

CAPÍTULO 5

TECNOLOGIAS DE SEQUENCIAMENTO: DO SANGER AO NGS

Carolina Bustamante

Especialista em Genética Molecular na OC Precision Medicine –
Grupo Oncoclínicas, Rio de Janeiro

Mariano Zalis

Diretor de Pesquisa e Desenvolvimento na OC Precision
Medicine – Grupo Oncoclínicas, Rio de Janeiro

Tecnologias de sequenciamento: do Sanger ao NGS

A capacidade de sequenciar o DNA constitui uma das etapas mais importantes para o estabelecimento da Medicina de Precisão. No entanto, o progresso desde a descoberta da estrutura tridimensional do DNA em 1953 por Watson e Crick até o Sequenciamento de Nova Geração é o resultado de um grande esforço da comunidade científica. Desde que o DNA foi isolado pela primeira vez em 1869 por Friedrich Mietscher, foram realizadas diversas tentativas para sequenciar esta molécula longa e de fita dupla, mas o primeiro esboço de sequenciamento começou com moléculas de RNA, porque eram mais fáceis de serem obtidas, compostas de fita simples e mais curtas que o DNA. Tal esboço criou bases para o desenvolvimento de técnicas de sequenciamento do DNA e, eventualmente em 1972, Walter Fiers realizou o primeiro sequenciamento completo do DNA de um gene, que codifica a proteína capsidial do bacteriófago MS2¹.

No entanto, o grande avanço que alterou para sempre o progresso da tecnologia de sequenciamento de DNA veio em 1977, com o desenvolvimento da “terminação de cadeia” de Sanger², com a utilização de didesoxinucleotídeos (ddNTPs), que carecem do grupo hidroxila 3', necessário para a extensão das cadeias de DNA, culminando em fitas de DNA de comprimentos diferentes e permitindo sua separação por tamanho. Sanger, então, realizou quatro reações paralelas contendo uma base diferente de ddNTP em cada, aplicou o resultado em quatro poços distintos de um gel de poliacrilamida e usou a autorradiografia para inferir qual era a sequência de nucleotídeos. A precisão, a robustez e a facilidade desta técnica levaram o método de terminação de cadeia didesoxi – ou simplesmente, sequenciamento de Sanger – a se tornar a tecnologia mais comum usada para sequenciar o DNA nos 30 anos seguintes¹.

Uma série de melhorias foi criada nos anos seguintes, o que envolveu principalmente a substituição da radiomarcagem por detecção com base fluorométrica (permitindo a realização de apenas uma reação) e a detecção aprimorada por meio de eletroforese capilar. Essas duas melhorias contribuíram para o desenvolvimento de máquinas de sequenciamento de DNA cada vez mais automatizadas e, subsequentemente, para a produção de máquinas comerciais, como as da linha ABI PRISM, produzida pela Applied Biosystems, que permitia o sequenciamento simultâneo de diversas amostras. O sequenciamento automático, aliado à técnica de *shotgun*, que permitia a clonagem de fragmentos longos de DNA, aceleraram a conclusão do Projeto Genoma Humano, ajudando a produzir o primeiro rascunho deste empreendimento gigantesco anos antes do previsto¹.

O sequenciamento de Sanger e a publicação da sequência de referência do genoma humano ofereceram novas estratégias e oportunidades para o conhecimento dos genomas do câncer, culminando na identificação de mutações acionáveis, preditivas de resposta ou de resistência a tratamentos no câncer. No início, esta abordagem foi utilizada para um pequeno número de

genes devido às baixas quantidade e qualidade do DNA obtido de células tumorais. Mesmo assim, esse esforço inicial ajudou na descoberta de mutações acionáveis no gene EGFR em 2004, que previam a resposta aos inibidores de tirosina quinase (TKIs) no tratamento de adenocarcinoma de pulmão. A descoberta das mutações em EGFR foi responsável pelo lançamento da Era da Medicina Personalizada no câncer de pulmão de não pequenas células (CPNPC) avançado, levando a uma mudança dramática no cenário terapêutico desta doença, ao permitir a seleção de pacientes para tratamento baseado nas características moleculares dos seus tumores³. Também em 2004 foi aprovado o tratamento com o primeiro anticorpo monoclonal anti-EGFR para câncer colorretal metastático⁴. Mas logo depois, em 2006, foi relatada redução da eficácia do tratamento em pacientes com mutações nos códons 12 e 13 do gene KRAS. Estudos subsequentes demonstraram que pacientes com mutações nos genes NRAS, BRAF e PIK3CA também apresentavam resistência à terapia anti-EGFR⁵.

O sequenciamento de Sanger também foi utilizado para mapear as mutações no genoma tumoral de 11 pacientes com câncer colorretal e 11 com câncer de mama em um empreendimento pioneiro. Embora tenha utilizado um número limitado de amostras, tal trabalho foi o primeiro a caracterizar as mutações presentes no genoma do câncer e a identificar mutações *drivers* como ponto de vulnerabilidade para desenvolvimento de novas terapias-alvo⁶.

Apesar do sucesso do sequenciamento de Sanger, essa técnica se mostrou ineficiente e de alto custo para projetos de larga escala. Apesar de ser uma metodologia com grande acurácia, produzindo sequências de alta qualidade com até 900 pares de base, não é eficaz para detecção de alterações de baixa frequência alélica, sendo ineficiente para a detecção de alterações somáticas. Adicionalmente, esta metodologia é normalmente capaz de detectar um alvo por vez, não tendo uma boa relação custo-benefício para detecção de múltiplos alvos ou de genes inteiros⁷.

Foi então que, a partir de 2005, surgiram as primeiras tecnologias de Sequenciamento de Nova Geração (NGS), também conhecidas como sequenciamento de DNA paralelo massivo ou sequenciamento de alta performance⁸. Além de melhorar a compreensão do câncer, o NGS promoveu o nascimento de uma nova forma de tratar pacientes, que hoje chamamos de Medicina de Precisão. No contexto da Oncologia, isso significa adaptar tratamentos para as características de cada paciente e de cada alteração genômica encontrada em seu tumor. Foi possível sair do conceito de medicamentos *one-size-fits-all*, ou seja, aquele que serve a todos, para a estratificação progressiva de pacientes de acordo com seus subtipos de doença, características clínicas e biomarcadores presentes em seus tumores para a seleção de uma terapia-alvo personalizada⁹.

Sequenciamento de Segunda Geração

Simultaneamente ao desenvolvimento do sequenciamento de Sanger, outra técnica desenvolvida preparou o terreno para a primeira onda de sequenciadores de nova geração. Este método diferia dos já existentes por não utilizar dNTPs ou oligonucleotídeos marcados com rádio ou fluorescência. Em vez disso, era utilizado um método luminescente para medir a síntese de pirofosfato. Esta abordagem foi utilizada para inferir a sequência, medindo a produção de pirofosfato à medida que cada nucleotídeo era incorporado ao DNA molde. Esta técnica tinha como vantagem o uso de nucleotídeos não marcados, em vez dos dNTPs fortemente modificados usados nos protocolos de terminação de cadeia, e podia ser observada em tempo real. A maior dificuldade era determinar precisamente a sequência de homopolímeros, ou seja, sequências repetitivas de um único nucleotídeo. O pirosequenciamento foi posteriormente licenciado para a 454 Life Sciences em 2005, quando evoluiu para a primeira grande tecnologia comercial de NGS.

As máquinas de sequenciamento produzidas pela 454 Life Sciences (posteriormente adquirida pela Roche) foram uma mudança de paradigma, pois permitiram o sequenciamento de amostras em paralelo de forma escalável e com automação significativamente maior do que os instrumentos padrões de eletroforese capilar. O equipamento da empresa 454 era capaz de produzir até 2 milhões de leituras e 80Mb por dia, representando um aumento em 100 vezes na produção de sequências de DNA se comparado à tecnologia existente na época¹⁰. Essa tecnologia aumentou o rendimento de sequenciamento por ordem de magnitudes, permitindo aos pesquisadores sequenciar completamente o genoma humano muito mais rapidamente e com menor custo.

O processo utilizado pela 454 para sequenciamento foi a PCR em emulsão, em que moléculas de DNA são primeiro anexadas a microesferas por meio de sequências adaptadoras, que então passam por uma PCR em emulsão de água em óleo (em PCR). Neste processo, cada gotícula possui uma molécula de DNA ligada a uma microesfera e os componentes necessários para a PCR, funcionando como um microrreator e multiplicando a molécula de DNA, que passa a revestir a superfície da microesfera. Essas esferas revestidas com DNA são então levadas a uma placa de picotitulação, composta de micropoços, junto com esferas contendo um coquetel enzimático e dNTPs. À medida que as dNTPs são incorporadas à fita de DNA, ocorre uma cascata enzimática resultando em um sinal bioluminescente, que é detectado por uma câmera acoplada abaixo dos poços.

Uma série de sequenciadores surgiram após o sucesso do 454 (figura 1). Um deles foi o da empresa Solexa, criadora do método de amplificação em ponte (*bridge amplification*), posteriormente adquirida pela Illumina. A inovação desta nova plataforma de sequenciamento consiste na síntese de DNA diretamente na *flowcell*, uma superfície sólida de vidro com microcanais em seu interior. A *flowcell* é dividida em linhas contendo diversos adaptadores fixados em sua

superfície. Por sua vez, os fragmentos de DNA também são ligados a adaptadores em suas extremidades, permitindo sua imobilização por hibridização aos adaptadores da *flowcell*. Na primeira etapa, são adicionados nucleotídeos não marcados e uma polimerase sintetiza uma fita complementar à imobilizada no suporte. O DNA é posteriormente desnaturado e a fita que não está ligada ao suporte é removida por lavagem. Em seguida, ocorre a formação de pontes, com o adaptador da outra extremidade da fita se hibridizando a um outro oligo na superfície da *flowcell*. O DNA é, então, novamente amplificado, formando uma fita complementar nova, mas desta vez, cada fita está ligada ao suporte. Todo este processo é repetido, gerando múltiplas cópias do DNA. No fim do processo, a DNA polimerase utiliza os adaptadores como *primer*, se liga e estende, adicionando nucleotídeos marcados que emitem uma fluorescência, reconhecida pelo equipamento (figura 1). A primeira versão de equipamento lançada foi o Genome Analyzer em 2006, posteriormente substituída pelo HiSeq, uma máquina capaz de comprimento e profundidade de leitura ainda maiores. Atualmente, a empresa possui um grande número de equipamentos com diferentes capacidades de sequenciamento, sendo o Novaseq o mais atual, com capacidade de gerar até 6Gb com 20 bilhões de bases sequenciadas entre 19 e 40 horas.

Veja esse conteúdo na íntegra
acessando a OC Academia:
<http://ocacademia.com>

