaptado)

| Estratégia | Ferramenta |
|--|---|
| Alinhamento e pré-processamento | |
| Alinhamento da leitura | BWA-MEM, Bowtie 2, minimap2, Novoalign |
| Marcação de duplicatas | Picard tools, Sambamba, SAMBLASTER |
| Criação do arquivo BAM | Samtools, GATK |
| Métricas do sequenciamento | BEDTools, Picard tools, QualiMap2 |
| Controle de qualidade da amostra | KING, VerifyBamID |
| Identificação de variantes | |
| SNVs/indels herdadas | FreeBayes, GATK HaplotypeCaller, Platypus, Samtools/BCFtools |
| Variantes somáticas | deepSNV, MuSE, MuTect2, SomaticSniper, Strelka2, VarDict, VarScan2 |
| CNVs | cn.MOPS, CONTRA, CoNVEXM ExomeCNV, ExomeDepth,XHMM |
| Variantes estruturais | DELLY, Lumpy, Manta, Pindel, SVMerge |
| Fusões gênicas (RNA-seq) | fusionCatcher, fusionMap, mapSplice, SOAPfuse, STAR-Fusion, TopHat-Fusion |
| Revisão/armazenamento de variante | |
| Visualização e revisão | Artemis, Integrative Genomics Viewer |
| Manipulação de arquivo VCF/BCF | BCFtools |

BAM binary alignment/map, SNV single nucleotide variant, CNV copy number variant, VCF variant call format, BCF binary variant call format

Tabela 2. Exemplos de ferramentas disponíveis para interpretação e predição do efeito de variantes. Fonte: CHANG, LI, 2013. (Adaptado)

| Descrição | Ferramenta |
|--|--|
| Predição de função da variante | Polyphen-2, SIFT, CHASM |
| Interpretação e priorização | ANNOVAR |
| Catálogo de variantes somáticas no câncer | COSMIC |
| Ferramentas online para visualizar, integrar e analisar a genômica do câncer e dados clínicos associados | UCSC Cancer Genomics Browser |
| Armazenamento de dados de mutações, número de cópias, expressão e metilação provenientes de projetos do TCGA, TARGET, COSMIC, GSK e NCI60; Ferramentas para visualizar genômica a nível amostral e alterações na transcrição em diversos tipos de câncer | Cancer Genome Workbench |
| Recomendações para interpretação de variantes | HGVS (<i>Human Genome Variation Society</i>) |
| Derivação da sequência de codificação | RefSeq database |
| Base de dados de SNPs | dbSNP |
| Base de dados de mutações em genes humanos | HGMD |

5 NGS NA ONCOLOGIA

A introdução do NGS na oncologia permitiu avanços importantes no entendimento do processo de carcinogênese desde sua iniciação, desenvolvimento, até processos subsequentes como a metástase, desenvolvimento de tratamentos personalizados a partir da farmacogenômica, melhor identificação de mutações *driver* e *passenger* e de demais tipos de mutação, suas características estruturais e efeitos na expressão gênica (KAMPS et al., 2017; BERGER, MARDIS, 2018). A partir do conhecimento do perfil genético do paciente e da doença é possível obter uma perspectiva mais assertiva acerca do prognóstico e determinar o tratamento que atenderá da melhor forma as necessidades de cada caso (WADDEN et al., 2022), principais usos do NGS na clínica juntamente ao monitoramento da resposta à intervenção estabelecida e o possível surgimento de mutações *driver*, ou seja, mutações somáticas que conferem alguma vantagem direta ou indireta ao desenvolvimento e proliferação de células tumorais (MORGANTI et al., 2018).

KAMPS et al. (2017) reuniu um extenso compilado de estudos que utilizaram das diversas técnicas de NGS (WGS, WES, painéis e *targeted-sequencing*) em suas análises do perfil somático de diferentes tipos de câncer. No caso de análise somática, uma alta cobertura é essencial para a detecção de variantes presentes nas células neoplásicas/subclones, visto que a frequência pode ser baixa (BERGER, MARDIS, 2018). Atualmente, o método mais utilizado na rotina clínica são os painéis multigênicos, pois são mais rápidos e baratos, mais abrangentes que o TS, possuem cobertura bastante elevada quando comparados à WGS e WES, o que é muito vantajoso em amostras tumorais devido à já mencionada possibilidade de variantes em baixa frequência e são direcionados especificamente às regiões de interesse, evitando a produção desnecessária de um arquivo de interpretação mais complexa e que necessite de muito espaço de armazenamento (LASSEN et al., 2021; BERGER, MARDIS, 2018).

Um desafio que acompanha a aplicação de WGS, WES e painéis multigênicos é a pequena quantidade de material genético passível de extração, em especial para amostras de biópsia ou punção aspirativa por agulha fina, casos nos quais o TS torna-se uma opção bastante viável (BERGER, MARDIS, 2018). Outros fatores a serem considerados incluem a heterogeneidade do tecido, celularidade e grau de pureza devido ao risco de infiltração por células normais, podendo reduzir drasticamente a fração variante esperada (MELDRUM et al., 2011; KOBOLDT, 2020). O armazenamento da amostra também é de grande importância,

idealmente o sequenciamento de DNA proveniente de amostras congeladas é indicado, porém sabe-se que a maioria dos laboratórios/hospitais armazena os tecidos em blocos fixados em formalina e embebidos em parafina (FFPE, *Formalin-Fixed Paraffin-Embedded*), método adequado para análise histológica, porém que pode afetar a qualidade do sequenciamento ao degradar o material - atualmente, a otimização dos métodos de preparo da amostra pré-sequenciamento procura também suprir essa questão, de forma a manter a sensibilidade, especificidade e reprodutibilidade das análises e resultados (BERGER, MARDIS, 2018).

5.1 Síndromes de Predisposição Hereditária

As síndromes de predisposição hereditária ao câncer são definidas pela presença de variantes patogênicas de origem germinativa que aumentam a suscetibilidade a determinados tipos de câncer num indivíduo e em sua família. Estima-se que até 10% dos cânceres tenham um componente germinativo, resultando em mais de 300 síndromes hereditárias caracterizadas; duas das mais estudadas são a Síndrome Hereditária do Câncer de Mama e Ovário (HBOC) e Síndrome de Lynch (LS) (RODRÍGUEZ et al., 2020). Os principais genes associados à HBOC são *BRCA1/2*, descobertos na década de 1990, porém em virtude do NGS foi possível a identificação de mais genes que aumentam a suscetibilidade ao câncer de mama ao mesmo tempo em que estão relacionados às síndromes de Li Fraumeni, Peutz-Jeghers, Cowden e câncer gástrico difuso hereditário, são eles *TP53*, *STK11*, *PTEN* e *CDH1* (VAN MARCKE et al., 2018). Quanto à LS, as principais mutações dizem respeito aos genes *MLH1*, *MSH2*, *MSH6* e *PMS2*, relacionados à correção do mal pareamento de DNA (sistema mais conhecido como *mismatch repair*, MMR), elevando em até 80% o risco de desenvolvimento de câncer colorretal, além do risco aumentado de câncer de endométrio (KAMPS et al., 2017; NGEOW, ENG, 2016). Para pacientes portadores de síndromes hereditárias bem definidas como as mencionadas, existem diretrizes de profilaxia e tratamento estruturadas, que incluem por exemplo a indicação de mastectomia e/ou ooforectomia preventiva para pacientes com HBOC e colectomia total com ileostomia para pacientes com LS, levando em conta particularidades de cada paciente na avaliação e indicação ou não do procedimento cirúrgico (NGEOW, ENG, 2016)

Os painéis multigênicos são atualmente a estratégia mais utilizada no diagnóstico, sua abrangência pode variar mas normalmente são divididos em genes de alta penetrância ($\geq 50\%$ de risco de desenvolvimento de câncer ao longo da vida), penetrância moderada (20 a 50% de risco ou risco aumentado em até 4x comparado à população) e baixa penetrância

(RODRÍGUEZ et al., 2020), conforme demonstrado por KAMPS et al., 2017, ao reunir 90 genes mais utilizados nos painéis disponíveis atualmente. Portanto, uma determinada variante pode estar relacionada a diversos tipos de câncer com diferentes graus de penetrância, podendo configurar por exemplo um fator de alto-risco para câncer de estômago e baixo-risco para câncer de mama; assim, esses indivíduos possuem risco aumentado de desenvolver tumores sincrônicos e metacrônicos ao longo da vida (KAMPS et al., 2017; NGEOW, ENG, 2016). Além disso, variantes germinativas não associadas diretamente com o desenvolvimento de câncer também podem atuar como co-oncogenes ao mediar e interagir com mutações somáticas (QING et al., 2020).

5.2 Oncologia Personalizada

O NGS permite a personalização do tratamento a partir do momento em que se conhece o perfil mutacional do paciente e da própria doença, visto que pacientes com tumores histologicamente parecidos podem ter um perfil mutacional totalmente diferente e, portanto, responderão também de maneira diferente a um determinado tratamento (CHANG, LI, 2013). Da mesma forma, pacientes com diferentes tipos de câncer e perfil mutacional semelhante podem ser submetidos ao mesmo tratamento, utilizando por exemplo inibidores de determinada proteína ou via molecular. Por conta disso, é comum que a elegibilidade de pacientes para estudos se baseie na mutação carregada, e não no tipo de câncer desenvolvido (BERGER, MARDIS, 2018). Por meio das particularidades moleculares de cada caso é possível presumir quais estratégias terapêuticas serão mais adequadas e para quais a doença provavelmente apresentará resistência, escolher de forma mais assertiva os quimioterápicos a serem utilizados de acordo com o alvo (ou biomarcador) específico, resultando numa ação mais eficaz e menos tóxica e consequentemente causando menos efeitos colaterais, além de economizar tempo, recursos e preservar a qualidade de vida do paciente ao evitar tratamentos não-específicos que em muitos casos acabam por não gerar resultados satisfatórios (MORGANTI et al., 2018; KAMPS et al., 2017; LASSEN et al., 2021). Apesar de ideal, a escolha terapêutica com base no perfil mutacional ainda apresenta desafios devido à instabilidade genômica e heterogeneidade do tecido tumoral e à necessidade de se entender de forma mais clara a relação entre genótipo e fenótipo (NGEOW, ENG, 2016). A figura 2 esquematiza a otimização dos resultados quando o perfil genético de cada paciente é levado em conta, e apresenta uma perspectiva para o futuro por Hussen et al. (2022), que consiste na

avaliação individual de resposta à terapia combinada, evitando o desenvolvimento de resistência e ajustando as doses conforme o necessário.

Um método que vem ganhando espaço em vista de sua coleta minimamente invasiva é a biópsia líquida, que consiste no sequenciamento de material proveniente de células tumorais circulantes (CTC) ou DNA tumoral circulante (ctDNA), fragmento circulante derivado do tumor e não associado às células (MORGANTI et al., 2018). Ambos podem ser utilizados na otimização do tratamento, pois atuam como biomarcadores no acompanhamento da resposta à quimio e imunoterapia, detecção de tumor residual e monitoramento da progressão da doença, especialmente ao possibilitar a detecção precoce em pacientes assintomáticos bem como do processo de colonização na metástase e refletir a heterogeneidade do tumor, visto que podem derivar de diferentes locais (KAMPS et al., 2017; TSIMBERIDOU et al., 2020; LASSEN et al., 2021). No entanto, é importante notar que é uma técnica que ainda pode apresentar elevada taxa de falsos-positivos e falsos-negativos dependendo do caso (YOHE, THYGARAJAN, 2017).

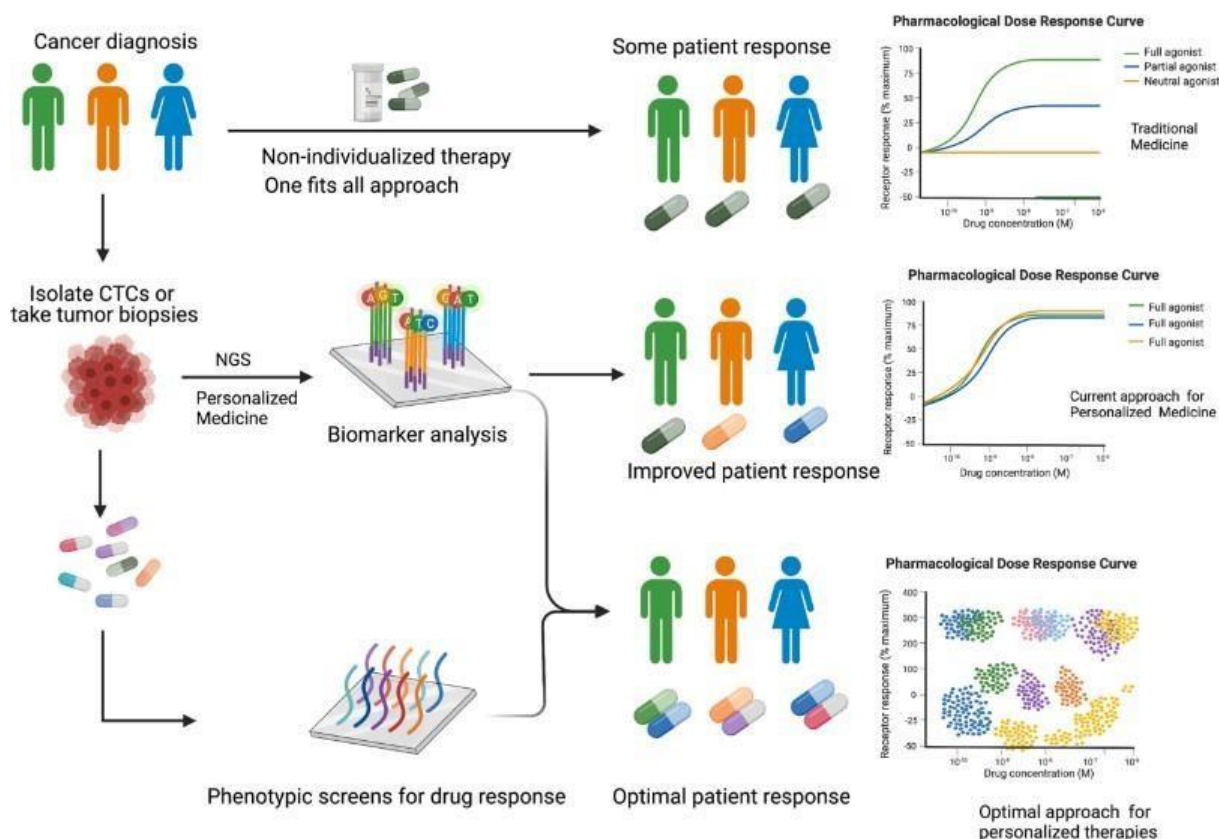


Fig. 2: Perspectivas atuais e futuras de tratamento do câncer. Atualmente, a maioria dos pacientes não recebe tratamento personalizado, porém o cenário tem mudado com a disponibilidade do NGS, aumentando as curvas de resposta aos fármacos. Com a aplicação de terapias combinadas e avaliação individual da resposta será possível aprimorar ainda mais a resposta, evitando desenvolvimento de resistência e ajustando o tratamento conforme necessário. Fonte: HUSSEN et al., 2022.

6 CONCLUSÃO

Nas últimas décadas, o desenvolvimento de tecnologias mais acessíveis e rápidas de sequenciamento genético resultou no aumento exponencial na quantidade de informações disponíveis acerca do genoma humano e até mesmo dos até então inexplorados genomas de outros animais e plantas. Esses conhecimentos, possíveis graças ao investimento em pesquisa científica, revolucionaram a medicina ao possibilitar um entendimento aprofundado de doenças, processos metabólicos, características estruturais e funcionais dos genes, informações que tem utilidade direta na estruturação de novos campos de estudo, como a farmacogenômica, que investe no desenvolvimento de caminhos terapêuticos e quimioterápicos a partir do perfil genético da população de interesse. A medicina personalizada também é consequência direta do NGS. A identificação de genes relacionados à predisposição a determinadas doenças permite o aconselhamento genético para tomada de decisões, aplicação de fármacos em sua maior eficácia, seleção de métodos de profilaxia conforme diretrizes estabelecidas pelos órgãos competentes, diagnósticos precoces que influenciam diretamente na expectativa e qualidade de vida dos pacientes.

Em vista do extenso caminho percorrido pela ciência baseada em evidências, o sequenciamento de nova geração é uma tecnologia bastante recente e que possui ainda muito potencial de aprimoramento. As tecnologias de segunda geração são atualmente as mais utilizadas por diversos fatores discutidos ao longo deste trabalho, porém as tecnologias de terceira geração prometem nos próximos anos trazer possibilidades ainda inexploradas com a redução de custo e tempo de sequenciamento, o que também só será viável a partir do momento que os próprios equipamentos forem aprimorados e se tornarem mais sensíveis, diminuindo sua taxa de erros. As perspectivas e expectativas para o futuro são muitas, como o melhor entendimento de regiões intrônicas, descoberta de novos alvos terapêuticos, desenvolvimento de estratégias epidemiológicas em associação com GWAS, entre outras. Já existem resultados otimistas, como apresentado por WADDEN et al., 2022 no trabalho *“Ultra-rapid somatic variant detection via real-time targeted amplicon sequencing”* ao desenvolver uma técnica de sequenciamento ultra-rápido para utilização durante procedimentos cirúrgicos, por exemplo. Com a disseminação do NGS espera-se uma redução nos custos das análises, de forma que se tornem mais acessíveis e possam levar benefícios para cada vez mais pessoas.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- AGGARWAL, S. Role of whole exome sequencing for unidentified genetic syndromes. **Current Opinion in Obstetrics & Gynecology**, v. Publish Ahead of Print, 18 jan. 2021.
- BERGER, M. F.; MARDIS, E. R. The emerging clinical relevance of genomics in cancer medicine. **Nature Reviews Clinical Oncology**, v. 15, n. 6, p. 353–365, 29 mar. 2018.
- BEWICKE-COPLEY, F. et al. Applications and analysis of targeted genomic sequencing in cancer studies. **Computational and Structural Biotechnology Journal**, v. 17, p. 1348–1359, 7 nov. 2019.
- CHANG, F.; LI, M. M. Clinical application of amplicon-based next-generation sequencing in cancer. **Cancer Genetics**, v. 206, n. 12, p. 413–419, dez. 2013.
- GOODWIN, S.; MCPHERSON, J. D.; MCCOMBIE, W. R. Coming of age: ten years of next-generation sequencing technologies. **Nature reviews. Genetics**, v. 17, n. 6, p. 333–51, 2016.
- HRDLICKOVA, R.; TOLOUE, M.; TIAN, B. RNA-Seq methods for transcriptome analysis. **Wiley Interdisciplinary Reviews: RNA**, v. 8, n. 1, p. e1364, 19 maio 2016.
- HU, T. et al. Next-generation sequencing technologies: An overview. **Human Immunology**, v. 82, n. 11, 19 mar. 2021.
- HUSSEN, B. M. et al. The emerging roles of NGS in clinical oncology and personalized medicine. **Pathology - Research and Practice**, v. 230, p. 153760, 1 fev. 2022.
- JACKSON, S. E.; CHESTER, J. D. Personalised cancer medicine. **International Journal of Cancer**, v. 137, n. 2, p. 262–266, 12 maio 2014.
- KAMPS, R. et al. Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. **International journal of molecular sciences**, v. 18, n. 2, p. 308, 2017.
- KOBOLDT, D. C. Best practices for variant calling in clinical sequencing. **Genome Medicine**, v. 12, n. 1, 26 out. 2020.
- LAPPALAINEN, T. et al. Genomic Analysis in the Age of Human Genome Sequencing. **Cell**, v. 177, n. 1, p. 70–84, mar. 2019.
- LASSEN, U. N. et al. Precision oncology: a clinical and patient perspective. **Future Oncology**, v. 17, n. 30, p. 3995–4009, out. 2021.
- MARTÍNEZ-JIMÉNEZ, F. et al. A compendium of mutational cancer driver genes. **Nature Reviews Cancer**, v. 20, n. 10, p. 555–572, 1 out. 2020.
- MELDRUM, C.; DOYLE, M. A.; TOTHILL, R. W. Next-generation sequencing for cancer diagnostics: a practical perspective. **The Clinical Biochemist. Reviews**, v. 32, n. 4, p. 177–195, 1 nov. 2011.

METZKER, M. L. Sequencing technologies — the next generation. **Nature Reviews Genetics**, v. 11, n. 1, p. 31–46, 8 dez. 2009.

MORGANTI, S. et al. Complexity of genome sequencing and reporting: Next generation sequencing (NGS) technologies and implementation of precision medicine in real life. **Critical Reviews in Oncology/Hematology**, v. 133, p. 171–182, jan. 2019.

MURRAY, M. F. et al. DNA-based screening and population health: a points to consider statement for programs and sponsoring organizations from the American College of Medical Genetics and Genomics (ACMG). **Genetics in Medicine**, v. 23, n. 6, p. 989–995, jun. 2021.

MUZZEY, D.; EVANS, E. A.; LIEBER, C. Understanding the Basics of NGS: From Mechanism to Variant Calling. **Current Genetic Medicine Reports**, v. 3, n. 4, p. 158–165, 4 set. 2015.

NAKAGAWA, H.; FUJITA, M. Whole genome sequencing analysis for cancer genomics and precision medicine. **Cancer Science**, v. 109, n. 3, p. 513–522, 26 fev. 2018.

NGEOW, J.; ENG, C. Precision medicine in heritable cancer: when somatic tumour testing and germline mutations meet. **npj Genomic Medicine**, v. 1, n. 1, 13 jan. 2016.

PETERSEN, B.-S. et al. Opportunities and challenges of whole-genome and -exome sequencing. **BMC Genetics**, v. 18, n. 1, 14 fev. 2017.

PROKOP, J. W. et al. Genome sequencing in the clinic: the past, present, and future of genomic medicine. **Physiological Genomics**, v. 50, n. 8, p. 563–579, 1 ago. 2018.

QING, T. et al. Germline variant burden in cancer genes correlates with age at diagnosis and somatic mutation burden. **Nature Communications**, v. 11, n. 1, 15 maio 2020.

RICHARDS, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. **Genetics in medicine : official journal of the American College of Medical Genetics**, v. 17, n. 5, p. 405–24, 2015.

RODRÍGUEZ, N. et al. Genomic profiling in oncology clinical practice. **Clinical and Translational Oncology**, v. 22, n. 9, p. 1430–1439, 24 jan. 2020.

SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences**, v. 74, n. 12, p. 5463–5467, 1 dez. 1977.

STARK, R.; GRZELAK, M.; HADFIELD, J. RNA sequencing: the Teenage Years. **Nature Reviews Genetics**, v. 20, n. 11, p. 631–656, 24 jul. 2019.

TSIMBERIDOU, A. M. et al. Review of precision cancer medicine: Evolution of the treatment paradigm. **Cancer Treatment Reviews**, v. 86, p. 102019, jun. 2020

VAN MARCKE, C. et al. Prevalence of pathogenic variants and variants of unknown significance in patients at high risk of breast cancer: A systematic review and meta-analysis of gene-panel data. **Critical Reviews in Oncology/Hematology**, v. 132, p. 138–144, dez. 2018.

WADDEN, J. et al. Ultra-rapid somatic variant detection via real-time targeted amplicon sequencing. **Communications Biology**, v. 5, n. 1, p. 1–12, 15 jul. 2022.

XUAN, J. et al. Next-generation sequencing in the clinic: promises and challenges. **Cancer Letters**, v. 340, n. 2, p. 284–295, 1 nov. 2013.

YOHE, S.; THYAGARAJAN, B. Review of Clinical Next-Generation Sequencing. **Archives of Pathology & Laboratory Medicine**, v. 141, n. 11, p. 1544–1557, nov. 2017.

ZHONG, Y. et al. Application of Next Generation Sequencing in Laboratory Medicine. **Annals of Laboratory Medicine**, v. 41, n. 1, p. 25–43, 1 jan. 2021.