

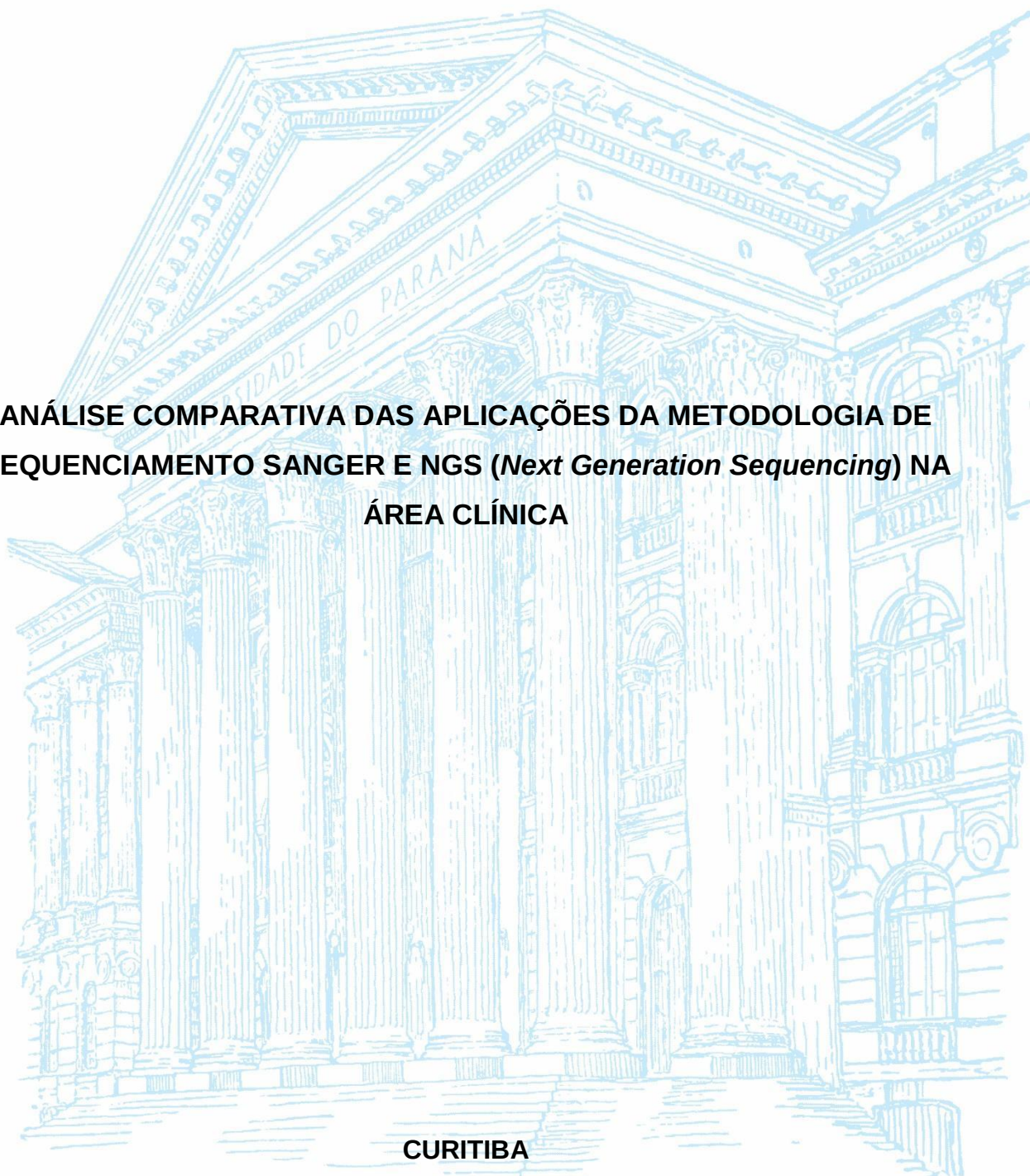
UNIVERSIDADE FEDERAL DO PARANÁ

LUCAS RAFAEL DA SILVA

**ANÁLISE COMPARATIVA DAS APLICAÇÕES DA METODOLOGIA DE
SEQUENCIAMENTO SANGER E NGS (*Next Generation Sequencing*) NA
ÁREA CLÍNICA**

CURITIBA

2019



LUCAS RAFAEL DA SILVA

**ANÁLISE COMPARATIVA DAS APLICAÇÕES DA METODOLOGIA DE
SEQUENCIAMENTO SANGER E NGS (*Next Generation Sequencing*) NA
ÁREA CLÍNICA**

Monografia apresentada como requisito
parcial à conclusão do Ciências Biológicas -
Bacharel, Setor de Ciências Biológicas,
Universidade Federal do Paraná.

Orientador: Prof. Maria da Graça Bicalho
Co-orientador: Marcelo Irajá Mion

RESUMO

As técnicas de sequenciamento de DNA vêm se popularizando cada vez mais na área de análises clínica. Anteriormente, muito trabalhosas e caras essas técnicas ficavam restritas a grandes laboratórios de alto poder aquisitivo. Com o desenvolver da tecnologia, as metodologias de sequenciamento vêm ficando cada vez mais acessíveis. Uma das principais técnicas de sequenciamento é a de Sanger, a qual se baseia na inclusão de dideoxinucleotídeos marcados, que são adicionados ao final de cada fragmento de maneira aleatória. Uma das grandes evoluções no quesito sequenciamento, foi o advento das plataformas conhecidas como NGS (*next generation sequencing*) que trazem como inovação a grande capacidade de sequenciamento de longas extensões de nucleotídeos com milhares de pares de bases no genoma ou em genes de interesse. Este trabalho analisou as principais aplicações das técnicas no contexto clínico e buscou realizar um comparativo entre ambas as técnicas tomando como base a percepção de usuários. Como forma de averiguar a real percepção de usuários que já tiveram experiências com ambas técnicas foi aplicado um questionário abordando todos os pontos envolvidos, ou seja, parte técnica de bancada, de manuseio de sequenciador e aspectos analíticos e econômicos. Ao final do estudo pode-se constatar uma grande superioridade das metodologias NGS em todos os quesitos abordados quando comparados com a metodologia de Sanger.

Palavras chaves: sequenciamento genômico, NGS, Sanger, análises clínicas

ABSTRACT

The DNA sequences techniques are becoming increasingly popular in the field of clinical analysis. Previously very laborious and expensive, these techniques were restricted to large high-income laboratories. With the development of technology, sequencing methodologies have grown increasingly accessible. One of the most important sequencing techniques is that of Sanger, which is based on the inclusion of labeled dideoxynucleotides, randomly added to the end of each fragment. One of the major breakthroughs in the sequencing field was the advent of a platform known as NGS (*Next Generation Sequencing*), which has a remarkable capacity of sequencing long extensions of DNA sequences. This work analyzed the main applications of the techniques in the clinical context and sought to make a comparison between both techniques based on the perception of users. To calculate the real average of users who already had experience with both techniques, a questionnaire was applied addressing issues, such as, the workbench, the sequencer handling and the costs expenses. Our conclusion is that the use of NGS is more accurate for prognosis, and for the identification of a broad spectrum of mutations when compared to the Sanger methodology.

Key words: Genomic sequencing, NGS, Sanger, clinical analysis

SUMÁRIO

1. INTRODUÇÃO.....	5
2. OBJETIVOS	
2.1. Objetivos gerais.....	7
2.2. Objetivos específicos.....	7
3. REVISÃO DE LITERATURA	
3.1. Sequenciamento Sanger.....	8
3.2. Sequenciamento NGS.....	13
3.2.1. Plataforma 454.....	13
3.2.2. Plataforma Illumina.....	17
3.2.3. Plataforma SOLiD.....	21
3.2.4. Plataforma Ion.....	22
3.3. Aplicações na microbiologia.....	25
3.4. Aplicações na virologia.....	27
3.5. Aplicações na histocompatibilidade.....	29
3.6. Aplicações na Oncologia.....	32
4. METODOLOGIAS	
4.1. Revisão de literatura.....	35
4.2. Entrevistas.....	35
4.3. Análises de dados.....	35
5. RESULTADOS E DISCUSSÕES.....	36
6. CONCLUSÕES.....	45
7. REFERÊNCIAS.....	46
8. ANEXOS.....	56

1. INTRODUÇÃO

As técnicas de sequenciamento de DNA têm como principal objetivo determinar a sequência dos nucleotídeos Adenina (A), Timina (T), Citosina (C) e Guanina (G) de um gene ou outra região do genoma. Um dos grandes marcos no ramo de sequenciamento de DNA foi a publicação da sequência da molécula de insulina, que foi determinada no início da década de 1950 pela técnica de Sanger [1]. No início, as técnicas de sequenciamento costumavam ser muito caras e trabalhosas, e isso muitas vezes acabava por limitar sua utilização a grandes laboratórios com alto poder aquisitivo [2].

Como grande inovação, a técnica de sequenciamento de nova geração (do inglês *Next-Generation Sequencing* - NGS) constitui uma estratégia de combinação de preparação de *template*/biblioteca, sequenciamento e geração de imagem, além de alinhamento e montagem de um genoma e/ou genes específicos [3]. Esta técnica reduziu consideravelmente os custos, possibilitando sua popularização e, permitindo que muitos laboratórios pudessem ter acesso a essa nova metodologia. Além desses aspectos, uma grande inovação nos procedimentos de sequenciamento genômico foi a disponibilidade de plataformas multiplex, oportunizando o sequenciamento de vários genes com apenas uma amplificação por PCR, ou seja, otimizando o tempo prático de execução da técnica [4].

Dentro das análises clínicas o emprego do NGS vem permitindo a produção e liberação de diagnósticos mais confiáveis e seguros, além da redução do tempo da análise de resultados. A biologia do câncer teve uma grande evolução com o advento desta técnica, possibilitando a identificação de muitos oncogenes e de genes supressores tumorais [5]. Outro importante emprego desta metodologia, são os testes clínicos de detecção de doenças infecciosas, atuando principalmente no sequenciamento de RNAr bacteriano, além de contar com protocolos já bem estabelecidos para sequenciamento viral [6,7]. A metodologia NGS também vem sendo aplicada nos testes de histocompatibilidade, tendo grande importância nos

processos de busca por doadores compatíveis tanto para doações de órgãos sólidos como também para medula óssea [8].

2. OBJETIVOS

2.1. Objetivos gerais

Estabelecer um comparativo entre a metodologia de Sanger e a técnica inovadora de NGS nas suas diversas aplicações.

2.2. Objetivos específicos

2.2.1. Avaliar a percepção dos usuários quanto às dificuldades apresentadas no processo de amplificação nas metodologias de Sanger e NGS.

2.2.2. Avaliar a percepção dos usuários quanto às dificuldades apresentadas no processo de reações de sequenciamento na metodologia Sanger e de preparo da biblioteca em NGS.

2.2.3. Avaliar a percepção dos usuários quanto às dificuldades apresentadas no processo de manuseio dos sequenciadores utilizados na metodologia de Sanger e NGS.

2.2.4. Avaliar a percepção dos usuários quanto a economia de insumos apresentados na metodologia de NGS quando comparado com a metodologia de Sanger.

2.2.5. Avaliar a percepção dos usuários quanto a economia de tempo apresentado na metodologia de NGS quando comparado com a metodologia de Sanger.

2.2.6. Avaliar a percepção dos usuários quanto às facilidades apresentadas no processo de análise de resultados obtidos a partir da metodologia de Sanger e da metodologia de NGS.

2.2.7. Avaliar a percepção dos usuários quanto a melhoria no número de ambiguidades geradas e resolução nas metodologias de NGS quando comparada com a metodologia de Sanger.

2.2.8. Avaliar a percepção dos usuários quanto a melhoria geral na metodologia de NGS quando comparada com a metodologia de Sanger.

3. REVISÃO LITERÁRIA

As técnicas de sequenciamento genômico, visam estabelecer as sequências em que os nucleotídeos do DNA e/ou RNA.

Desde o início do sequenciamento, no final da década de 70, as técnicas foram passando por inúmeras adaptações e evolução. [9]. As primeiras técnicas apenas permitiam o sequenciamento de pequenos fragmentos (na casa dos kilobases) além de serem pouco acessíveis devido ao preço elevado e dificuldades técnicas. Em contraposição, as técnicas atuais permitem o sequenciamento de genomas completos com baixo custo e com dificuldade infinitamente menor [1].

O sequenciamento do DNA tem sido utilizado em inúmeras aplicações, tais como: obtenção da expressão gênica, caracterização evolutiva de uma dada população ou comunidade, a estruturação e função de um gene alvo, aplicações na Biologia forense, além do foco deste trabalho que é a aplicação dentro das análises clínicas [10].

Abaixo estão descritas as principais técnicas utilizadas para o sequenciamento de DNA ao redor do mundo, desde aquelas com baixo poder de resolução (com baixa capacidade de sequenciamento) e as técnicas mais modernas, que permitem um sequenciamento em grande escala.

3.1 Sequenciamento de Sanger

O método de Sanger é um procedimento amplamente utilizado para sequenciamento de DNA, e que foi elaborado pelo bioquímico britânico Frederick Sanger e colaboradores na década de 70.

Esta técnica se baseia no princípio da marcação radioativa de fragmentos de DNA que são sintetizados a partir de uma fita filha [11]. Devido a esse aspecto de marcação, a técnica de Sanger somente foi possível devido ao desenvolvimento da PCR (Reação em Cadeia da Polimerase) por Mullis em 1986 [12]. A PCR consiste

na síntese de novas fitas de DNA *in vitro* e se utiliza basicamente de um DNA molde, da enzima Polimerase, *primers*, os desoxinucleotídeos além de tampões que irão permitir que a reação ocorra, e que a enzima polimerase desempenhe, de forma correta, sua função.

Em síntese, a amostra a ser sequenciada é combinada com os *primers*, a enzima Polimerase e os nucleotídeos formadores da dupla cadeia (aATP, aTTP, dGTP e dCTP). Além disso são marcados os nucleotídeos que são considerados "terminadores de cadeia" com corante radioativo e esses também são adicionados e concentração menor que a dos nucleotídeos clássicos [13].

Esta mistura será submetida ao procedimento de PCR que consiste em um primeiro aquecimento da amostra para que o DNA seja desnaturado e, posteriormente resfriado. A partir desse momento, haverá a anelação dos *primers* na fita de DNA molde. Após o processo de anelamento, a temperatura é aumentada para que a enzima polimerase possa realizar sua função em condições ótimas e rendimento máximo. A DNA polimerase continuará a adicionar nucleotídeos à cadeia até que um dideoxinucleotídeo seja adicionado ao invés de um clássico. Quando isto ocorre, nucleotídeos clássicos não poderão mais ser adicionados, resultando na finalização da fita com a presença de um dideoxinucleotídeo previamente marcado radioativamente.

O procedimento anteriormente mencionado, irá ocorrer ao longo de vários ciclos. Dessa forma, admite-se que ao final de todos os ciclos, um dideoxinucleotídeo terá sido incorporado em todas as posições do DNA alvo em pelo menos uma reação, apresentando a amostra diferentes comprimentos de fragmentos contendo dideoxinucleotídeos [14]. As extremidades dos fragmentos serão marcadas com corantes capazes de indicar o último nucleotídeo corado, a partir do DNA original. Ao final, por meio dos fragmentos marcados com corantes é possível identificar quais os últimos nucleotídeos adicionados a uma dada sequência [15].

O produto obtido na etapa anterior será submetido a técnica de eletroforese, ou seja, o produto será aplicado em um gel (normalmente de poliacrilamida devido

ao seu alto poder resolutivo) e submetido a uma descarga elétrica, o que determinará que os diferentes tamanhos de fragmentos de DNA se movimentem de maneira distinta no gel formando as bandas [16]. Com isso, será possível visualizar fragmentos que diferem por até mesmo um nucleotídeo. A análise do gel, deverá ser iniciada pelos fragmentos com menor número de nucleotídeos em sua formação, o que permitirá determinar a sequência de nucleotídeo da fita de DNA quando comparada com um padrão conhecido de sequência [17]. O procedimento está simbolicamente representado na

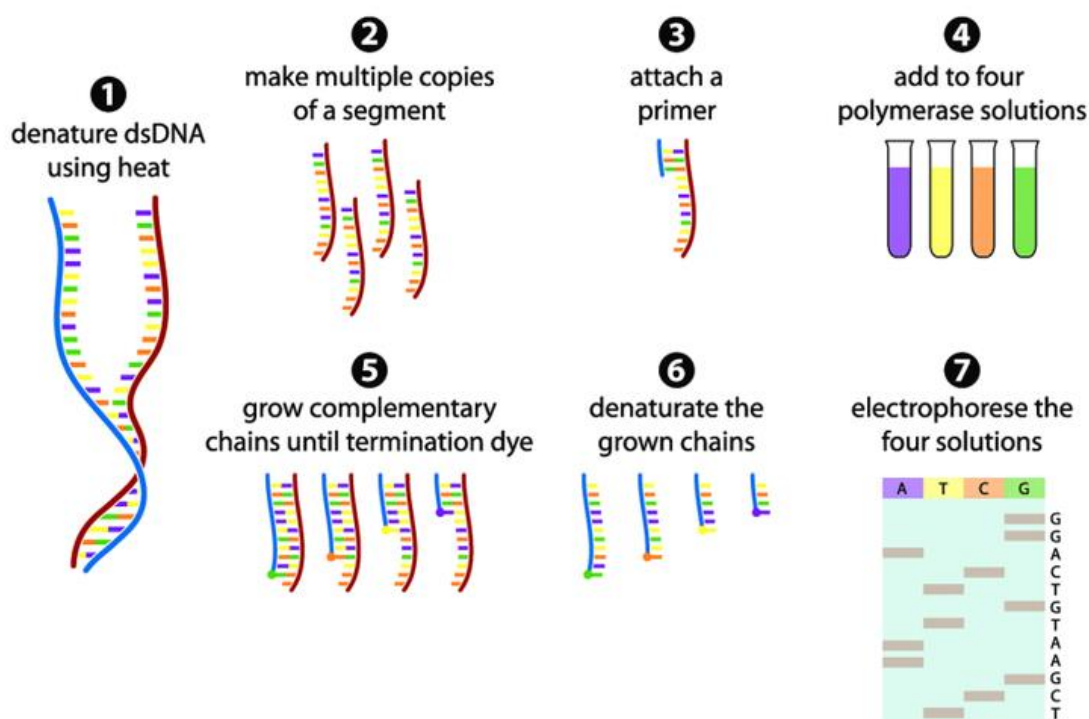


Figura 1- Metodologia de Sanger utilizando gel de poliacrilamida. (1) desnaturação do DNA com alta temperatura para ação da enzima polimerase. (2) Criação de múltiplas cópias do DNA. (3) Anelamento do *primer*. (4) Adição em quatro soluções distintas contendo os dNTPs. (5) Obtenção das cadeias até que um dNTP seja adicionado a cadeia. (6) desnaturação da dupla cadeia. (7) Resultado obtido em gel de poliacrilamida. Fonte: Gauthier, Michel G, 2008 com modificações

A técnica de Sanger se popularizou muito e, atualmente é aplicada em várias áreas da biologia molecular utilizando-se de distintos protocolos. Com o avanço

tecnológico, esta técnica passou por inovações principalmente no quesito automação [18]. Com o advento das técnicas semi-automatizadas e completamente automatizadas, o princípio da técnica se manteve, porém com maior facilidade de execução o que a tornou mais simples além de ocorrer uma redução nos custos, tornando-a mais acessível. No método semi-automatizado, um grande avanço foi decorrente da incorporação de dideoxinucleotídeos com corantes específicos e, com a capacidade de serem excitados quando submetidos a diferentes comprimentos de ondas [19]. Na metodologia inicial, utilizava-se de 4 tubos distintos (uma para cada dideoxinucleotídeo). Com esta nova técnica foi possível realizar toda reação em um único tubo de reação. Assim o produto poderia ser aplicado em apenas uma canaleta de gel, diferentemente do método original, onde a utilização de 4 canaletas do gel, demandava do laboratorista exposição a um produto tóxico por tempo mais elevado [20]. Com isso, foi possível otimizar a técnica a ponto de que a quantidade de amostras analisadas em cada corrida pudesse ser 4 vezes maior. Além disso, com o tempo desenvolveram-se plataformas de sequenciamento, como por exemplo o ABI 377, que tornou possível o sequenciamento de até 48 fragmentos de DNA em um curto espaço de tempo [21-23].

Uma forma alternativa à utilização dos géis poliacrilamida devido ao seu alto poder tóxico, é a realização de uma eletroforese do tipo capilar, o que caracterizou a época de “Automação do Método de Sanger” [24]. Com esta técnica, os fragmentos são submetidos a um longo e fino tubo que estará previamente em contato com uma matriz de gel. Os fragmentos serão detectados no fim do tubo a partir de um laser, que permitirá a detecção do corante e sua correspondente base nitrogenada [25]. Esta técnica se baseia no mesmo princípio de leitura dos fragmentos de DNA No gel poliacrilamida, onde os menores fragmentos passarão pelo laser por primeiro e, serão lidos conforme o aumento no tamanho dos fragmentos. O detector irá fazer o registro das bases constituintes do DNA com os picos de intensidade de fluorescência que é correspondente a cada base [26]. No final, a sequência do DNA original irá ser obtida a partir desses picos de fluorescência registrados de cada última base nitrogenada presente nos fragmentos

que serão lidos, um após outro, pelo laser detector. O esquema de leitura a partir de capilar está esquematizado na Figura 2.

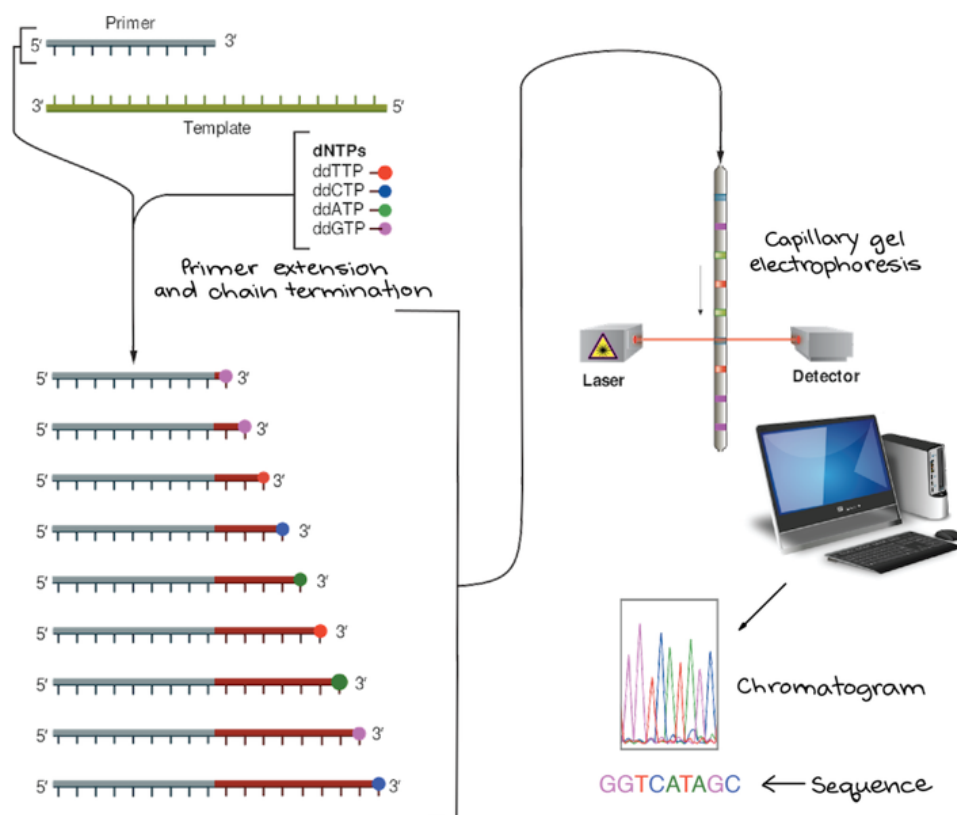


Figura 2- Metodologia de Sanger a partir de eletroforese em capilar. Os fragmentos de DNA serão submetidos a um fino capilar onde a detecção será feita a partir de um laser. Imagem disponível em: <https://pt.khanacademy.org/science/biology/biotech-dna-technology/dna-sequencing-pcr-electrophoresis/a/dna-sequencing>

Toda técnica possui limitações e com a metodologia de Sanger não é diferente. Este método tem uma baixa eficiência, tendo a capacidade de sequenciar fragmentos de até no máximo 900 pb. Isso impossibilita a sua utilização em projetos que requerem o sequenciamento de grandes sequências de DNA [27]. Devido a essa baixa capacidade, este método normalmente é utilizado para o sequenciamento de DNA plasmidial ou, no sequenciamento de genes ou sequências de DNA com poucos pares de bases [28]. Além disso, mesmo com o

processo de semi-automação e automação, continua sendo uma técnica cara de sequenciamento quando comparada com técnicas mais recentes e mais eficientes, isso inviabiliza a utilização por pequenos laboratórios, com baixo poder aquisitivo [29].

3.2. Sequenciamentos de Nova Geração (*Next Generation Sequencing - NGS*)

Após o sequenciamento do genoma humano, surgiram novas plataformas e técnicas referidas como sequenciadores de segunda geração. A utilização destes sequenciadores iniciou-se a partir do ano de 2005 e, desde então essas metodologias vem se consolidando cada vez mais no mercado. Um grande avanço em relação às técnicas tradicionalmente praticadas, foi a sua grande capacidade de sequenciamento de longas extensões do DNA com milhares de pares de bases no genoma ou em genes de interesse [30].

Muitas empresas vêm desenvolvendo sequenciadores para estas técnicas. Uma das pioneiras nesta área foi a Roche que se utilizou da metodologia de pirosequenciamento para desenvolver seu produto. Algumas das principais plataformas de *NGS* serão apresentadas a seguir.

3.2.1. Plataforma 454

Esta plataforma foi a primeira do grupo de sequenciadores de nova geração a ser disponibilizada ao mercado, por volta do ano de 2005. Esta técnica tem como fundamento o princípio do pirosequenciamento que, em termos gerais, tem como princípio a liberação de uma molécula de pirofosfato quando um determinado nucleotídeo é incorporado durante o processo de sequenciamento. Dessa forma, difere das técnicas de Sanger que tem como princípio a incorporação de ddNTPs [31]. A empresa que trouxe essa novidade foi a Roche, que inseriu no mercado o

sequenciador 454 GS20, porém a máquina que realmente se consolidou no mercado utilizando a plataforma 454 foi a GS FLX® (Figura 3) [30].



Figura 3 - Sequenciador 454 GS FLX® fabricado pela empresa Roche

Um grande diferencial quando comparada com as técnicas anteriormente utilizadas, é que não é necessário a clonagem do DNA através da utilização de um vetor, facilitando consideravelmente a execução da técnica. Além disso, houve também um barateamento quando valores são confrontados com as outras técnicas utilizadas [32].

Para esta plataforma, a técnica é executada em basicamente três etapas. A primeira etapa consiste em uma preparação de biblioteca a partir da fragmentação de maneira aleatória do DNA. A seguir, adaptadores são adicionados e ligados, em ambas extremidades dos fragmentos de DNA que permitirão a seleção dos fragmentos de interesse e a formação de uma biblioteca de DNA de fita simples, com fragmentos com tamanhos entre 300 e 800pb (Figura 4) [33].

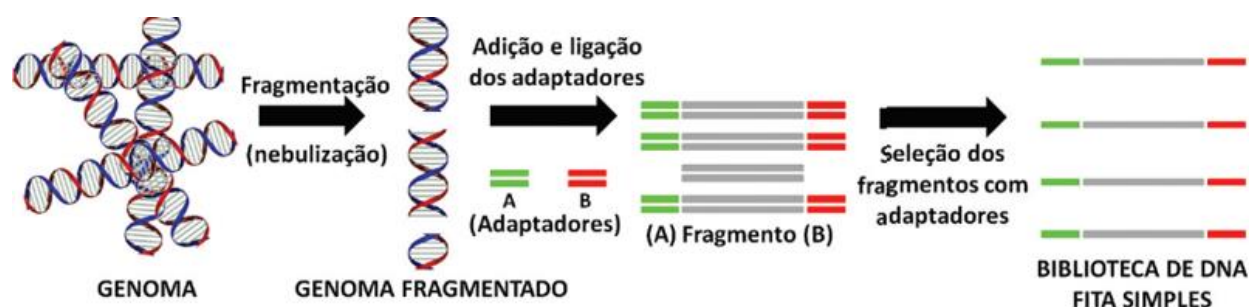


Figura 4 - O genoma é completamente fragmentado e após isso são adicionados adaptadores em ambas a extremidades. Após isso ocorrerá a seleção dos fragmentos de interesse para a formação de uma biblioteca de DNA fita simples. Imagem disponível em Ciências Genômicas [30] - Adaptada

Após a formação da biblioteca, os fragmentos serão ligados e submetidos a *beads* ou esferas magnéticas, que contém sequências complementares específicas a determinados fragmentos, permitindo a hibridização de sequências complementares [34]. Após a captura dos fragmentos pelas *beads*, numa gota oleosa ocorrerá a PCR em emulsão e assim a amplificação de inúmeros fragmentos [35] (Figura 5).

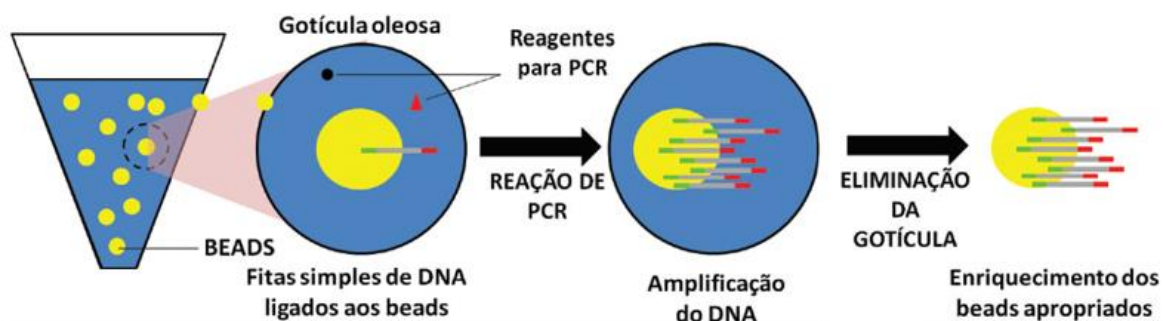


Figura 5 - Os fragmentos da biblioteca são ligados a esferas magnéticas (*beads*). Em uma gota oleosa irá ocorrer a PCR e consequente número de cópias. Imagem disponível em Ciências Genômicas [30] – Adaptada

Depois da obtenção das cópias dos fragmentos, estes serão a técnica de pirosequenciamento. Cada nucleotídeo incorporado às sequências possui em suas extremidades uma molécula de pirofosfato, que irá liberar luz em diferentes comprimentos de ondas, dependendo do nucleotídeo ao qual está acoplada [36]. Cada pirofosfato iluminado gerará uma forma de pirograma, que será traduzido por

um *software* adequado a fim de se obter a sequência nucleotídica presente no DNA inicial (Figura 6).

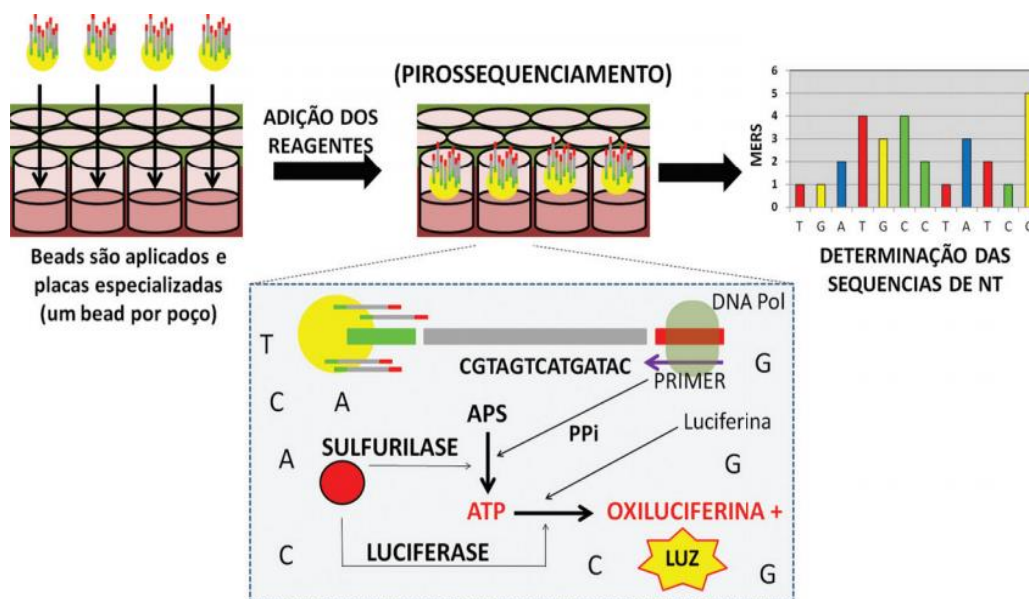


Figura 6- As *beads* se alocam de maneira individual em poços no suporte de sequenciamento. Logo após, ocorrerão as reações de pirosequenciamento. Quando um nucleotídeo é incorporado ocorre a liberação de um pirofosfato e consequente emissão de luz que é registrado, formado um pirograma para cada poço. Após isso os pirogramas serão decodificados por software. Imagem disponível em Ciências Genômicas [30] - Adaptada

Uma importante limitação desta técnica, é a baixa capacidade de detectar a presença de homopolímeros. Esta técnica se baseia na fluorescência emitida durante a incorporação de um nucleotídeo em uma dada sequência, levando a uma repetição da inserção para confirmação da sequência nucleotídica e isso acaba levando a uma imprecisão [37]. A plataforma 454 costuma ter um valor muito mais elevado quando comparada com técnicas mais atuais e com um poder maior de precisão [38]. Além disso, outra limitação é a detecção de sequências provenientes da ligação de dois ou mais fragmentos distintos.

Apesar das restrições, a técnica é muito utilizada no sequenciamento, principalmente de procariotos, os quais possuem um genoma pequeno, onde os erros causados costumam ser menores. Um importante ponto favorável da técnica

é a presença de alto *throughput*, ou seja, há uma alta taxa de rendimento com esta técnica, além de disponibilidade de uma vasta opção de criação de bibliotecas [39].

3.2.2. Plataforma Illumina

A plataforma Illumina é atualmente considerada o “padrão ouro” dentro das análises relacionado às técnicas de NGS. O desenvolvimento da plataforma se deu com a união de importantes empresas dentro do segmento, que são Solexa, Lynx Therapeutics, Manteia Predictive Medicine e Illumina [30]. Possui uma grande versatilidade por utilizarem diferentes equipamentos, como por exemplo MiSeq, NextSeq e HiSeq (Figura 7). Esta técnica popularizou-se por ser de fácil manuseio quando comparada com outras técnicas e, também por seu preço acessível possibilitando a um número maior de laboratórios a utilização das técnicas NGS [38].



Figura 7 - Sequenciadores MiSeq, NextSeq e HiSeq (na sequência)

A plataforma Illumina tem uma semelhança com a metodologia de Sanger, uma vez que em ambas as técnicas utilizam a marcação de nucleotídeos com fluoróforo que serão adicionadas à sequência nucleotídica durante o processo de alongação [40]. A análise de resultados se dá através do registro dessa fluorescência e, a partir da qual haverá a instauração da sequência do DNA de interesse.

A primeira etapa requer a montagem da biblioteca, ou seja, realizar a fragmentação do DNA a ser sequenciado. Após essa fragmentação será necessária

a seleção dos fragmentos viáveis os quais serão acoplados a adaptadores em ambas as extremidades para uma posterior fixação da placa [41] (Figura 8).

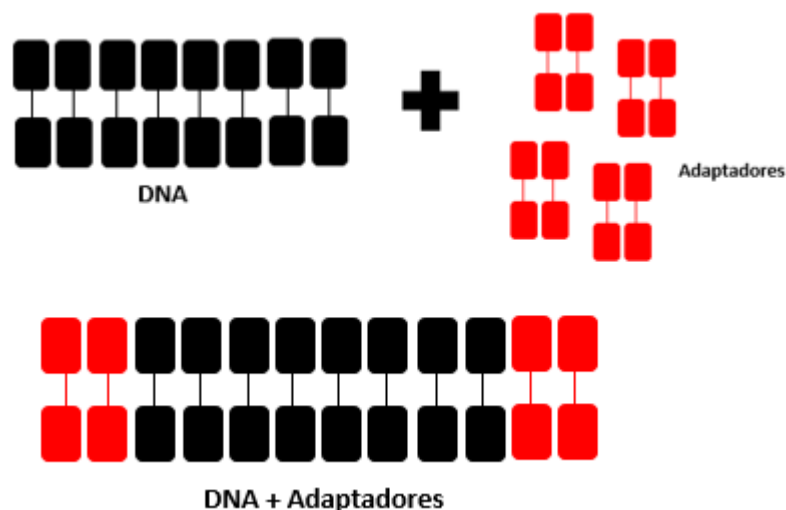


Figura 8 -Esquema da ligação dos fragmentos de DNA constituintes da biblioteca a adaptadores

A etapa de fixação na placa permitirá que os fragmentos do DNA e adaptadores ligados sejam aderidos a uma placa de vidro (*flowcell*) que contém em toda sua superfície outros adaptadores aos quais se conectarão os adaptadores dos fragmentos do DNA. Essa estratégia técnica impede que o fragmento de DNA estabeleça qualquer tipo de interação com a placa [42]. A ligação ao adaptador da placa tem como função permitir o início da extensão de uma nova cadeia nucleotídica com a utilização de polimerases [38]. No início, apenas serão incorporados na cadeia nova nucleotídeos que não possuem marcação de fluorescência. Com esta fita nova já sintetizada, ocorrerá um anelamento com a fita molde formando uma espécie de ponte que caracterizará o processo de amplificação (Figura 9).

Na continuidade da técnica ocorre um aumento da temperatura da placa que contém os fragmentos de DNA fixados pelos adaptadores. Esse aumento de temperatura irá causar uma desnaturação na dupla fita de DNA. Com esta desnaturação rompem-se as pontes de hidrogênio e a consequente linearização das fitas [43]. As fitas lineares de DNA irão ligar-se complementarmente aos

primers, dando início ao segundo ciclo de amplificação. Este processo se repetirá várias vezes, completando aproximadamente 35 ciclos, resultando num número exponencial de cópias do DNA molde, também denominado de *cluster* [44].

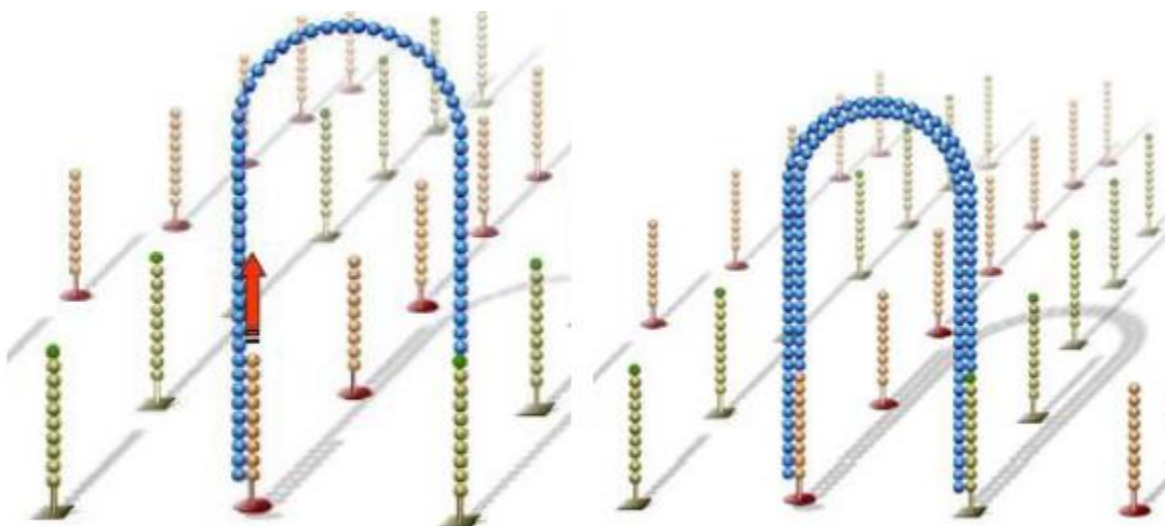


Figura 9 – Formação da ponte e aumento do número das sequências

Após a obtenção desse alto número de cópias de DNA, adiciona-se à placa uma solução que contém 4 nucleotídeos marcados com substância fluorescente. Esses nucleotídeos serão adicionados nas cadeias sintetizadas, a partir da atividade da polimerase em diferentes regiões em cada um dos *clusters*. Na máquina de leitura, haverá um disparo de luz sobre esses *clusters* e, com isso os nucleotídeos marcados irão se excitar e emitir luz, gerando uma imagem desses picos de luminescência [45]. As bases modificadas e que não possuem um grupo hidroxila na posição 3' do carbono do anel de sacarose, funcionam como espécie de “trava”, o que impede a atividade de alongamento da cadeia nucleotídica mediada pela polimerase e determina o término da reação de alongação da cadeia. Após essa primeira etapa de registro dos picos de luminescência, realiza-se uma lavagem com o objetivo de retirar está “trava” da região 3' e com ela o fluoróforo e alguns agentes em excesso, permitindo o início de um novo ciclo. Estes ciclos serão gerados até que um dado fragmento seja completamente sequenciado [46,47] (Figura 10).

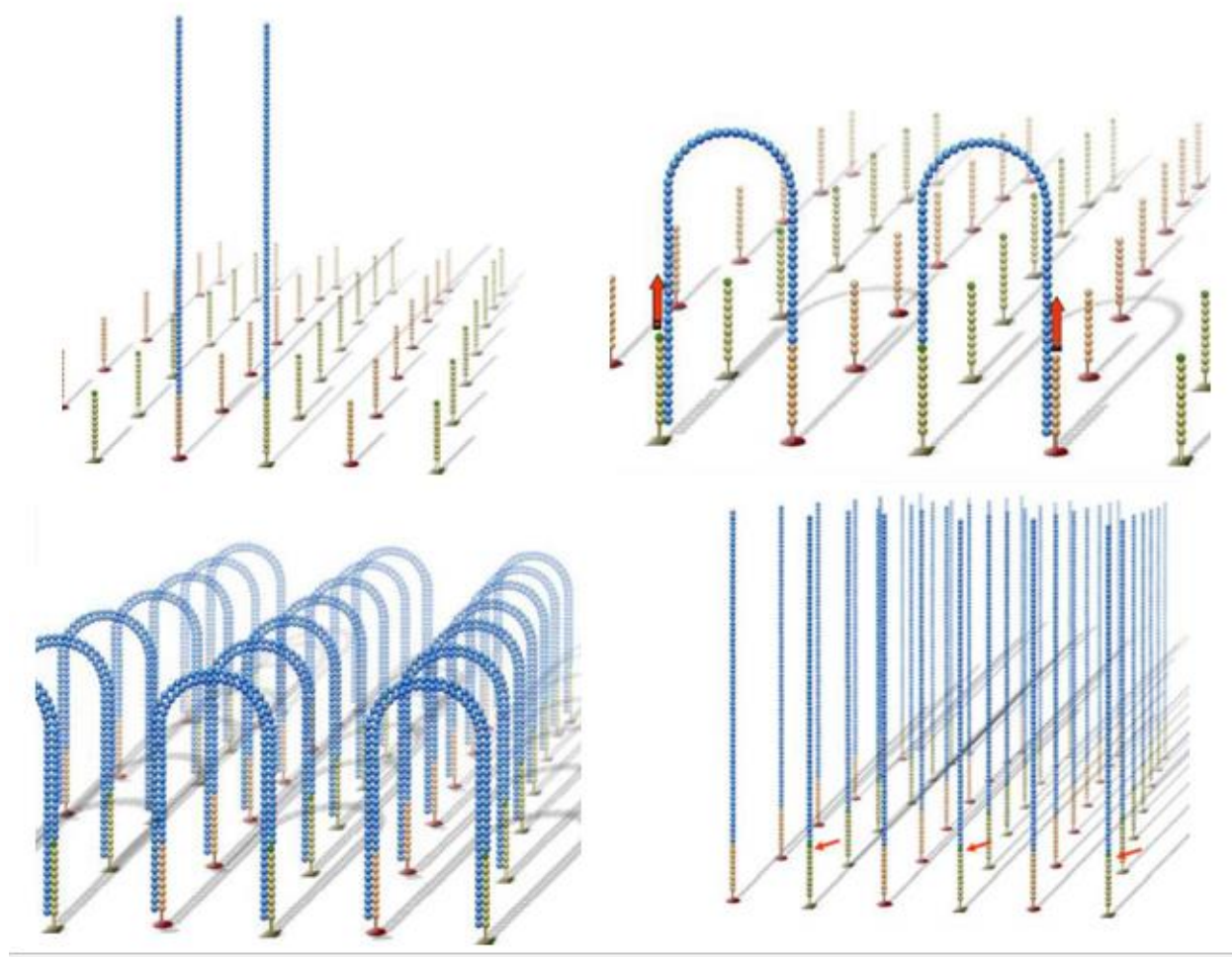


Figura 10 - Acontece a linearização da fita com o aumento da temperatura, e os adaptadores livres se ligam a outros adaptadores na placa dando início a um novo ciclo. Com a formação do cluster haverá milhões de cópias do mesmo fragmento. Em seguida são adicionados os dideoxynucleotídeos (ddNTP's) terminadores que contém o fluoróforo. A enzima polimerase irá incorporar o ddNTP's no fim de cada cadeia. A incidência de luz sobre a placa irá excitar o fluoróforo permitindo a detecção das bases a partir de um software.

Uma grande vantagem desta técnica é que ela permite que a placa de leitura possa ser dividida em várias regiões. Isto possibilitou o desenvolvimento de uma plataforma multiplex, que permite que sequências de diferentes genes possam ser analisadas em uma mesma corrida o que diminui custos e tempo de execução da técnica [38,48,49,50]. Além disso, a plataforma vem se aperfeiçoando e possibilitando o sequenciamento de genomas completos o que desperta interesse e procura pelo mercado consumidor.

3.2.3. Plataforma SOLiD

A plataforma SOLiD que é a abreviação de “*Sequencing by Oligonucleotide Ligation and Detection*”, foi desenvolvida em meados dos anos 2006 pela empresa Applied Biosystems. No entanto, apenas no ano seguinte foi desenvolvido o primeiro sequenciador adaptado para a plataforma SOLiD. Essa primeira máquina ficou conhecida como ABI Solid e, desde então os sequenciadores vem sofrendo atualizações e melhorias em suas estruturas com a nova empresa Thermo Fisher, como é o caso do SOLiD 4 e do sequenciador atualmente considerado “top de linha”, o 5500 XL (Figura 11) [51].



Figura 11 -Sequenciadores que utilizam a plataforma SOLiD

Como nas outras técnicas já mencionadas, o primeiro passo para início do processo de sequenciamento é realizar a preparação da biblioteca. Uma diferença das demais técnicas é que, na plataforma SOLiD existem dois tipos distintos de biblioteca [52]. No tipo “*tag únicas*” os fragmentos preparados são acoplados a adaptadores denominadas P1 e P2, considerados universais. Outro tipo de biblioteca é a de “*tags duplas*” que também é conhecida como “*mate-pair*” onde é possível a fragmentação do DNA em segmentos maiores podendo chegar até 10Kb. Além dos adaptadores inseridos no tipo de biblioteca anterior, nesta, há a inserção de um adaptador no meio da cadeia e cuja função é conectar duas cadeias distintas de DNA [30]. A escolha do tipo de biblioteca a ser utilizada no processo dependerá da finalidade da aplicação.

Nas *beads* utilizadas, há uma sequência específica onde os adaptadores anteriormente adicionados aos fragmentos de DNA irão se encaixar. Após este acoplamento estes complexos serão submetidos a uma PCR de emulsão que resultará num número exponencial de cópias do fragmento original [53].

Uma peculiaridade desta técnica, é a utilização da enzima DNA ligase diferindo das outras duas técnicas de NGS que são embasadas no trabalho da DNA-polimerase [54].

Cinco etapas são necessárias para o sequenciamento, que se baseia na hibridização de sondas fluorescentes com os fragmentos de DNA. Primeiramente ocorrerá a ligação do *primer* ao adaptador P1 que irá servir de suporte para o acoplamento da primeira sonda. Nesta sonda há uma sequência nucleotídica que irá se ligar de forma covalente com os últimos nucleotídeos do *primer* utilizado, esta ligação ocorrerá com a utilização da DNA ligase [30]. Após a formação deste complexo, com a utilização de luz irá ocorrer a excitação da molécula de fluoróforo que se encontra na extremidade da sonda (contém dinucleotídeos) que emitirá fluorescência que será computadorizada em diferentes cores. De forma a eliminar este fluoróforo já excitado, ocorre a clivagem que irá liberá-lo caracterizando o fim do primeiro ciclo [54]. Este processo ocorrerá inúmeras vezes até que todo o DNA molde seja coberto por sondas específicas contendo dinucleotídeos.

O segundo passo compreende a desnaturação das fitas de DNA e o anelamento de um *primer* menor do que o utilizado na primeira etapa. Isto determina que os dinucleotídeos sejam incorporados em regiões distintas daqueles que foram incorporados na primeira etapa. Esta etapa é repetida três vezes (etapas três, quatro e cinco) e, em cada uma delas o tamanho do *primer* inicial é progressivamente diminuído. Um complexo de cores formados a partir da inserção dos dinucleotídeos em diferentes regiões representará a informação contida ao longo do DNA molde [55] (Figura 12).

Após todo o trabalho será necessário realizar uma decodificação dos resultados obtidos. Essa decodificação é realizada utilizando como base as cores de cada nucleotídeo e é realizada a partir do conhecimento da base referente a

posição zero do sequenciamento. Assim, pode-se reconhecer o tipo de nucleotídeo e sua posição na sequência a partir das cores obtidas no resultado.



Figura 12 - Sistema de cores que representa a informação contida ao longo da extensão do DNA inicial que se deseja a sequência.

Como esta plataforma permite sequenciar duas vezes o mesmo nucleotídeo, os erros durante o processamento tendem a diminuir quando comparados com as metodologias anteriormente citadas, sendo esse um dos principais atributos da técnica [52]. Um dos principais problemas enfrentados com esta plataforma, são os autos *throughput* (exige a transferência de muitos arquivos) e isso acaba exigindo uma capacidade computacional muito grande na hora de realizar a montagem de um genoma sequenciado. Além disso, os arquivos gerados são muito pesados e isso exige computadores que tenham capacidade de processar esse número expressivo de dados e informações [56].

3.2.4. Plataforma Ion

Como forma de melhorar ainda mais as técnicas de NGS, a empresa Life Technologies vem aprimorando e atualizando os sequenciadores (Figura 13) [9]. Uma grande novidade no mundo do sequenciamento são os Ion(s), com metodologia fundamentada na capacidade de medir a quantidade de hidrogênio que é liberado na incorporação de uma base nitrogenada, já que cada uma difere na quantidade liberada durante a incorporação [57]. Além disso, essa inovação diminui

consideravelmente o custo do sequenciamento além de contar com técnicas bem mais simples e, em sua maioria automatizada que favorecem sua implementação no mercado [38].



Figura 13 - Sequenciadores da linha Ion Torrent (Ion Chef e Ion S5)

No processamento, as amostras são colocadas em *chips* que contém milhares de nanoporos com medidores de pH que farão a distinção dos nucleotídeos. Nesta plataforma, as bases nitrogenadas vão sendo adicionadas e o ciclo de adição das bases é repetido muitas vezes [58]. Quando ocorre a incorporação correta de uma base na sequência de DNA, a polimerase irá fazer a consolidação desta base e, conseqüentemente, haverá a liberação de um íon H^+ , aumentando de forma imperceptível o pH do meio; porém nos poros essa alteração será percebida e ao final será compilado os dados obtidos em cada poro para se estabelecer a sequência do DNA alvo [59 - 62].

Um dos principais aspectos, que tornam esta metodologia atrativa aos gestores de laboratórios, é o preço dos reagentes e do próprio sequenciador, que são bem mais acessíveis do que os anteriormente citados. Além disso esta técnica tem uma facilidade muito maior do que as outras, e por isso é possível sequenciar longas sequências gênicas e, até mesmo alguns genomas em poucas horas ou em um dia [48]. Porém, assim como a plataforma 454, esta técnica não é adequada para o sequenciamento de regiões homopoliméricas, sendo essa uma das principais desvantagens dessa metodologia.

3.3. Aplicações na Microbiologia

Na área da microbiologia as técnicas de sequenciamento vêm sendo empregadas principalmente para detecção de bactérias causadoras de patologias, fornecendo informações a partir das quais se possa prever tratamento com uma maior eficácia, além de caracterizar possíveis evoluções de quadro dos pacientes. Um outro parâmetro que vem sendo atualizado com o advento das técnicas de sequenciamento e, suas constantes modernizações, são as relações evolutivas entre as bactérias e isso se torna importante principalmente na busca de novos tratamentos e na identificação de colônias. Além disso, os métodos de sequenciamento vêm ajudando na caracterização de novas espécies de bactérias, em especial daquelas que podem ser de interesse de saúde pública.

Um dos genes que desperta interesse clínico em seu sequenciamento é aquele que codifica o rDNA 16S, que é um componente da subunidade 30S dos ribossomos presentes em todos os ribossomos procariontes. Além disso é considerado uma sequência altamente conservada dentro das espécies e, por isso pode ser utilizado para se determinar um laudo durante uma infecção bacteriana, por exemplo. No estudo realizado por Woo e colaboradores [63] foi utilizado o sequenciamento dos primeiros 527pb presentes na sequência do DNAr 16s como forma de separar 37 cepas de bactérias que fenotipicamente eram indistinguíveis. Com o sequenciamento, foi possível identificar a espécie de 30 cepas, sendo que 5 delas foram identificadas erroneamente a nível de gênero e 2 a nível de espécie. O rDNA 16S serviu como base no trabalho de Lau e colaboradores [64] para a realização da identificação da infecção bacteriana contraída por dois pacientes onde os testes apontaram uma ambiguidade do agente infeccioso. No primeiro caso, os testes convencionais de detecção bacteriana apontavam com 98,9% como sendo *Aureobacterium* ou *Corynebacterium aquaticum*. Quando o rDNA 16S da colônia foi extraído e sequenciado o resultado obtido revelou 99,4% de identidade com banco de dados com atribuição de uma terceira espécie, a *Microbacterium oxydans*. No segundo caso do estudo, os testes tradicionais apontavam uma identidade de 99,7% como sendo a causadora da infecção *Cellulomonas* ou *Microbacterium*. Por

meio do sequenciamento foi confirmado que se tratava de uma bactéria do gênero *Mycobacterium*. Em ambos os casos os tratamentos somente poderão ser efetivos após a correta identificação do organismo causador.

Com relação ao gene codificante para o RNA 16S [65], Hummelen e colaboradores em seu estudo verificaram o padrão de bactérias causadoras de vaginose em pacientes portadoras de HIV. Utilizaram a plataforma Illumina e constataram uma diversidade de 90% nas sequências das regiões 16S, representando uma alta diversidade nas bactérias causadoras da vaginose associadas ao HIV. Ao final do estudo, os autores concluem que o perfil das bactérias vaginais em mulheres infectadas com o HIV apresenta uma vasta variabilidade genética, tanto em mulheres com ou sem vaginose. Além disso, observaram que a recorrência de vaginose em mulheres poderia resultar da antibioticoterapia usada no tratamento que, de certa forma acaba favorecendo a instalação de bactérias causadoras da infecção. Métodos de sequenciamento foram fundamentais para estabelecer esses padrões.

Greub e colaboradores [66] tiveram como objetivo sequenciar o genoma completo das bactérias *Parachlamydia acanthamoebae* visando identificar possíveis proteínas imunogênicas. Para este estudo utilizaram a plataforma Illumina. Essa abordagem se constitui numa importante ação profilática para se prever os produtos formados, também para prever medicamentos adequados e que possam combater infecções causadas pela *Parachlamydia acanthamoebae*. Ao total, foram identificadas 18 proteínas com poder de patogenicidade na sequência genômica dessa bactéria. Por se tratar de indivíduos com pouco conhecimento sobre a estrutura genômica e seus potenciais produtos patogênicos, esse trabalho trouxe importantes contribuições para literatura, tornando a identificação dos produtos patogênicos relacionado a bactéria mais fáceis e de identificação do melhor tratamento possível de ser aplicado.

Alguns trabalhos vêm utilizando o sequenciamento de hsp65 para a identificação rápida de micobactérias. No trabalho realizado por Ringuet e colaboradores [67], que utilizaram sequenciamento do DNA pelo método de Sanger, de todas as cepas utilizadas observou-se pouca diferença entre as sequências

nucleotídica das cepas, as quais diferiam em média de 2% entre si. Uma diferença significativa no sequenciamento de *hsp65* foi na diferenciação das espécies *M. chelonae* e *M. abscessus* que possuem uma grande dificuldade de identificação visual e por sequenciamento do rDNA 16S. Um estudo semelhante foi realizado no Brasil por Senna e colaboradores [68] que utilizaram o sequenciamento *hsp65* para realizar a identificação 28 cepas distintas. Esses pesquisadores também concluíram que as sequências das espécies do gênero eram muito semelhantes entre si e que, além disso, apresentavam uma diferença filogenética muito pequena.

No ano de 2005 Bartau e colaboradores, [69] sequenciaram o DNA de sete genes *housekeeping* obtidos de 40 isolados clínicos de *A. baumannii* recuperados de surtos epidemiológicos que acometeram a Espanha e a Alemanha entre 1990 e 2001. Seu objetivo era entender melhor os aspectos relacionados à epidemiologia. Observou-se uma variação entre 6 a 12 nt e um total de 20 perfis alélicos variando entre as cepas de *A. baumannii*. Ao final do estudo os autores concluem que o emprego do sequenciamento é um grande aliado para prever e reconhecer clones resistentes e com potencial epidemiológico além de permitir um monitoramento mais eficiente da disseminação.

Além das aplicações já citadas, as técnicas de sequenciamento têm permitido a caracterização de bactérias resistentes a uma grande variedade de antibióticos [70-73].

3.4. Aplicações na Virologia

Dentro da área de detecção de vírus patogênicos as técnicas para diagnóstico vêm sofrendo uma revolução, passando das técnicas clássicas como cultura celular e anticorpo-antígenos para técnicas mais sensíveis de biologia molecular, como por exemplo a reação em cadeia de polimerase [74]. As técnicas de NGS tem sido importante para a detecção rápida de inúmeros patógenos [75]. A utilização de técnicas de sequenciamento genômico viral, podem ajudar a compreender melhor a variabilidade existente entre os diferentes vírus e a partir

disso ajudar no desenvolvimento de medicamentos e vacinas mais eficientes e eficazes [76, 77].

No estudo realizados por Hang e colaboradores em 2013 [78], os pesquisadores procuraram estabelecer uma técnica de sequenciamento que permitisse determinar o genótipo causador das resistências do vírus da Hepatite B (HBV), devido ao grande impacto na hora do aconselhamento clínico. Para isso, coletaram-se plasmídeos de pacientes e submeteram estes a um tratamento com os principais medicamentos usados para tratamento dessa patologia (Lamivudina, Adefovir, Entecavir, Telbivudina, Tenofovir e Emtricitabina). Como resultado, observou-se que nos 395 plasmídeos coletados, um total de 52 plasmídeos continham uma mutação que determinava resistência ao fármaco tradicionalmente utilizado no tratamento, explicando assim o fato de alguns pacientes não responderem aos fármacos utilizados nos tratamentos tradicionais. Os autores concluem que, o emprego de plataformas de sequenciamento sensíveis o bastante para a detecção de possíveis mutações em vírus que acometem pacientes, é de fundamental importância para um aconselhamento clínico eficaz, já no início do tratamento.

Alguns estudos vêm tentando estabelecer as características genômicas do Papiloma vírus humano, que além do câncer do colo de útero pode acarretar outras doenças tanto em mulheres como também em homens. No trabalho de Barzon. L. e colaboradores [80] testou-se a sensibilidade da plataforma 454 para detecção do HPV. Como resultado, foi possível a detecção de HPV do tipo 16 e 18 em baixa concentração de DNA, além de realizar a detecção de outras diferentes formas de HPV, o que seria difícil por outras técnicas. Concluíram que o NGS pode ser um método muito eficaz na detecção de subtipos de HPV por sua alta sensibilidade. Além disso, o sequenciamento do HPV vem se tornando importante no desenvolvimento de vacinas mais eficazes [81]

Além da caracterização de vírus já conhecidos, o advento das técnicas de sequenciamento e, em especial o NGS, vêm impactando na descoberta de novos vírus causadores de patologias aos humanos. Estima-se que pelo menos 320 mil vírus específicos a mamíferos ainda estão por ser descobertos [77]. Recentemente,

alguns vírus como o Vírus Ebola [82] , Virus Nipah [83], rotavirus Sin Nombre, SARS, e alguns causadores da influenza (H1N1, H7N9) [84] provocaram surtos epidemiológicos e tiveram suas sequências especificadas graças ao NGS e, os pacientes puderam receber um tratamento correto [6].

Um grande estudo que descobriu novos vírus de grande importância patológica humano foi o realizado por Gilda Grard e colaboradores [79]. Em uma aldeia da África com surtos de febre hemorrágica aguda, realizou-se um trabalho de pesquisa utilizando sequenciamento para apurar a causa deste surto. A doença era inicialmente caracterizada como uma febre alta que logo evoluiu para hemorragia nas mucosas e dentro de três dias era capaz de levar o paciente a morte. Como resultado, foi encontrado um novo arbovírus que compartilha apenas 34% do genoma quando comparado com a filogenia do grupo. A caracterização desses novos vírus é de fundamental importância na busca de possíveis vetores e consequente tratamento, que até o momento encontra-se escasso.

3.5. Aplicação na Histocompatibilidade

O complexo principal de histocompatibilidade (MHC) é o principal agente regulador de reconhecimento do que é pertencente ao organismo ou não, ou seja, faz o reconhecimento do próprio do não próprio [85]. Cada série de organismo possui um MHC característico, nos humanos este complexo é conhecido como HLA (*Human Leukocyte Antigen*) cujos genes são os mais polimórficos existentes. Este complexo tem uma fundamental importância nas realizações de testes de compatibilidade que precedem um transplante. Considerando-se a grande variabilidade HLA entre indivíduos, as técnicas de sequenciamento têm auxiliado nesse aspecto por serem capazes de prever com um maior índice de certeza os alelos HLA e, com isso, diminuir os riscos de rejeição no transplante [86].

Os testes de sequenciamento vêm sendo utilizados na identificação de alelos que definem a histocompatibilidade e na identificação de mutações e de novos alelos. No sistema gênico HLA, existe duas classes tradicionais de genes que impactam no prognóstico dos transplantes. Os genes de Classe I, clássicos são

HLA-A, *HLA-B* e *HLA-C* além dos não clássicos *HLA-E*, *HLA-F*, *HLA-G*. Os genes de Classe II é composto pelos loci *HLA-DR*, *HLA-DQ*, *HLA-DP*. Desde a popularização dos métodos de sequenciamento, a detecção de novos alelos se tornou muito mais comum e, com grande impacto na liberação de laudos pois permite uma maior assertividade na determinação alélica e um menor índice na rejeição de órgãos, devido à incompatibilidade HLA. O gráfico abaixo (Figura 14) mostra a evolução na descoberta de novos alelos com a popularização das técnicas de sequenciamento, em especial a partir do ano 2009 com as técnicas de NGS.

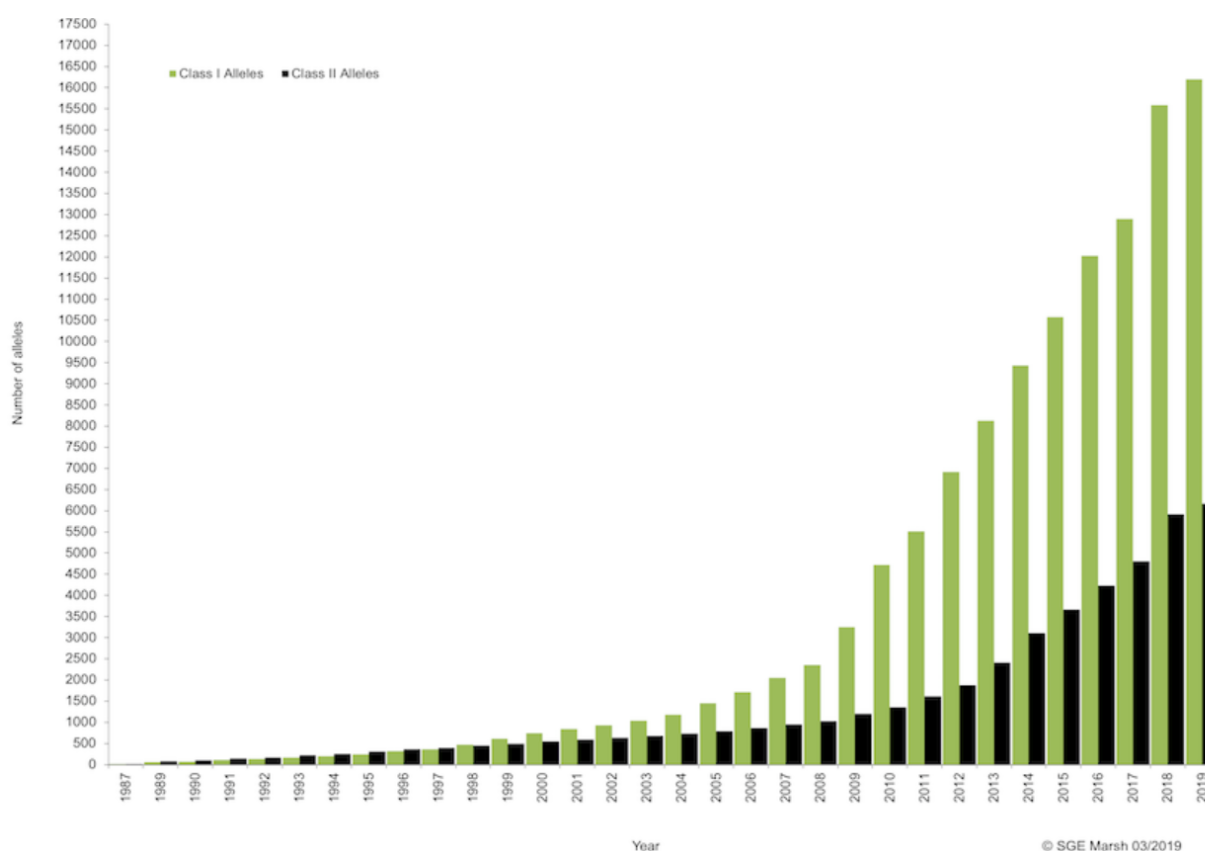


Figura 14 - Evolução no número de alelos HLA classe 1 (barras verdes) e classe 2 (barras pretas) ao longo do tempo. Disponível em: <http://hla.alleles.org/nomenclature/index.html>

Um exemplo do emprego das técnicas de sequenciamento na descoberta de mutações e a consequente geração de novos alelos, é o trabalho de Kuzmich [87]. Neste trabalho foi detectado o alelo *HLA-DQB1*06:210* em uma paciente portadora de leucemia que se preparava para um transplante de medula. Foi utilizado a

metodologia de Sanger, a qual detectou um *mismatch*, ou seja, havia uma substituição de G por T na sequência nucleotídica do gene o que impactava na troca de uma serina por uma alanina (mutação não sinônima) na cadeia polipeptídica correspondente. Apenas com essa modificação foi possível distinguir do alelo DQB1*06:03:01 e definir o novo alelo DQB1*06:210. Dos mesmos autores [88], foi identificado também, com a metodologia da Sanger, um novo alelo HLA-B*44, que foi denominado HLA-B*44:02:45. Este novo alelo detectado em uma doadora russa voluntária para medula óssea. Este novo alelo possui uma substituição nucleotídica na posição 243 do gene onde uma G foi substituída para T, porém não ocorreu nenhuma modificação no aminoácido codificado (mutação sinônima). Mesmo não ocorrendo a mudança de aminoácido na cadeia polipeptídica correspondente, houve uma substituição nucleotídica na sequência codificadora do alelo *HLA-B*44:02:01:01* o que justifica e define o novo alelo HLA-B*44:02:45.

Nos testes tradicionais de tipagem HLA para transplantes, os índices de geração de ambiguidades gerados nos resultados de análises são consideravelmente frequentes. As técnicas de sequenciamento vêm ajudando a resolver tais ambiguidades, exemplo da aplicação de sequenciamento para resolução de ambiguidades é o trabalho desenvolvido por Holcomb e colaboradores [89]. Foram realizados os testes de sequenciamento NGS com a plataforma 454 para resolver ambiguidades encontradas em 20 amostras, sendo tipados os loci de classe 1 clássicos (A, B e C) e os *loci* de classe 2 clássicos (DPB1, DQA1, DQB1, DRB1 e DRB3/4/5). A partir desse sequenciamento, os autores conseguiram resolver cerca de 95,3% das ambiguidades encontradas anteriormente com testes clássicos, demonstrando a eficácia do método.

Outro conjunto de genes que desempenham importante papel para a histocompatibilidade são os genes KIR (*killer cell immunoglobulin-like*) que codificam para receptores encontrados nas membranas das células natural *killer* (NK). Essas moléculas estabelecem conexão com os *HLA* de classe I afim de exercer ou não suas funções de ativar as células NK para que suas ações citolítica ou secretora de citocinas sejam ativadas ou inibidas [90,91]. A presença de alguns genes KIR específicos tem demonstrando uma grande associação com a

suscetibilidade de pacientes a doenças infecciosas e principalmente a doenças autoimunes. Tendo em vista a importância desses genes, o sequenciamento de DNA para tipagem dos alótipos de KIR vem sendo empregados dentro da clínica [92].

Uma alternativa terapêutica que vem sendo utilizada para o tratamento de leucemias e outras doenças técnica é o transplante de células progenitoras hematopoiéticas. Observou-se que tanto as infecções virais no pós-transplante bem como as células leucêmicas tendem a diminuir quando o doador possui haplótipos de genes KIR ativadores, e para isso as técnicas de sequenciamento podem ser aplicadas para seleção desses doadores. Além disso pode ser aplicado para selecionar doadores de células-tronco com maior compatibilidade para pacientes com algumas malignidades. [93,94].

3.6 Aplicação na Oncologia

Na Oncologia, o sequenciamento de DNA torna-se uma ferramenta importante para se entender melhor as anomalias que causam o desenvolvimento de um quadro oncológico, bem como o seu desenvolvimento. Além disso, as metodologias de sequenciamento vêm sendo empregadas para prever a suscetibilidade de um dado paciente a apresentar possíveis predisposição mutações herdadas como, por exemplo, *BRCA1* e *BRCA2* que estão associados ao câncer de mama [95]. Também podem ser empregados para a decisão do tratamento, pois com o sequenciamento é possível prever as via de sinalização que compõe cada célula [96].

Outro método frequentemente utilizado nos testes oncológicos é a imunohistoquímica e, no trabalho de Sjögren [97] buscou-se comparar este método com a técnica de sequenciamento para verificar mutação no gene p53, que é um supressor tumoral. Mutações nesse gene tem sido associada a um estado um pior prognóstico no câncer de mama. Nesse estudo foram utilizadas 316 amostras de DNA de pacientes com tumores primários para a realização de sequenciamento de cDNA e para os testes de imunohistoquímica. A detecção da presença de mutação no p53 por NGS foi mais eficaz e 30% superior ao método de imuno-histoquímica.

Semelhante ao trabalho anterior, Agarwal [98] e colaboradores também confrontaram a técnica de imunohistológica com a de sequenciamento de DNA, para identificar mutações no gene IDH1 em gliomas. Amostras que se mostraram negativas (20%) para a mutação pelos testes imunohistoquímicos foram caracterizadas positivamente pelo sequenciamento.

O avanço das técnicas de sequenciamento, alertou para a complexidade genética envolvida no contexto oncológico, dentro das famílias. Castera e colaboradores [99] buscaram analisar um padrão de mutação em genes em pacientes com câncer de mama e câncer de ovário com suspeita de herança familiar e, para isso utilizaram o sequenciamento de DNA. Foram testados 708 pacientes e constatou-se um padrão de mutação em cerca de 14 genes. De forma semelhante, Tung [100] realizou o sequenciamento do DNA de 1781 pacientes com câncer de mama. Observou que além da mutação nos genes clássicos associado com o câncer de mama e que representaram 9,3% dos pacientes, haviam também 4,2% dos pacientes que continham outras mutações em genes distintos. Em geral os mais observados foram PALB2, CHEK2 e ATM. Um importante estudo realizado no contexto de aconselhamento familiar foi o de Aloraifi e colaboradores [101]. Os autores sequenciaram 312 genes de amostras de 13 pacientes onde encontraram variantes alélicas associadas com a patologia nos genes ATM, RAD50, PALB2, CHEK2 e TP53. Com esse resultado, aumenta o conhecimento sobre genes alvos e, de certa forma ajuda no aconselhamento em famílias que possam conter alterações nesses principais genes e que podem ter acesso a um resultado precoce e rápido, ou seja, beneficiando-se de uma orientação médica adequada.

Além de câncer de mama e ovário, alguns outros tipos de câncer também podem ter uma ligação forte com a hereditariedades, como por exemplo o colorretal, endometrial e ginecológico. Cragum e colaboradores [102] também testaram modelos de painéis de genes e conseguiram detectar mutações associadas com esses tipos de câncer em cerca de 10 % dos indivíduos com câncer colo retal no total de 586 pacientes com histórico familiar. Semelhantemente ao anterior, Fray e colaboradores [103] utilizaram o histórico de 127 pacientes com câncer ginecológico e, que tinham registro de incidências na família. Pela técnica do sequenciamento

analisaram o DNA desses pacientes pois os testes arte então utilizados não revelaram qualquer alteração. Como resultado, o sequenciamento detectou que cerca de 7% dos pacientes eram portadores de uma mutação associada ao câncer ginecológico.

Nos tempos atuais o sequenciamento vem sendo usado para prever os riscos de uma criança vir a desenvolver uma anomalia ligada ao câncer antes mesmo da fertilização. Essas técnicas são conhecidas como diagnóstico pré-implantação [104]. Essas técnicas normalmente utilizam a análise de SNP e as técnicas de NGS podem ser empregadas para analisar tanto os distúrbios de um único gene ou a nível de cromossomo [105].

4. METODOLOGIAS

4.1. Revisão da literatura

Esta revisão foi realizada a partir de artigos obtidos de plataformas como Periódico Capes, Google Acadêmico e pelo *Science Direct*. Os artigos obtidos além de serem bases para a seção “Revisão da Literatura” foram utilizados também na Introdução e Discussão.

4.2. Entrevistas

Afim de se levantar a real percepção de usuários quando comparadas as técnicas de Sanger e NGS, foi realizado um breve questionário (Anexos) onde cada tópico pudesse ser avaliado com notas variando de 1 a 5.

O questionário foi desenvolvido a partir dos principais pontos que são passíveis de servirem como comparação entre as duas tecnologias. Os pontos abordados foram o tempo de bancada, o manuseio dos sequenciadores utilizados (considerado apenas a forma automatizada em ambas técnicas) e ao processo de análise (ambiguidades, resoluções e formas de liberação de resultados).

Como forma de aumentar o número amostral, o questionário foi publicado na rede social profissional *LinkedIn*, onde profissionais dos diversos setores da área de análises clínicas puderam ter acesso, o que facilitou a comparação dos resultados obtidos. Ao total, 50 participantes, que já tiveram algum tipo de experiência com ambas técnicas responderam ao questionário.

4.3. Análise dos resultados obtidos.

A compilação das respostas a partir do questionário, possibilitou o agrupamento e análise dos dados e, no software Excel foram plotados gráficos para melhor visualização dos resultados obtidos.

5. RESULTADOS E DISCUSSÃO

Um total de 50 participantes com experiências com as técnicas de sequenciamento de Sanger e com metodologias de NGS responderam o questionário. Para fins comparativos, alguns dos questionamentos irão ser apresentados adiante em forma e gráficos conjuntos.

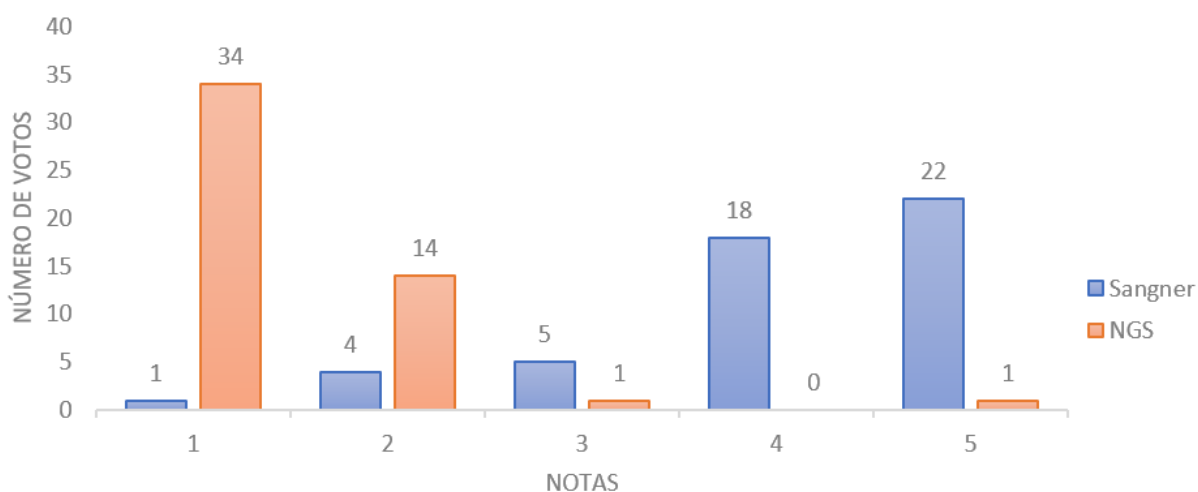
Devido ao fato de que atualmente existem diversas plataformas NGS no mercado, os usuários foram questionados, primeiramente em qual/quais plataformas possuem experiência. As opções dadas inicialmente continham as plataformas anteriormente abordadas (454, Illumina, SOLiD e Ion). Foi também previsto uma opção e espaço para casos onde houvesse experiência com outras plataformas. As plataformas com mais apontamentos entre os usuários que responderam o questionário foram a plataforma Illumina e a Ion, com respectivamente 36 e 28 apontamentos. As plataformas 454 e SOLiD foram as menos apontadas pelos usuários, tendo um total de 8 e 9 apontamentos respectivamente. Além dessas plataformas previamente fornecidas como opções, foram mencionadas também a PacBio com 3 apontamentos e a plataforma Oxford Nanopore com 2.

Do total de 50 usuários que responderam o questionário, 20 (40%) pertenciam a área de Histocompatibilidade, 12 (24%) as áreas de Microbiologia, 8 (16%) Virologia, 6 (12%) Oncologia e, alguns outros usuários em menor quantidade pertencentes a área de Genética Médica (2%) e Genética Forense (2%)

Em relação a dificuldade no processo de amplificação inicial (Gráfico 1) foram atribuídas notas de 1 a 5, onde 1 representa baixa dificuldade durante o processo e 5 uma alta dificuldade. Quando questionados sobre este aspecto na metodologia de Sanger, um total de 1 (2%) usuário marcaram a opção “1”, 4 (8%) marcaram a opção “2”, 5 (10%) marcaram a opção “3”, 18 (36%) usuários marcaram a opção “4” e 22 (44%) marcaram a opção “5”. Já quando questionados sobre a amplificação inicial na metodologia de NGS, um total de 34 (68%) marcaram a opção “1”, 14 (28%) marcaram a opção “2”, 1 (2%) marcaram a opção “3”, 0 (0%) usuários marcaram a

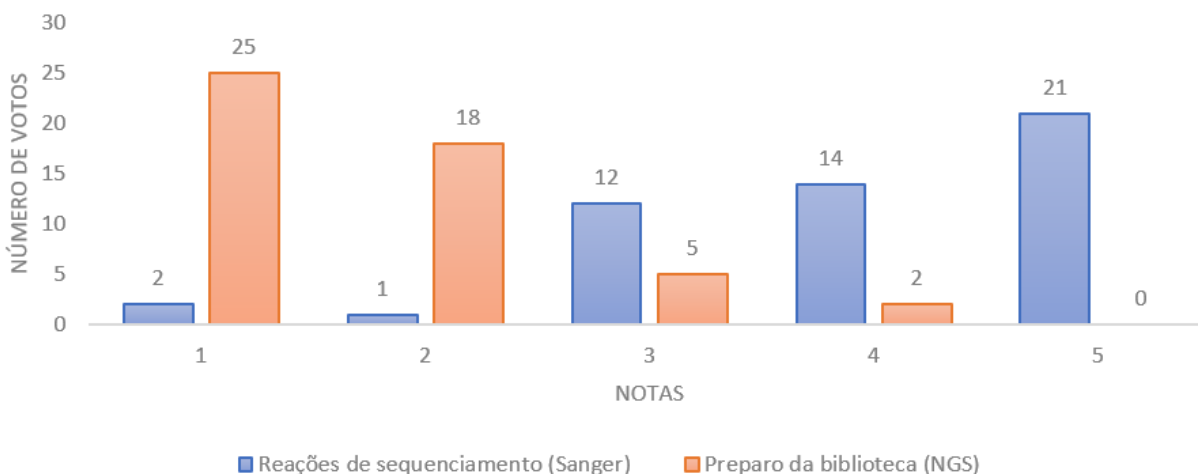
opção “4” e 1 (2%) marcaram a opção “5”. A nota média obtida na metodologia Sanger foi de 4,12 enquanto a média obtida em NGS foi de 1,4.

Gráfico 1. Percepção dos usuários quanto ao nível de dificuldade no processo de amplificação inicial em Sanger e NGS



Em relação a dificuldade no processo de reações de sequenciamento (Gráfico 2) na metodologia Sanger e ao de preparo de biblioteca no NGS foram atribuídas notas de 1 a 5, onde 1 representa baixa dificuldade durante a técnica e 5 uma alta dificuldade. Com relação às reações de sequenciamento em Sanger, um total de 2 (4%) usuário marcaram a opção “1”, 1 (2%) marcaram a opção “2”, 12 (24%) marcaram a opção “3”, 14 (28%) usuários marcaram a opção “4” e 21 (42%) marcaram a opção “5”. Com relação ao procedimento durante a preparação da biblioteca em NGS, um total de 25 (50%) usuário marcaram a opção “1”, 18 (36%) marcaram a opção “2”, 5 (10%) marcaram a opção “3”, 2 (4%) usuários marcaram a opção “4” e 0 (0%) marcaram a opção “5”. A média das notas obtidas no método de Sanger foi de 4,02 e na técnica de NGS foi 1,68.

Gráfico 2. Percepção dos usuários quanto ao nível de dificuldade durante as reações de sequenciamento no método de Sanger e no preparo da biblioteca no NGS



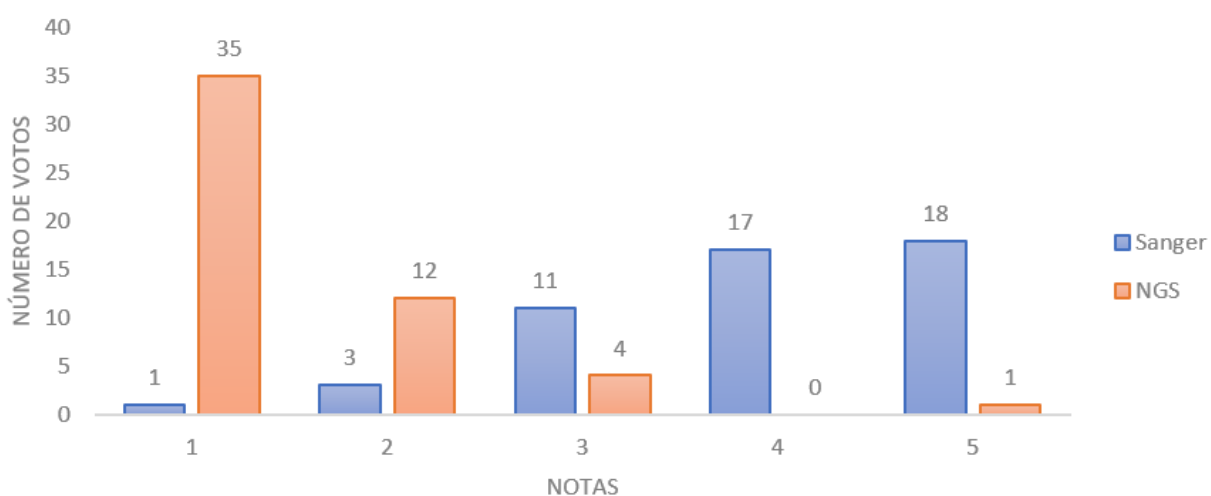
Uma das grandes inovações que as plataformas NGS trazem como novidade é a possibilidade de realizar a tipagem genética de mais de um gene durante uma única corrida. No contexto clínico, isso é de fundamental importância tendo em vista que normalmente há uma necessidade de análise da sequência de diferentes genes para chegar a um resultado coerente.

Esse advento das conhecidas plataformas *multiplex*, acarretam em uma diminuição no tempo total de bancada [106]. Normalmente, o processo de amplificação inicial é otimizado pela utilização de um compilado de primers que são capazes de amplificar diferentes regiões genômicas necessitando apenas de um processo de PCR [107]. Ao contrário disso, a metodologia de Sanger normalmente demanda uma quantidade de reações muito extensas, isso devido ao fato de não poderem contar com amplificações simultâneas, ou seja, cada gene deve ser amplificado de forma independente.

Outro aspecto abordado na pesquisa foi em relação a dificuldade de manuseio do sequenciador (Gráfico 3), os usuários deram notas variando de 1 a 5 onde 1 representa uma baixa dificuldade e 5 alta dificuldade. Quando questionados sobre a utilização do sequenciador utilizado no método de Sanger, um total de 1 (2%) usuário marcaram a opção “1”, 3 (6%) marcaram a opção “2”, 11 (22%) marcaram a opção “3”, 17 (34%) usuários marcaram a opção “4” e 18 (36%)

marcaram a opção “5”. Já quando questionados sobre a dificuldade no manuseio do sequenciador utilizado na metodologia NGS, um total de 35 (70%) usuário marcaram a opção “1”, 12 (24%) marcaram a opção “2”, 2 (4%) marcaram a opção “3”, 0 (0%) usuários marcaram a opção “4” e 1 (2%) marcaram a opção “5”. A média das notas obtidas com relação ao manuseio do sequenciador utilizado em Sanger foi de 3,96 enquanto aos utilizados em NGS foi de 1,52.

Gráfico 3. Percepção dos usuários quanto ao nível de dificuldade no manuseio dos sequenciadores utilizados nos ensaios de Sanger e NGS

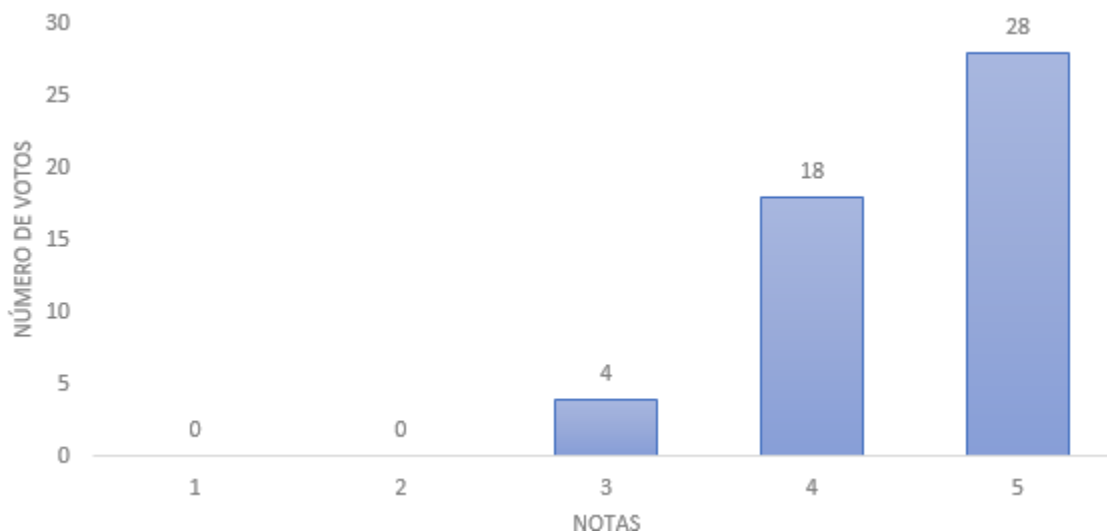


É sabido que a algumas plataformas de sequenciamento via método de Sanger exigem maior manuseio do equipamento do que outras, isso levando a uma maior dificuldade no método, como por exemplo ABI 3710 e ABI 3100 [108]. Além disso, é sabido também que o tempo em que as amostras precisam estar no sequenciador podem variar muito, podendo levar até uma noite inteira, dessa forma aumentando o tempo necessário para análise. Os sequenciadores NGS, em especial Illumina e Ion, diminuem consideravelmente a participação humana na configuração do *template* e inicialização das máquinas [109]. O tempo de processamento das amostras nos sequenciadores também diminuem muito quando comparado ao método de Sanger, podendo chegar até apenas 5 horas [110].

Outro tópico que devesse levar em contato é com relação à economia de consumíveis que tende de ser maior nos ensaios de NGS. Os usuários entrevistados deram notas de 1 a 5 com relação percepção deles com a economia

dos consumíveis (Gráfico 4), onde 1 significa uma baixa economia e 5 uma alta economia. Como resultado, nenhum dos entrevistados assinalou a opção “1” e também “2”, 4 (8%) marcaram a opção “3”, 18 (36%) usuários marcaram a opção “4” e 28 (56%) marcaram a opção “5” resultando uma nota média de 4,48.

Gráfico 4. Percepção dos usuários quanto a economia de insumos na técnica de NGS quando comparado ao método de Sanger

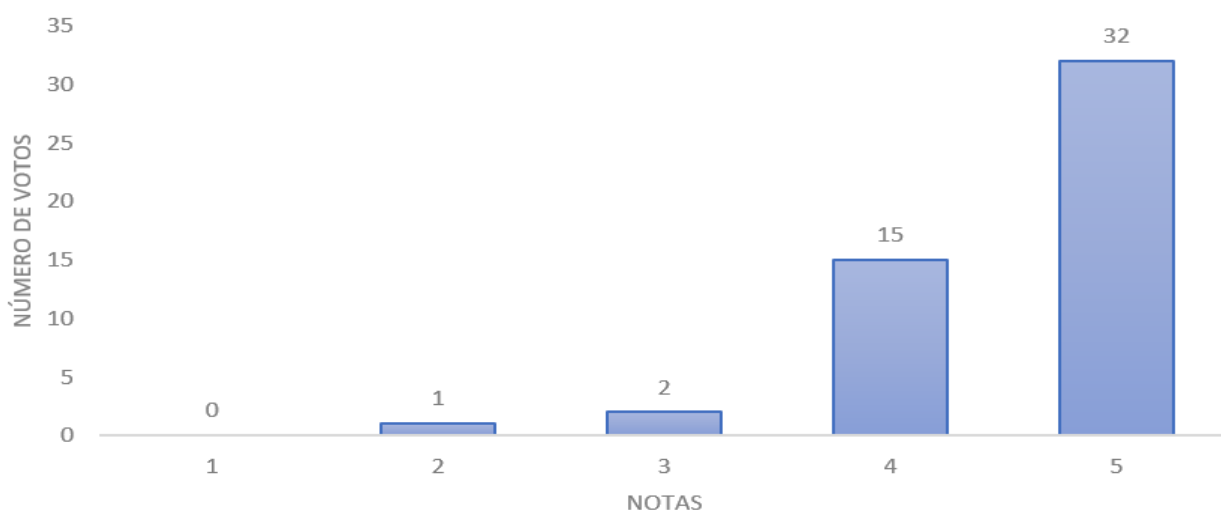


Além da economia de consumíveis, outro aspecto é a economia no tempo de bancada da técnica de NGS quando comparada com a técnica de Sanger. Novamente, os usuários deram notas de 1 a 5 com relação a percepção da economia de tempo (Gráfico 5) obtida nas técnicas de NGS, onde 1 significa uma baixa economia e 5 uma alta economia. Como resultado, obtemos que um total de 0 (0%) usuário marcaram a opção “1”, 1 (2%) marcaram a opção “2”, 2 (4%) marcaram a opção “3”, 15 (30%) usuários marcaram a opção “4” e 32 (64%) marcaram a opção “5”, a nota média da encontrada para este quesito foi de 4,56.

Como existem as plataformas *multiplex*, o *workflow* da metodologia de NGS permitem o preparo da biblioteca de forma conjunta de todos os genes que necessitam ser analisados, ou pelo de uma grande quantidade de uma única vez. Novamente, esse processo otimiza a rotina laboratorial fazendo com que haja não só uma economia no tempo levando a liberação de laudos de forma mais rápida como também a diminuição de utilização de insumos, como por exemplos os

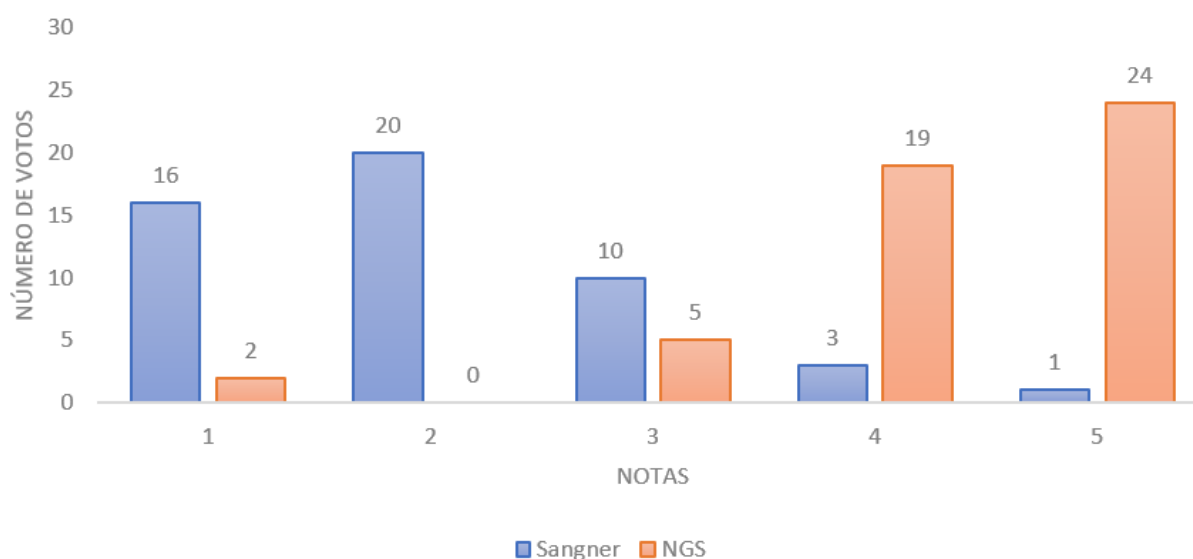
plásticos e reagente. As reações de sequenciamento pós a PCR na metodologia de Sanger precisam ocorrer separadamente para cada gene analisado, dessa forma demandando um alto tempo de bancada e utilização de insumos o que torna muito trabalhoso para o laboratorista.

Gráfico 5. Percepção dos usuários quanto a economia de tempo na execução da técnica de NGS quando comparado ao método de Sanger



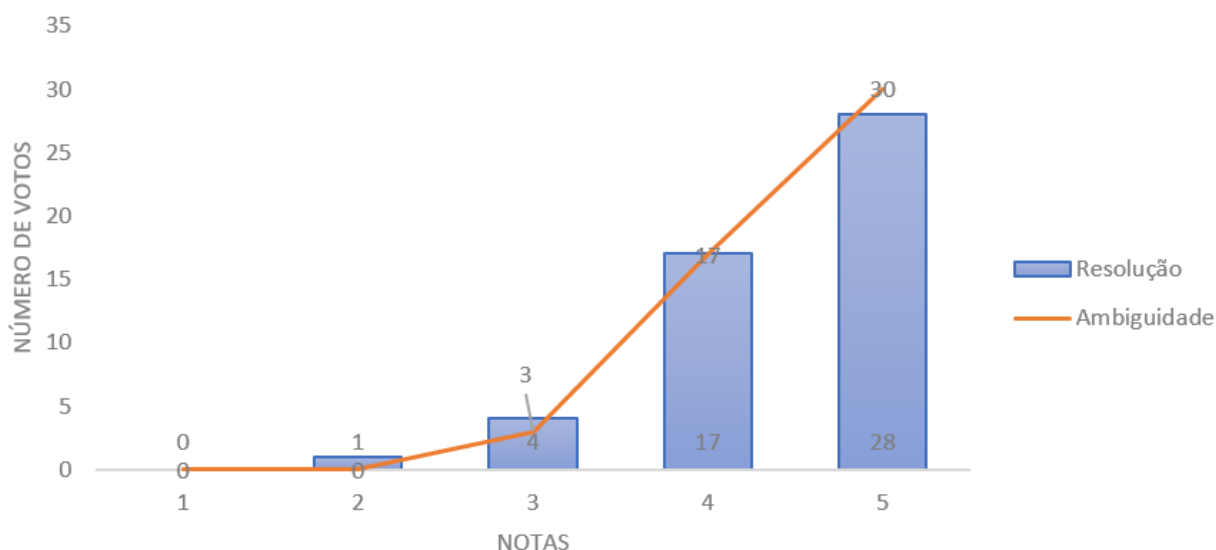
Durante o processo de liberação de resultados um dos processos mais importantes é a análise dos dados obtidos. Os *softwares* de análise dos dados de NGS e Sanger tendem a ser diferentes. Os usuários foram questionados sobre a facilidade de análise tanto no NGS e Sanger (Gráfico 6), classificando a facilidade de análise de 1 a 5 onde 1 representa uma baixa facilidade e 5 uma alta facilidade. Na análise Sanger, obtemos que um total de 16 (32%) usuário marcaram a opção “1”, 20 (40%) marcaram a opção “2”, 10 (20%) marcaram a opção “3”, 3 (6%) usuários marcaram a opção “4” e 1 (2%) marcaram a opção “5” onde a média de nota foi de 2,06. Já quando os usuários foram questionados da análise de NGS, obtemos que um total de 2 (4%) usuário marcaram a opção “1”, 0 (0%) marcaram a opção “2”, 5 (10%) marcaram a opção “3”, 19 (38%) usuários marcaram a opção “4” e 24 (48%) marcaram a opção “5” obtendo assim uma média nas nota de 4,26.

Gráfico 6. Percepção dos usuários quanto à facilidade durante o processo de análise de dados em NGS e em Sanger



A resolução de ambiguidades é algo necessário para resultados assertivos e em conjunto com o aumento da resolução da metodologia. Esses dois aspectos foram investigados a partir do questionário (Gráfico 7). Os usuários apontaram suas percepções das melhorias da tecnologia de NGS quando comparados com os resultados obtidos em metodologia Sanger para os parâmetros de Resolução de ambiguidades e resolução das técnicas. Com a relação à melhoria na resolução de ambiguidades por NGS em relação ao Sanger, os usuários classificaram de 1 a 5 onde 1 significa que as resoluções das ambiguidades não são diferentes entre as técnicas e 5 quando for altamente diferente. Como resultado, nenhum usuário marcou a opção “1” e “2”, 3 (6%) marcaram a opção “3”, 17 (34%) usuários marcaram a opção “4” e 30 (60%) marcaram a opção “5” onde a média da nota foi de 4,54. Já com relação a resolução em ambas técnicas, os usuários apontaram em escala de 1 a 5 a melhoria da resolução em NGS quando comparado com Sanger, onde 1 significa baixa diferença e 5 quando houver muita diferença. Como resultado, obtemos que um total de 0 (0%) usuário marcaram a opção “1”, 1 (2%) marcaram a opção “2”, 4 (8%) marcaram a opção “3”, 17 (34%) usuários marcaram a opção “4” e 28 (56%) marcaram a opção “5” onde a média da nota foi de 4,54.

Gráfico 7. Percepção dos usuários a melhoria da resolução e eliminação de ambiguidades na técnica de NGS em relação ao método de Sanger



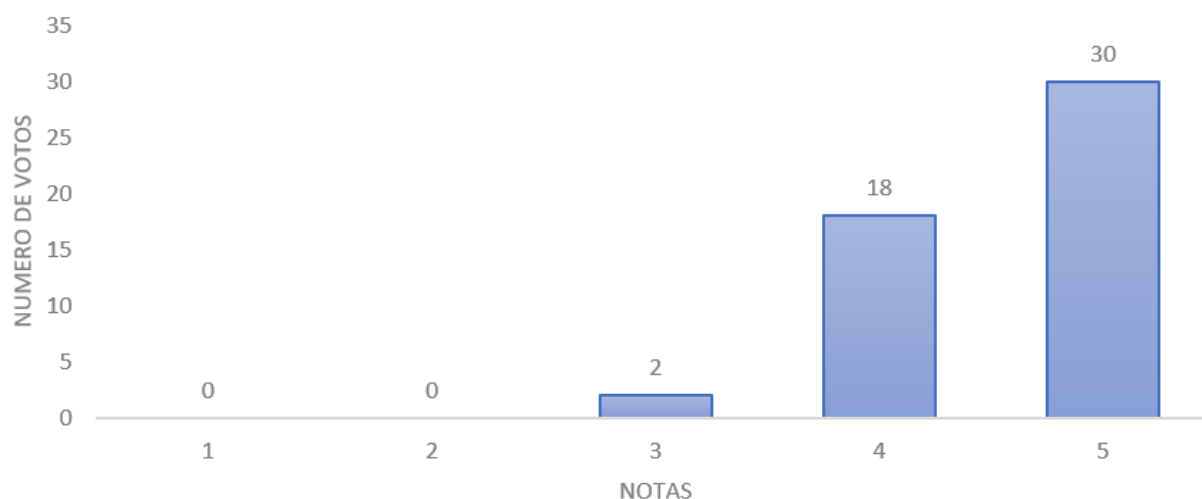
Como apontado anteriormente, as resoluções entre as metodologias são muito discrepantes entre si. Em Sanger as capacidades de detecção de bases são poucas e quanto nas plataformas de NGS a detecção de bases é muito alto, logo isso tem um impacto muito considerável na resolução que cada técnica possui, mostrando uma superioridade nas de NGS [114]. Com uma maior resolução as ambiguidades existentes tendem a ser muito menores também, isso por mostrar uma cobertura maior genômica, podendo chegar a análise de regiões intrônicas, chegando em resultados mais concisos, isso é de fundamental importância na tipagem de genes que possui um alto grau de polimorfismos e paralelismos, como o os genes HLA, onde a presença de ambiguidades devido a sequência muito semelhantes é comum. [111 - 115].

Além da diminuição das ambiguidades, o aumento da resolução aumenta também a facilidade de análise de dados. Em NGS os resultados oferecidos por softwares em raros casos necessitam de alteração manual dos resultados já oferecidos. Já na metodologia de Sanger, os dados tendem a exigir uma análise

mais minuciosa devido a existência de muitos backgrounds que podem dificultar o analista a liberar o resultado.

Por fim, os usuários foram questionados quanto às melhorias em gerais da metodologia NGS quando comprado com o Sanger, novamente os usuários apontaram de 1 a 5 onde 1 significa poucas melhorias e 5 muitas melhorias (Gráfico 8). Como resultado, nenhum usuário assinalou as opções “1” e “2”, 2 (4%) marcaram a opção “3”, 18 (36%) usuários marcaram a opção “4” e 30 (60%) marcaram a opção “5” onde a média da nota foi de 4,56.

Gráfico 8. Percepção dos usuários quanto às melhorias em termos gerais da metodologia NGS quando comparado ao Sanger



6. CONCLUSÃO

A partir dos dados presentes na literatura e na percepção mostrada pelos participantes do presente estudo é possível concluir que existe uma superioridade nas técnicas de NGS quando comparado às de Sanger. As técnicas de NGS mostraram superioridade em todos os quesitos analisados, bancada, análise e resolução dos dados. Por contar com plataformas multiplex, NGS otimiza muito o tempo de liberação de resultados e diminui a utilização de insumos plásticos, tornando as técnicas de sequenciamento muito mais barata e também diminuindo a contaminação ambiental que poderia ser causado pelo descarte inadequado dos plásticos. Atualmente a técnica de Sanger ainda é a mais empregada dentro do mercado, porém com o constante barateamento das técnicas de NGS é esperado que esse quadro se reverta em pouco tempo.

O constante avanço da tecnologia ainda deve melhorar muito os testes de sequenciamento dentro das análises clínicas. A evolução dentro dos testes de sequenciamento genômico é clara, diminuindo os custos e o tempo para elaborações dos testes e consequente liberação de laudos, porém esses testes ainda não são empregados em testes que necessitam uma necessidade de liberação de resultados de forma rápida, como por exemplo a tipagem HLA de um doador falecido, necessitando de testes menos sensíveis e consequentemente mais rápidos. Com a evolução da tecnologia é esperado que em breve isso se contorne, diminuindo ainda mais o tempo para liberação de resultados a partir de metodologias de sequenciamento que demandem ainda menos tempo e menos investimento financeiro para o seu emprego.

7. REFERÊNCIAS

- [1] Jay Shendure *et al.* "DNA sequencing at 40: past, present and future" *Nature*, V. 550, 345 – 553.
- [2] Sboner A, *et al.*: The real cost of sequencing: higher than you think! *Genome Biology* 2011, 12:125.
- [3] Metzker M. "Sequencing technologies — the next generation" *Nature Reviews*, 2010, v.11, 31-46.
- [4] Onsongo *et al.*: Implementation of Cloud based Next Generation Sequencing data analysis in a clinical laboratory. *BMC Research Notes* 2014, v7:314.
- [5] Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW: Cancer genome landscapes. *Science* 2013, v339:1546–1558.
- [6] Salipante SJ, *et al.* "Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections." *PLoS One* 2013, 8:e65226.
- [7] Barzon L. *et al.* "Next-generation sequencing technologies in diagnostic virology" *Journal of Clinical Virology*, 2013, v58, 346-350
- [8] M. Kotowski *et al.* "The Importance of New Generation Sequencing (NGS) HLA Typing in Renal Transplantation Preliminary Report" *Transplantation Proceedings*, 2018, v50, 1605-1615
- [9] Heath J. e Chain B. "The sequence of sequencers: The history of sequencing DNA" *Genomics* Volume 107, Issue 1, January 2016
- [10] Shendure J. e Aind E.L "Review Article | Published: 08 November 2012 The expanding scope of DNA sequencing" *Nature Biotechnology*, 2012, v30, pages 1084–1094
- [11] Sanger F. e Coulson A.R "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase" *JMB*, V.94, Issue 3, 25 May 1975
- [12] Mullis K *et al* "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." *Cold Spring Harb Symp Quant Biol.* 1986;51
- [13] Ansorg W *et al* "One label, one tube, Sanger DNA sequencing in one and two lanes on a gel." *Nucleic Acids Res.* 1990 Jun 11; v18(11): 3419–3420.

- [14] Livro BIOLOGIA MOLECULAR BASICA, 5ª edição – Arnaldo Zaha e colaboradores
- [15] Alberts, B., et.al. 2002. Molecular Biology of the Cell. 4th ed. New York. Garland Science. página 505.
- [16] Cleveland D.W et al "Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis." The Journal of Biological Chemistry, v252, 1102-1106.
- [17] Sanger. F e Coulson AR "The use thin acrylamide gels for DNA sequencing" FEBES, v87, 1978
- [18] Zimmermann et al "Automated Sanger dideoxy sequencing reaction protocol" FEBES Letters, 1988 v233, Issue 2, Pages 432-436
- [19] Tang S. et al "Semi-Automatic In Silico Gap Closure Enabled De Novo Assembly of Two Dehalobacter Genomes from Metagenomic Data" PLOS ONE, 2012 December v21
- [20] Gingereas T. et al "A semi-automated method for the reading of nucleic acid sequencing gels" Nucleic Acids Research, v10, Issue 1, 11 January 1982, Pages 103–114,
- [21] Toonen R. e Hughes S. "Increased Throughput for Fragment Analysis on an ABI PRISM ® 377 Automated Sequencer Using a Membrane Comb and STRand Software" BioTechniques, 2001, v31:1320-1324
- [22] Smith T et al "PCR-based setup for high-throughput cDNA library sequencing on the ABI 3700 automated DNA sequencer" Biotechniques 2001, v30(2):394.
- [23] Koumi. P et al "Evaluation and validation of the ABI 3700, ABI 3100, and the MegaBACE 1000 capillary array electrophoresis instruments for use with short tandem repeat microsatellite typing in a forensic environment" Electrophoresis, v25, Issue1, 2004, Pages 2227-2241
- [24] Zimmermann J. et al "Automated Sanger dideoxy sequencing reaction protocol" FEBES Letter, v233, Issue 2, 20 June 1988, Pages 432-436
- [25] Lagally E.T. et al "Fully integrated PCR-capillary electrophoresis microsystem for DNA analysis" Lab on a Chip, 2001, v1, 102–107

- [26] Lu. W et al "Multi-target PCR analysis by capillary electrophoresis and laser-induced fluorescence" *Nature*, 1994, v368, 269–271
- [27] Fakruddin Md e Chowdhury A. "Pyrosequencing-An Alternative to Traditional Sanger Sequencing" *American Journal of Biochemistry and Biotechnology*, 2012, v8 (1): 14-20
- [28] Hamilton A. "Concordance between whole-exome sequencing and clinical Sanger sequencing: implications for patient care" *Molecular Genetics & Genomic Medicine* 2016, v4(5): 504–512
- [29] Wiley-Blackwell "Next Generation Genome Sequencing: Towards Personalized Medicine", 2008
- [30] Leandro Marcio Moreira "Ciencias genomicas: fundamentos e aplicações" Sociedade brasileira de genética, PUC-Goiás, 2013
- [31] Luo C. et al "Direct Comparisons of Illumina vs. Roche 454 Sequencing Technologies on the Same Microbial Community DNA Sample" *PLOS ONE*, 2012, v7, Issue 2
- [32] Rothberg J. M e Leamon J.H "The development and impact of 454 sequencing" *Nature Biotechnology*, 2008, v26, pages 1117–1124.
- [33] Meyer, M., Stenzel, U. & Hofreiter, M. Parallel tagged sequencing on the 454 platform". *Nat. Protocols* 3, 2008, 267–278
- [34] D., Luo, S., et al. "Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays." *Nat. Biotechnol.* 2000, v18, 630–634.
- [35] Ellegren H "Sequencing goes 454 and takes large-scale genomics into the wild" *Molecular Ecology*, 2008, v17, Issue7 Pages 1629-1631
- [36] Vera J.C et al "Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing" *Molecular Ecology*, 2008, v17, Issue7, 1636-1647
- [37] Galan M. et al "Methodology article A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies" *BMC Genomics*, 2010, v11:296
- [38] Stoddard J.L "Targeted NGS: a cost-effective approach to molecular diagnosis of PIDs" *Front. Immunol.*, 2014, v3:531

- [39] Scheible M et al "Short tandem repeat sequencing on the 454 platform" Forensic Science International: Genetics Supplement Series, 2001, v3, Issue 1, 357-358
- [40] DiGuistini S. et al "De novo genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data" Genome Biology, 2009, v10:R94
- [41] Quail M.A et al "A large genome center's improvements to the Illumina sequencing system" Nature Methods, 2008, v5, 1005–1010
- [42] Holt R. A e Jone S.J.M "The new paradigm of flow cell sequencing" Genome Res. 2008, v18, 839-846
- [43] Grada A. e Weinbrecht k. "Next-Generation Sequencing: Methodology and Application" Journal of Investigative Dermatology, 2013, v133, e11.
- [44] Kozarewa I, et al "Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes" Nature Methods, 2009, v6, 291–295.
- [45] Tindall E. A et al "Interpretation of custom designed Illumina genotype cluster plots for targeted association studies and next-generation sequence validation" BMC Research Notes, 20103 v39.
- [46] Kircher M. et al "Addressing challenges in the production and analysis of illumina sequencing data" BMC Genomics, 2011, v12:382.
- [47] Bokulich N. A " Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing" Nature Methods, 2013, v10, 57–59.
- [48] Roberts R. J "The advantages of SMRT sequencing" Genome Biology, 2013, v14:405.
- [49] Park P. J "ChIP–seq: advantages and challenges of a maturing technology" Nature Reviews Genetics, 2009, v10, 669–680.
- [50] Nelson M.C. "Analysis, Optimization and Verification of Illumina-Generated 16S rRNA Gene Amplicon Surveys" PLOS ONE, 2014, v9, Issue 4.
- [51] Applied Biosystems: Applied Biosystems homepage. [<http://www.Appliedbiosystems.com>], Acesso em 08/05/2019 às 13:45.
- [52] Hedges D.J "Comparison of Three Targeted Enrichment Strategies on the SOLiD Sequencing Platform" Platform. PLoS ONE, 2011, v6(4): e18595.

- [53] Hultman T, et al "Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support" *Nucleic Acids Research*, 1989, v17, Issue 13, 4937–4946.
- [54] Mardis E. "The impact of next-generation sequencing technology on genetics" *Trends in Genetics*, 2008, v24, Issue 3.
- [55] Zolet C.T et al "Capítulo 8 Polimorfismo de Nucleotídeo único (SNP): metodologias de identificação, análise e aplicações" 201, disponível em: https://www.researchgate.net/publication/320624547_Capitulo_8_Polimorfismo_de_Nucleotideo_unico_SNP_metodologias_de_identificacao_analise_e_aplicacoes.
- [56] Ari Ş., Arikan M. "Next-Generation Sequencing: Advantages, Disadvantages, and Future". *Plant Omics: Trends and Applications*. Springer, Cham, 2016.
- [57] Quail M. A et al "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers" *BMC Genomics*, 2012, v13:341
- [58] Wotlei A D. "Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform" *Journal of Microbiological Methods*, 2012, v91, Issue 1.
- [59] Loman N et al "Performance comparison of benchtop high-throughput sequencing platforms" *Nature Biotechnology*, 2012, v30, pages 434–439.
- [60] Shedure J.A "Overview of DNA Sequencing Strategies" *Current Protol*, 2011, v96, Issue1.
- [61] Deagle B. E et al "Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count?" *Molecular Ecology* 2013, v13, Issue4.
- [62] Shokralla S. et al "Next-generation sequencing technologies for environmental DNA research" *Molecular Ecology*, 2012, v21(8),1794-805.
- [63] Woo P.C.Y et al "Usefulness of the MicroSeq 500 16S Ribosomal DNA-Based Bacterial Identification System for Identification of Clinically Significant Bacterial Isolates with Ambiguous Biochemical Profiles", *JOURNAL OF CLINICAL MICROBIOLOGY*, 2003, 1996–2001

- [64] Lau. S.K.P et al "Catheter-Related Microbacterium Bacteremia Identified by 16S rRNA Gene Sequencing", JOURNAL OF CLINICAL MICROBIOLOGY, 2002, 2681–2685
- [65] Hummelen. R et al "Deep Sequencing of the Vaginal Microbiota of Women with HIV" PLoS ONE, 2010, v5, Issue 8.
- [66] Greub. G et al "High Throughput Sequencing and Proteomics to Identify Immunogenic Proteins of a New Pathogen: The Dirty Genome Approach" PLoS ONE 2009, v4, Issue 12.
- [67] Ringuet. H et al "hsp65 Sequencing for Identification of Rapidly Growing Mycobacteria" JOURNAL OF CLINICAL MICROBIOLOGY, 1999, 852–857.
- [68] Senna S.G et al "Sequencing of hsp65 Gene for Identification of Mycobacterium Species Isolated from Environmental and Clinical Sources in Rio de Janeiro, Brazil", JOURNAL OF CLINICAL MICROBIOLOGY, 2008, 3822–3825.
- [69] Bartau S. G et al "Development of a Multilocus Sequence Typing Scheme for Characterization of Clinical Isolates of *Acinetobacter baumannii*" JOURNAL OF CLINICAL MICROBIOLOGY, 2005, 4382–4390.
- [70] Costa. V.M.D et al "Antibiotic resistance is ancient" Nature, v477, 457–461.
- [71] White O. et al "Genome Sequence of the Radioresistant Bacterium *Deinococcus radiodurans* R1" Science, 1999, 286, Issue 5444, 1571-1577.
- [72] Takiff H.E et al "Cloning and nucleotide sequence of *Mycobacterium tuberculosis* gyrA and gyrB genes and detection of quinolone resistance mutations." ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, 1994, 773-780.
- [73] Ashton P.M. et al "MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island" Nature Biotechnology, 2015, v33, 296–300.
- [74] Jayme Parker & Jack Chen "Application of next generation sequencing for the detection of human viral pathogens in clinical specimens" Journal of Clinical Virology, 2017, v86, 20–26
- [75] D. Bzhalava, et al "Unbiased Approach for Virus Detection in Skin Lesions" PLOS ONE, 2013, v8, Issue 6.

- [76] M.K. Sahoo, et al., "Detection of cytomegalovirus drug resistance mutations bynext-generation sequencing", *J. Clin. Microbiol*, 2013, v51 (11), 3700–3710.
- [4] Hang. Y et al "Analysis of hepatitis B virus genotyping, and drug resistance gene mutations based on massively parallel sequencing" *Journal of Virological Methods*, 2013, v193, 341– 347.
- [77] Anthony SJ. et al "A Strategy To Estimate Unknown Viral Diversity in Mammals" *MBio*, 2013 v4(5), Issue 5.
- [78] Dong J et al. "Emerging pathogens: challenges and successes of molecular diagnostics". *J Mol Diagn*, 2008, v10, 185-197.
- [79] Gard. G et al "A Novel Rhabdovirus Associated with Acute Hemorrhagic Fever in Central Africa" *PLOS Pathogenens*, 2012, v8, Issue 9.
- [80] Barzon L. et al "Human papillomavirus genotyping by 454 next generation sequencing technology." *J Clin Virol*, 2011, v52(2), 93-97.
- [81] Galloway. D "Papillomavirus vaccines in clinical trials" *The Lancet*, 2003, v3, Issue 8, 469-475.
- [82] Quick J. et al "Real-time, portable genome sequencing for Ebola surveillance" *Nature volume*, 2016, v530, 228–232.
- [83] Chan. YP et al "Complete nucleotide sequences of Nipah virus isolates from Malaysia." *Journal of General Virology*, 2001, v82, 2151–2155.
- [84] Trebbien R. et al "Development of oseltamivir and zanamivir resistance in influenza A(H1N1)pdm09 virus" *Euro Surveill*, 2017, v22(3).
- [85] MichaelEdidin "MHC antigens and non-immune functions" *Immunology Today*, 1983, v4, Issue 10, 269-270.
- [86] Montgomery R. A et al "HLA in transplantation" *Nature Reviews Nephrology*, 2018, v14, 558–570.
- [87] Kuzmich E. et al "Identification of a new HLA-DQB1*06 allele, HLA-DQB1*06:210, by monoallelic Sanger sequencing" *HLA Immune Response Genetics*, 2017, v90, 132-133.
- [88] Kuzmich E. et al "Detection of a new HLA-B*44 allele, HLA-B*44:02:45, by monoallelic Sanger sequencing" *HLA Immune Response Genetics*, 2017, v90, 124.

- [89] Halcomb C.L. et al "A multi-site study using high-resolution HLA genotyping by next generation sequencing" *Tissue Antigens*, 2011, v77, 206–217
- [90] Yawata. M et al "Roles for HLA and KIR polymorphisms in natural killer cell repertoire selection and modulation of effector function" *The Journal of Experimental Medicine*, 2006, v203, 633–645.
- [91] Long E. O. et al "Killer cell inhibitory receptors: diversity, specificity, and function" *Immunological Reviews*, 1997, v155, Issue 1, 135-144
- [92] Bashirova AA et al "The killer immunoglobulin-like receptor gene cluster: tuning the genome for Defense". *Annu Rev Genomics Hum Genet*, 2016, v 7, 277–300.
- [93] Tomblyn M et al "Decreased Infections in Recipients of Unrelated Donor Hematopoietic Cell Transplantation from Donors with an Activating KIR Genotype" *Biology of Blood and Marrow Transplantation*, 2010, v16, Issue 8, 1155-1161.
- [94] Martin MP et al "Innate partnership of HLA-B and KIR3DL1 subtypes against HIV1". *Nat Genet*, 2007, v39, 733–740
- [95] Moorcraft S.Y, et al."Understanding next generation sequencing in oncology: A guide for oncologists", *Crit Rev Oncol/Hematol* 2015, v96(3), 463-474.
- [96] Ford D. et al."Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families". *The breast cancer linkage consortium*, *Am. J. Hum. Genet*, 1998, v62, 676–689.
- [97] K. Cibulskis et al., "Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples" *Nat. Biotechnol.* 2013, v31, 213–219.
- [98] Sjögren S et al "The p53 Gene in Breast Cancer: Prognostic Value of Complementary DNA Sequencing Versus Immunohistochemistry" *JNCI: Journal of the National Cancer Institute*, 1996, v88, Issue 3-4, 173–182.
- [99] Castera, L. et al. "Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes" *Eur. J. Hum. Genet*, 2014, v22, 1305–1313.
- [100] Tung, N. et al. "Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel". *Cancer*, 2015, v121, 25–33.

- [101] Aloraifi, F. et al. "Detection of novel germline mutations for breast cancer in non-BRCA1/2 families". *FEBS J*, 2015, v282, 3424–3437.
- [102] Cragun, D. "Panel-based testing for inherited colorectal cancer: A descriptive study of clinical testing performed by a US laboratory." *Clin. Genet*, 2014, v86, 510–520.
- [103] Frey, M.K "Rescreening for genetic mutations using multi-gene panel testing in patients who previously underwent non-informative genetic screening. *Gynecol. Oncol.* 2015, v139, 211–215.
- [104] Treff, N.R. et al. "Next-generation sequencing for preimplantation genetic diagnosis". *Fertil. Steril.* 2013, v99, e17–e18.
- [105] Huang, L. et al "Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications." *Annu. Rev. Genom. Hum. Genet*, 2015, v16, 79–102.
- [106] Boccoz A. S et al "Massively parallel and multiplex blood group genotyping using nextgeneration-sequencing" *Clinical Biochemistry*, 2018, v60, 71-76
- [107] Product Inser of AllType NGS, One lambda - disponível em "<https://www.onelambda.com/en/product/alltype-ngs-reagents.html>" consultado em 15/10/2019 às 17:16
- [108] Koumi P. et al "Evaluation and validation of the ABI 3700, ABI 3100, and the MegaBACE 1000 capillary array electrophoresis instruments for use with short tandem repeat microsatellite typing in a forensic environment" *Electrophoresis* 2004, v25, 2227–2241.
- [109] Tuononen K. "Comparison of Targeted Next-Generation Sequencing (NGS) and Real-Time PCR in the Detection of EGFR, KRAS, and BRAF Mutations on Formalin-Fixed, Paraffin-Embedded Tumor Material of Non-Small Cell Lung Carcinoma—Superiority of NGS" *Genes, Chromosomes and Cancer*, 2013, v52, Issue 5, 503-511.
- [110] Cousins M. "Comparison of a high-resolution melting assay to next-generation sequencing for analysis of HIV diversity" *Journal of clinical microbiology*, 2012, v50(9), 3054-3059.

- [111] Lind C, et al "Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing" Human Immunology, 2010, v71(10), 1033-1042
- [112] Duke J et al "P154 Enumeration based ambiguous genotype detection for identifying ambiguous HLA allele combinations in the presence of intra-exon phasing" Human Immunology, 2017, v78, 169
- [113] Melista, E et al "Towards a new gold standard - NGS corrections to sanger SBT genotyping results" Human Immunology, 2015, v76, 148
- [114] Profazer T et al "Comparing fragment size with read-length and their effect on HLA genotyping by next-generation sequencing (NGS)" Human Immunology, 2015, v76, 52
- [115] Duke J et al "Determining performance characteristics of an NGS-based HLA typing method for clinical applications" HLA, 2016, v87(3), 141-152

8. ANEXOS

7.1. Questionário aplicado:

NGS x Sanger

Caso tenha alguma dúvida sobre o questionário, pode me contactar através do e-mail luc.raf1997@gmail.com

Agradeço imensamente pelas respostas e pela ajuda nesse processo.

* Required

Qual a sua área de atuação?

Your answer

Em qual (quais) das plataformas de NGS você possui experiência/conhecimento ? *

☐ Illumina

☐ Ion

☐ 454

☐ SOLiD

Com relação ao SBT (Sanger), como você classificaria a dificuldade no processo de amplificação inicial? Considere 1 como BAIXA dificuldade e 5 como ALTA dificuldade. *

1	2	3	4	5
☐	☐	☐	☐	☐

Com relação a etapa de reações de sequenciamento na metodologia de Sanger, qual seria o nível de dificuldade? Considere 1 como BAIXA dificuldade e 5 como ALTA dificuldade. *

1	2	3	4	5
☐	☐	☐	☐	☐

Com relação ao sequenciador utilizado nos ensaios de Sanger, como você classificaria a dificuldade de manuseio do equipamento? Considere 1 como BAIXA dificuldade e 5 como ALTA dificuldade. *

1	2	3	4	5
☐	☐	☐	☐	☐

Com relação ao NGS, como você classificaria a dificuldade no processo de amplificação inicial? Considere 1 como BAIXA dificuldade e 5 como ALTA dificuldade. *

1	2	3	4	5
☐	☐	☐	☐	☐

Como você classificaria a dificuldade no processo de preparo da biblioteca na metodologia de NGS? Considere 1 como BAIXA dificuldade e 5 como ALTA dificuldade. *

1	2	3	4	5
☐	☐	☐	☐	☐

Com relação ao sequenciador utilizado nos ensaios de NGS, como você classificaria a dificuldade de manuseio do equipamento? Considere 1 como BAIXA dificuldade e 5 como ALTA dificuldade. *

1	2	3	4	5
☐	☐	☐	☐	☐

Quanto você caracteriza a economia na utilização de insumos utilizados nas técnicas de NGS quando comparado as técnicas de Sanger? Considere 1 como BAIXA diferença de econômica e 5 como uma ALTA diferença econômica *

1	2	3	4	5
☐	☐	☐	☐	☐

Quanto você caracteriza a economia de tempo utilizados nas técnicas de NGS quando comparado as técnicas de Sanger? Considere 1 como BAIXA economia de tempo e 5 como ALTA diferença na economia de tempo *

1	2	3	4	5
☐	☐	☐	☐	☐

Como você classificaria a diferença de resolução obtida pelo NGS quando comparado ao Sanger? Considere 1 se as diferenças de resolução NÃO FOR SIGNIFICATIVA e 5 como ALTAMENTE SIGNIFICATIVA. *

1	2	3	4	5
☐	☐	☐	☐	☐

Com relação a análise de dados obtidos a partir da metodologia Sanger, qual seria o nível de facilidade de análise? Considere 1 como BAIXA FACILIDADE de análise e 5 como GRANDE FACILIDADE de análise *

1	2	3	4	5
☐	☐	☐	☐	☐

Com relação a análise de dados obtidos a partir da metodologia NGS, qual seria o nível de dificuldade de análise? Considere 1 como BAIXA FACILIDADE de análise e 5 como GRANDE FACILIDADE de análise *

1	2	3	4	5
☐	☐	☐	☐	☐

Como você classificaria a resolução de ambiguidades por NGS quando comparado a metodologia de Sanger ? Considere 1 quando NÃO HAVER diferenças e 5 quando HOVER grande diferença positiva *

1	2	3	4	5
☐	☐	☐	☐	☐

Em termos gerais, como você classificaria as melhorias do NGS quando comparado ao Sanger? Considere 1 como POUCAS melhorias e 5 como MUITAS melhorias *

1

2

3

4

5

☐☐☐☐☐**SUBMIT**